

DE GRUYTER  
MOUTON

# EXPLORING STRUCTURES IN LANGUAGES AND LANGUAGE CONTACT

*Edited by Nataliya Levkovich,  
Julia Nintemann and Maike Vorholt*

DE  
G

Nataliya Levkovich, Julia Nintemann and Maike Vorholt (Eds.)  
**Exploring Structures in Languages and Language Contact**



# Exploring Structures in Languages and Language Contact



Studies in honour of Thomas Stolz.  
On occasion of his retirement.

Edited by  
Nataliya Levkovych, Julia Nintemann and Maike Vorholt

**DE GRUYTER**  
MOUTON

ISBN 978-3-11-149634-4  
e-ISBN (PDF) 978-3-11-149643-6  
e-ISBN (EPUB) 978-3-11-149679-5  
DOI <https://doi.org/10.1515/9783111496436>



This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. For details go to <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

**Library of Congress Control Number: 2025931606**

**Bibliographic information published by the Deutsche Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>.

© 2025 the author(s), editing © 2025 Nataliya Levkovich, Julia Nintemann and Maike Vorholt, published by Walter de Gruyter GmbH, Berlin/Boston, Genthiner Straße 13, 10785 Berlin  
The book is published open access at [www.degruyter.com](http://www.degruyter.com).

Cover image: vichie81/iStock/Getty Images Plus  
Typesetting: Integra Software Services Pvt. Ltd.  
Printing and binding: CPI books GmbH, Leck

[www.degruyter.com](http://www.degruyter.com)  
Questions about General Product Safety Regulation:  
[productsafety@degruyterbrill.com](mailto:productsafety@degruyterbrill.com)

---

“Le langage est source de malentendus.”

Antoine de Saint-Exupéry, *Le Petit Prince*

“C’est le temps que tu as perdu pour ta rose qui fait ta rose si importante.”

Antoine de Saint-Exupéry, *Le Petit Prince*

“You can’t always get what you want, but if you try sometimes, you’ll find you get what you need.”

The Rolling Stones, *You Can’t Always Get What You Want*



## Preface

It was clear to us that we – as (former) students and current colleagues of Thomas Stolz – wanted to honour this outstanding linguist with an edited volume to reflect his passion for writing and reading linguistic publications. However, it was less clear to us which topic would do justice to his broad field of expertise and interests. We aim at reflecting his wide range of interests in the title of this volume, while at the same time providing a thematic focus. We also wanted the volume to go beyond a mere collection so that the featured papers would contribute to current research in structural linguistics. Publishing this volume with De Gruyter highlights Thomas Stolz's long-standing relationship with this publishing house, both as an editor and an author of numerous works. We are indebted to Barbara Karlson and Birgit Sievert at De Gruyter Mouton who made this whole project possible in the first place.

Thomas Stolz's interest in linguistics was already awakened in his teenage years when he stumbled upon Benjamin Lee Whorf's "Language, Thought and Reality" in a box of books that belonged to his sister. After this fateful discovery, his career path was set. In 1977, he began his studies in linguistics at the Ruhr University Bochum and graduated in 1985 with a PhD in General and Comparative Linguistics. In his PhD dissertation (prized by the Ruhr University Bochum as the year's best), he explored issues of the grammar of Creole languages. His research interests broadened rapidly, and in his early publications, he covered not only issues of creolistics but also topics such as verbal and nominal morphology in individual languages, e.g. Icelandic, Latvian, Lithuanian, Welsh, and Romanian. In this early period of his linguistic career, Thomas Stolz focused on different aspects of grammaticalization as well as Markedness and Naturalness theories (*Markiertheits- und Natürlichkeitstheorien*). This is where his interest in linguistic typology and language contact arose. Thomas Stolz finished his habilitation on secondary inflection formation (in the context of morphologization) in 1991. In 1995, his outstanding research work was honoured with the Karl Arnold Prize of the North Rhine-Westphalian Academy of Sciences. His outstanding academic achievements enabled him to become the awardee of the Heisenberg program of the *Deutsche Forschungsgemeinschaft (DFG)* for the period of 1991–1995. This was when his work on two large typologically oriented research projects started, which resulted in numerous talks and publications – *Hispanization* (and *Romancization* in a wider sense) and *Typology of Comitatives*.

In 1998, Thomas Stolz became a full professor of General and Comparative Linguistics at the University of Bremen, after serving two years as an interim professor at the university. During his 27 years as Chair of Linguistics, he was in charge of the remarkable transformation of this discipline in Bremen, particularly in the fields

of typology and language contact. His teaching and mentoring activities supported and promoted young academics; in 2010, he established the working group and young scholars' network *Language Contact and Comparison (Sprachkontakt und Sprachvergleich)* at the University of Bremen. Thomas Stolz supervised ten PhD dissertations that were completed successfully, and he is currently supervising four dissertations that are still in preparation.

Besides the large typological projects mentioned above, Thomas Stolz conducted (and continues to conduct) other large research projects concerning – among others – various aspects of areal phonology, possession, reduplication, the areal typology of comparatives, gender assignment in loan words, function-word borrowing, the typology of spatial interrogatives, and special onymic grammar. The latter research project along with his research on the (absence of) place names in Northern Mariana's *gāni*-islands unites two disciplines which did not have many points of intersection until then – structural linguistics and onomastics. Overall, Thomas Stolz was able to obtain funding for ten of his research projects from the *Deutsche Forschungsgemeinschaft (DFG)*. In collaboration with Ingo H. Warnke, Thomas Stolz also established a new paradigm in linguistic research – *Colonial and Postcolonial Linguistics*; his special interest in this area lies in *Colonial Toponomastics*, which resulted in numerous talks and publications.

Until the beginning of 2025, Thomas Stolz authored 376 publications (including 14 monographs) and has given 315 talks on different occasions, most of which were international conferences and workshops. His contributions on various aspects of language contact, areal linguistics, and typology are pivotal, ground-breaking, and frequently cited as foundational.

Being a typologist, Thomas Stolz investigated hundreds of different languages from a typological perspective. However, he also authored various contributions on structural issues in individual languages such as Afrikaans, Classic Aztec, Faroese, Icelandic, Latvian, Lithuanian, and Romanian. He has a particular interest in Welsh (Celtic), Chamorro (Austronesian), and Maltese (Semitic), to which he has dedicated many publications and is considered one of the world's leading experts, especially on the latter two languages. Thomas Stolz is active in promoting Maltese linguistics and linguistic studies on the Maltese language through his numerous publications, talks, collaborations, teaching activities, and a funded research project. He initiated a cooperation between the universities of Bremen and Malta. In 2007, he founded the *International Association of Maltese Linguistics / L-Għaqda Internazzjonali tal-Lingwistika Maltija (GHILM)*. He has been the president of the association up to the present day. Since its launch in 2012, he is also the director of the Malta Centre at the University of Bremen – the only institution devoted to Maltese linguistics outside of Malta.

Thomas Stolz is the editor-in-chief of the leading A-rated typological journal *STUF – Language Typology and Universals* and its supplementary series *Studia Typologica*. He is a co-editor of several linguistic book series, one of them being *Koloniale und Postkoloniale Linguistik / Colonial and Postcolonial Linguistics*. During his career at the University of Bremen, he organised dozens of national and international linguistic conferences and workshops including the Annual Meeting of the *Societas Linguistica Europaea (SLE)* and the Annual Meeting of the *Deutsche Gesellschaft für Sprachwissenschaft (DGfS)*.

Thomas Stolz is one of the world's leading typologists and experts on language contact. His contributions have been, and continue to be, significant for the advancement of linguistics as a discipline, particularly in the aforementioned fields. His publications were inspiring and thought-provoking for many of his colleagues and students, some of whom took the opportunity to honour Thomas Stolz by taking part in this endeavour to acknowledge his achievements in linguistics.

The thirteen contributions of this volume give new insights into linguistic questions with a connection to the very broad research interests of Thomas Stolz. The authors selected their topics with the aim of reflecting the broad repertoire of his work, exploring various structural aspects of language contact and/or typology. In the following, we briefly summarise each paper.

**Marianne Mithun's** contribution is directly connected to one of the most important fields of Thomas Stolz's expertise – language contact. In this paper, the power of contact in shaping language structure is demonstrated through the example of tense categories in the languages of Northern California. Mithun shows that mechanisms of mutual influence of languages can be explored even in cases of scarce or non-existent historical documentation.

Hispanization as a contact phenomenon between Spanish and indigenous languages world-wide is one of the most prominent aspects of Thomas Stolz's studies in the field of language contact. **Eeva Sippola's** contribution on the Spanish-based Creole Ternate Chabacano takes up the topic of Hispanization processes in the Philippines. The article examines causative constructions and highlights valency-changing operations. Influences from Tagalog, such as voice and argument systems, also shape Chabacano causative expressions.

In his contribution on Turkish as a “heritage language” (i.e. non-dominant language) in contact with the socially dominant languages German and English, **Christoph Schroeder** scrutinises contact-induced clause combining phenomena – an aspect of language contact to which Thomas Stolz also dedicated parts of his research. Schroeder also critically evaluates terminological issues in heritage language research.

The case study by **Livio Gaeta** is dedicated to the Titsch variety of Walser German – a Germanic enclave language spoken in northern Italy. The focus of his

study lies on verbal morphology, more precisely on strong and weak verb classes and their remodelling. The connections to Thomas Stolz's research can be drawn not only at the level of language contact in general but also to Stolz's interest in contact-induced aspects in another Germanic enclave variety of Italy – Cimbrian.

The impact of language contact on the Austronesian language Nalik, spoken in New Ireland (Papua New Guinea), is the subject of **Craig Alan Volker's** article. The universal bilingualism of Nalik speakers with the socially dominant Tok Pisin has led to the convergence of these two languages in phonology, morphology, syntax, lexicon, and pragmatics. Volker discusses the impact these structural contact-induced changes in Nalik can have on the speakers' identity.

The potential impact of language contact in the development of multi-verb constructions is one of the aspects that **Tom Güldemann** examines in his study on juncture-verb constructions in the now extinct click language Kwadi. By looking into the verbal morphology, Güldemann provides further evidence for the genealogical relation of this poorly documented language of southwestern Angola (once considered as an isolate) to the southern African Khoe family.

As the short summaries suggest, issues of the impact of language contact on language structures are outlined in the first half of this volume. The subsequent papers are primarily concerned with typological issues – the other central area of Thomas Stolz's research.

**Julia Nintemann, Maike Vorholt, Deborah Arbes, and Kevin Behrens** – coincidentally the authors of the four pending dissertations supervised by Thomas Stolz – dedicate their contribution to a recurrent topic in Thomas Stolz's career: spatial relations in typological perspective. In their paper, configurational expressions in Place, Goal, and Source constructions are explored, and a hierarchy based on the morphological complexity of these constructions is established.

The typology of aspectuality is the subject of **Johanna Mattissen's** contribution. Mattissen investigates which features, beyond aspect marking, are interacting with aspectuality in a variety of languages with aspect and aspectoid systems. In the study, Mattissen explores different types of marking of aspectual values in aspectually head-marking languages in comparison to those language that have aspectually double- and dependent-marking.

**Silvia Luraghi** examines the use of space prepositions for the encoding of the semantic role of agent with passive verbs in various ancient and modern European languages. Luraghi demonstrates that Metaphors – such as Origin and Source – extend spatial prepositions to agents by mapping spatial relations onto agency. These topics are related to Thomas Stolz's studies on comitatives (in the sense of Companion Metaphor) and spatial relations.

**Paolo Ramat's** contribution deals with another prominent topic of Thomas Stolz's research – reduplication, repetition, and iteration in typological perspective.

The focus of Ramat's paper lies on tautological constructions of the type *Whatever will be, will be* in contrast to other types of reduplication or iteration, generalized constructions, and stereotyped idiomatic constructions. Ramat compares these types of constructions in a range of languages.

Questioning the data point of the Tungusic language Uilta (aka Orok) in the WALS chapter on Adjective-Noun constituent order, **Bernard Comrie**'s contribution offers not only a correction of the claim that Uilta has dominant Noun-Adjective order but also explores a methodology for investigating such claims. Comrie comes to a conclusion that will surely be endorsed by many typologists, including Thomas Stolz: "The moral is clear: The price of typology is eternal vigilance."

One of the challenges faced by typologists is the lack of uniform terminology in linguistic works. For typological research, it is essential to develop terms and definitions that are both suitable for typological purposes and inspired by typological studies. In his paper, **Martin Haspelmath** proposes a definition of "root" which is based on the root-types in different languages: it is a contentful morph that can occur as part of a free form without another contentful morph.

In a volume dedicated to Thomas Stolz, a contribution on Maltese is a must, as he dedicated a lot of his attention in the recent two decades of his linguistic career to this language. In **Michael Spagnol**'s contribution, Maltese stative verbs are thus taken under scrutiny and classified into four distinct categories. Spagnol also addresses questions of aspectology and the role of Maltese auxiliary verbs.

We are grateful to all contributors for their willingness to honour Thomas Stolz and for providing insights into current research topics in this volume.

Nataliya Levkovych, Julia Nintemann, and Maike Vorholt  
Bremen, January 2025



# Contents

## Preface — VII

Marianne Mithun

**The joys and challenges of contact effects without substance — 1**

Eeva Sippola

**Causatives and contacts in Chabacano — 21**

Christoph Schroeder

**What is heritage language research good for? A critical discussion, with examples from Turkish language contacts — 41**

Livio Gaeta

**Remodeling inflectional classes: Strong and weak verbs in Walser German — 63**

Craig Alan Volker

**The convergence of Nalik with Tok Pisin: Two languages becoming one linguistic repertoire — 93**

Tom Güldemann

**A Kwadi perspective on Khoe juncture-verb constructions — 117**

Julia Nintemann, Maike Vorholt, Deborah Arbes & Kevin Behrens

**OUT OF OUR MINDS AND ONTO THE PAPER: On the morphological complexity of configurational spatial relations — 147**

Johanna Mattissen

**Aspectual head/dependent marking in relation to verb/satellite framing — 199**

Silvia Luraghi

**The origin of agent markers – Metaphors underlying the extension of spatial prepositions in European languages — 243**

Paolo Ramat

**What(ever) will be, will be — 265**

Bernard Comrie

**Uilta (Orok) attributive adjective order: Grammar, text, typology — 275**

Martin Haspelmath

**Roots and root classes in comparative grammar — 307**

Michael Spagnol

***Taf x'naf?* A closer look at stative verbs in Maltese — 325**

**Tabula gratulatoria — 349**

**Language index — 351**

**Subject index — 355**

Marianne Mithun

# The joys and challenges of contact effects without substance

**Abstract:** A multitude of factors can shape the effects of language contact. The more we know about circumstances surrounding contact situations the better we can identify mechanisms by which features of one language can influence another. Where substance has been replicated, and there is detailed documentation of social and cultural characteristics over a long span of time, along with deep philological records, the mechanisms may be easier to discern. But for many languages such knowledge is nonexistent. Here some mechanisms are explored by which structure was apparently transferred without substance in the development of tense categories in languages of Northern California.

**Keywords:** category transfer, contact, linguistic areas, subjectivity, tense

## 1 Introduction

As shown richly in work by Thomas Stolz and others, the power of contact in shaping language has become ever clearer as we have learned more about more languages (Stolz 2002, 2005, 2006, 2008, 2010, 2012, 2015, and more). The majority of the world's citizens are multilingual. They may command their languages equally or not, they may use them in different contexts or not, they may accord them different prestige or not, and they may see them as markers of identity or not. The multilingualism may be relatively uniform across all members of a group, or different languages may be primary in different parts of it. The language under investigation may have been the first language of its speakers over a long period of time, or the result of recent language shift. In some cultures, code-switching, language mixing, and borrowing are the norm or even admired, while in others, efforts are made to keep the languages apart. In all cases, the presence of multiple languages within a single brain can affect each language to at least some degree.

Identifying contact effects on a language can be facilitated if the history, circumstances, and nature of the contact are known. Identifying transferred matter, particularly vocabulary, can be more straightforward than identifying affected

---

**Marianne Mithun**, University of California, Santa Barbara, e-mail: [mithun@linguistics.ucsb.edu](mailto:mithun@linguistics.ucsb.edu)

structure. But contact effects can sometimes be discerned even without such advantages. Here such a situation is illustrated in languages indigenous to the Northern California area of North America.

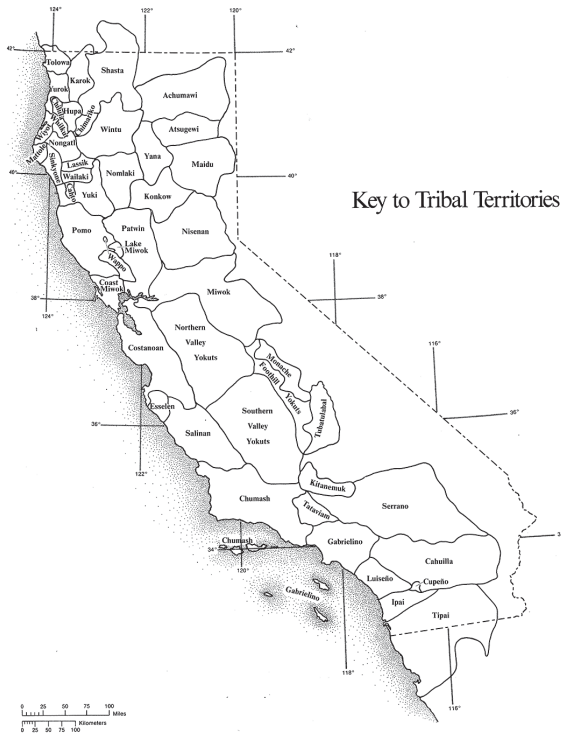
## 2 Indigenous California

What is now California is a well-known language area, but it shows considerable genealogical diversity: it is home to around twenty distinct indigenous genetic groups. Some of these are families that also include languages spoken outside the area as well, many are families with just few members, and some consist of just a single language isolate. Northern California is the center of a particularly strong cultural and linguistic area. It is not delimited by modern state boundaries: areal relations extend in all directions, and within it there are a number of sub-areas. The locations of the languages can be seen in Figure 1.

Archaeological evidence suggests that Yukians (Yuki and Wappo) entered the area around 9500–9000 B.C., then Pomoans and the Karuk, Chimariko, Shasta, Yana, Atsugewi, Achumawi, and Washo several thousand years later, followed by Wintun, Maidun, and Klamath-Modoc peoples several thousand years after that (Golla 2011). First contacts with Europeans were not until the late eighteenth century for some, and the mid or late nineteenth century for others. Among the earliest significant ethnographic descriptions are those in Powers (1877), followed by more detailed study at the beginning of the twentieth century under Kroeber, including his major 1925 *Handbook of the Indians of California*. There is thus little detailed information about social relations and interactions within and among groups over the millennia during which the languages took their modern shapes, crucial for understanding the nature of language contact in depth.

So far as can be known, most of the communities have been relatively small throughout their histories. Describing the social and political organization of California groups, Heizer (1978: 5) mentions the term *tribelet*.

The word was coined by Kroeber to indicate the basic, autonomous, self-governing, and independent sociopolitical group found all over the state. The term *village community* has also been used in the same sense. The tribelet consisted of the aggregation of people living in two or more (often up to a dozen) separate villages, acknowledging the leadership of a chief who usually resided in the largest and most important of the several settlements. The data on number and nature of the tribelets of some larger tribes (that is, linguistic units) is known with fair completeness. The Pomo, for example, were divided into 34 tribelets living on 3,370 square miles of land and numbering altogether about 8,000 persons (Steward 1943). The Achumawi were divided into 11 tribelets and their total numbers are calculated at 3,000 persons, their territory comprising about 6,000 square miles of plateau land (Kniffen 1928).



**Figure 1:** Indigenous Languages of California. Heizer (1978: ix).

The small sizes of communities resulted in traditions of exogamy, patrilocality, and multilingualism (Harold et al. 2016). But there are also traditions of place-based language use: one speaks the language where one is, without mixing. There has accordingly been relatively little lexical borrowing, apart from the incorporation of terms from Spanish for items brought in with the recent arrival of missionaries. Yet a striking number of structural features are shared across the area among unrelated languages.

In the absence of deep philological records, we can only posit hypotheses about the mechanisms by which parallel patterns developed, but the results are suggestive. Speakers accustomed to specifying certain distinctions in one of their languages may, often unconsciously, replicate the frequency of those distinctions in another language, using material from that language. Over a certain period of time, high-frequency distinctions can become routinized and even ultimately result in parallel morphological structures between languages in contact, though the markers of these distinctions will be different in form in the two languages. Effects of this process are illustrated here with tense systems.

## 3 Graded past tense

A number of languages indigenous to Northern California distinguish multiple past tenses.

### 3.1 Shasta

Shasta is a small set of very closely related languages or dialects once spoken in the northernmost part of California and into Oregon. The last speaker of any Shastan variety, Clara Wicks, died in 1978. Apart from wordlists, there is little documentation of any except Shasta proper, described by Silver (1966). Shasta declarative verbs are inflected with prefixes that mark mode, subject person and number, and tense. Silver distinguishes present, near past, and distant past tenses. She characterizes them as follows.

- The **present tense** describes an action or state of being at the time of speech. It is also used when talking about immediately past actions or states if the context for them is still in force at the time of speech. (Silver 1966: 125)
- The **near past** is used in conversation and anecdote to describe events relatively recent in time from the speaker's point of view; for example, in an anecdote involving himself, SS [speaker Sargeant Sambo] might use the near past to refer to an event that happened any time from the day before to fifty or sixty years ago. (Silver 1966: 127)
- The **distant past** prefixes are usitative or simply markers of distant past time . . . <p'> is distant past inferential. <t'w>, which has a variant <tw>, is distant past reportative. (Silver 1966: 129)

The past tense markers also distinguish evidentiality.

- <k'w> is near past inferential. <t> 'reportative' is used when describing events of which the speaker has no personal knowledge; however, there is a tendency to substitute <t> for the other forms in the near past and use it simply as a past time marker. (Silver 1966: 128)
- <t'w> and <t> are used in the narration of myths. They are also used in conversation and anecdote, along with the other near past and distant past forms. (Silver 1966: 129)

An important variable in such tense distinctions is subjectivity. Silver notes that speaker Sergeant Sambo used the near past for events that happened any time from

the day before to 50–60 years before. He was already 105 years old when he began working with her in 1957. It is also significant that the distant past markers are fused with inferential and reportative evidentials.

### 3.2 Chimariko

Immediately to the south of the Shasta is Chimariko. Chimariko is an isolate, last spoken in 1950. Fieldnotes from various researchers have been assembled in a grammar by Jany (2009). Like many other languages of the area it shows a rich tense/aspect system, though, as noted by Jany, it is not always easy to distinguish the two kinds of distinctions, due to the nature of the data. Among the verb suffixes are those in (1). Jany notes (2009: 104) that both *-neq* and *-nip* encode events that happened in the past, and both are often accompanied by a temporal adverb, *šur* ‘formerly’, *šuraku* ‘already’, or *moʔa* ‘yesterday’. Of significance here is the distinction between basic and remote (‘ancient’) pasts.

- (1) Chimariko past tenses (Jany 2009: 103–104)
- ak/-k* Past (completed action)
  - neq* Past (formerly)
  - nip* Past (‘already’), complete
  - taʔ* Ancient Past, perfective (Remote)

Chimariko thus distinguishes basic past and remote past tenses, but the forms do not resemble those of its immediate neighbor Shasta.

### 3.3 Achumawi and Atsugewi

Directly to the east of Shasta is Achumawi. Achumawi was earlier linked with its southern neighbor Atsugewi in a family termed Palaihnihan, though whether they constitute a family of their own remains under discussion (Nevin 2019). Achumawi was spoken in small communities along the Pit River. Nevin (p.c.) notes that verbs without overt tense markers are atemporal, used for both present and immediate past situations. A Recent Past is formed with a suffix *-íní*, and a Remote Past, used in myths, is formed with a prefix *ck-*. It also contains a Pluperfect suffix *-oʔoy*, which indicates a time prior to the main time under discussion.

- (2) Achumawi Pluperfect (Bruce Nevin, p.c.)
- a. *Qá cuucé tikusqáato'oy púk ckwammúci.*  
'The comb that she **had** set up plop fell down.'
  - b. *Q<sup>h</sup>ééwáté ís tiikááco'oyí twiyí qá itt<sup>h</sup>ú íssílóo.*  
'My ancestors **used to** live there.'

In Atsugewi, directly to the south of Achumawi, the basic finite verb is used for both present and immediate past. There is also a Past marker *-n-* or *-i?n-*, (de Angulo 1930: 96, Leonard Talmy, p.c.), which shows up in folk tales, perhaps cognate with the Achumawi Remote Past.

Both Achumawi and Atsugewi thus distinguish Recent and Remote Past tenses, though the forms of the markers resemble those of neither Shasta nor Chimariko.

### 3.4 Yana

Immediately to the south of Achumawi and the west of Atsugewi is the isolate Yana, with three dialects. So far as is known, the language was last spoken nearly a century ago. Like Shasta, Chimariko, Achumawi, and Atsugewi, Yana distinguishes basic and remote past tenses. The Yana markers are also associated with evidentiality as in the unrelated but neighboring Shasta. Among the Northern Yana verb suffixes Sapir lists those in (3).

- (3) Northern Yana past tense suffixes (Sapir 1922: 230, 232)
- ha* 'past, implies personal knowledge on part of speaker'
  - ?ni, -n?* 'remote past' (generally prefixed to other tense-modal elements)
  - n?ha-* 'quite some time ago, on personal knowledge of speaker'
  - n?ni-wara* 'remote non-quotative past'
  - n?ti* 'long ago, as they say' (used in myths)
  - n?t'i-wara* 'remote quotative past'

(The element *-wara* is apparently a Perfective.)

### 3.5 The periphery: Karuk and Yurok

Languages spoken to the west of Shasta show past tense distinctions, but these are somewhat less grammaticalized, expressed as enclitics and particles.

Immediately to the west of Shasta is the isolate Karuk. Karuk contains just one past tense suffix *-at*, but additional markers have been added to it to form enclitics for further distinctions.

(4) Karuk pasts (Bright 1957: 125)

*-at* Past

*=ahe:n* Anterior tense (Immediate)

designates time previous to that indicated by *-at* Past Tense. In narratives where tenseless forms predominate, however, anterior forms, like past tense forms, may be used in reporting an event immediately past. In either case, *-ahe:n* may be translated by the English pluperfect construction.

*=anik* Ancient tense (Remote)

designates a time more remote than that indicated by the past or anterior morpheme. Forms with this postfix [enclitic] are frequently used to begin stories, which are then continued with tenseless forms.

Distinctions of time are also indicated by two particles.

*?ip(a)* ‘near past’ is used principally in conjunction with the past tense suffix *-at*, the meaning of which it duplicates in part. It contrasts, however, with *mit(a)* ‘remote past’ which is also used with past tense verb forms. (Bright 1957: 126)

Yurok territory is further away from Shasta, to the west of Karuk. Robins (1958: 32) reports: “There are no formally differentiated tenses in the Yurok verb. Verb forms of any of the paradigms, except the imperative, may be interpreted as referring to past, present, or future time, according to the linguistic and non-linguistic contexts in which they occur.” There is, however, a substantial set of preverbal particles, many of which Robins translates with tense and aspect-like meanings. In addition to several simply glossed as ‘past time’ is *?ockic* ‘time just past’, presumably an immediate or recent past.

### 3.6 Elaboration: Washo

Washo territory is further to the south and east, straddling the modern boundary between California and Nevada, but it is considered part of the California area linguistically. Genealogically it is an isolate, not demonstrably related to any other language. In his 1964 grammar, Jacobsen describes a full set of past tense categories.

## (5) Washo pasts

*-leg* Recent Past

used to place the time of an event at an earlier point on the same day, or during the preceding night. (Jacobsen 1964: 631)

The adverbial expressions of time that have been found associated with verbs bearing this suffix include *watlí:ʔiŋ* ‘early in the morning’, *téšciw watlí:* ‘just this morning’, and *lélim* ‘at night, last night’. (Jacobsen 1964: 632)

*-ay?* Intermediate Past

indicates a time in the past earlier than the same day but not in the extremely distant past. (Jacobsen 1964: 636)

*-áyti?* Pluperfect

places the time of an event prior to that of another event in the event prior to that of another event in the past. (Jacobsen 1964: 636)

*-gul* Remembered Past

places the time of an event in the distant past but still within the lifetime of the speaker. (Jacobsen 1964: 636)

*-lul* Distant Past

places the time of an event in the distant past, before the lifetime of the speaker. (Jacobsen 1964: 636)

*-elem* Distant Tense

attested only in material from RJ and HP [speakers Roy James and Bertha Holbrook], occurs always followed by one of the three prefinal suffixes expressing past or future tenses: *-ay?* Intermediate Past, *-lul* Distant Past, or *-gab* Distant Future. It has the meaning of making the time of the action more distant, either in the past or in the future, than it would otherwise be. (Jacobsen 1964: 637)

### 3.7 Maidun

To the west is Maidun (or Maiduan) territory. The Maidun family consists of four main languages: Nisenan, Konkow (= Northwestern Maidu), Chico, and Maidu (= Northeastern Maidu = Mountain Maidu).

Nisenan is directly to the west of Washo. Anderson (2024) discusses the past tense suffixes of Southern Hill Nisenan and their uses documented by Uldall in unpublished fieldnotes (1950) and published oral narratives (Uldall and Shipley 1966). There are three past tense suffixes.

## (6) Nisenan pasts (Anderson 2024)

*-'as(i)* Immediate/Recent Past

Most “recent”: Seconds prior to the reference time

Most “distant”: During the preceding night

*-'a* Intermediate Past

Most “recent”: 1–2 days prior to the reference time

Most “distant”: 6 months prior to the reference time

*-t'o* Distant Past

Most “recent”: 1–2 years prior to the reference time

Most “distant”: Creation Time

Anderson notes that *-t'o* Distant Past does not occur in sentence elicitation, because of the nature of the task, which consisted of asking speakers to translate English past tense sentences without additional context.

To the northeast of Nisenan is the Maidun language Konkow. Ultan (1967) distinguishes several past tense markers, but notes that identifying their precise meanings is challenging.

## (7) Konkow pasts

*-ma* Past

In most examples, past tense seems to fit the situations but there are apparent exceptions, some of which are certainly due to deficiencies in translation, others perhaps not so. Furthermore, on the basis of the present analysis: 1) there are a few constructions which must represent tautologies, i.e. *ma* occurs twice or with *wono* ‘remote past’ in the same string, 2) *ma* must be accorded triple privilege of occurrence as a thematic suffix, a participle, and as an inflectional suffix. (Ultan 1967: 99)

*-in* Recent Past

Its primary function is to identify a non-finite verb, yet where it is necessary to make explicit a distinction in tense (between past and non-past), between two parallel or otherwise identical constructions, *-in* serves as a contrastive non-past (or recent) to *ma*. Aside from such situations, *-in* can and does occur with themes in any tense: past, remote past, future, and unmarked or recent. (Ultan 1967: 119)

*-wono* Remote Past

refers to situations which occurred at some time in the past which is felt to be remote by the speaker. This may have been in an absolute sense such as for an event which transpired before the speaker was born, or in a relative sense associated with an occurrence which is beyond the speaker’s range of memory. (Ultan 1967: 96–97)

Konkow is also directly to the south of Yana.

To the north of Nisenan and Konkow is Maidu. Shipley (1964) did not find the past distinctions seen in Nisenan and Konkow within the Maidu verb.

The present-past is used:

1. To describe a recently completed punctual action, but only if the context in which that action occurred persists at the time of speech. It is never used to describe events from the previous day; the passing of a night is conceived as an inevitable change of the context.
2. To describe some punctual action which is taking place at the time of speech.
3. To express a state of being at the time of speech.
4. To express the idea of present static location (Shipley 1964: 46–47).

There is, however, a periphrastic construction for remote pasts. *kyʔy'm* + the past punctual of *ka . . .* 'be' is used for remote past punctual time within the speaker's lifetime, for example,

(8) Maidu (Shipley 1964: 53–54)

a. *Tibím kylém p'y'm maʔá.*

'Long ago I was a little girl.'

b. *Nèkbék'ym béj, ník ʔejáj kyʔy'm maʔám.*

Thus did my father speak to me **that one time long ago.**'

No past tenses can thus be reconstructed for Proto-Maiduan. The fact that the Maidu remote past construction is a full word suggests that the distinction entered the language later than it did in Nisenan, spoken in an area adjacent to that of Washo, and in Konkow, directly to the north of Nisenan.

### 3.8 Wintun

West of the Maidun languages are those of the Wintun (Wintuan) family. Wintun consists of two major divisions, Northern Wintun (Wintu and Nomlaki), and Southern Wintun (Patwin and Southern Patwin). In her 2006 reconstruction of Proto-Wintun, Shepherd found no past tense distinctions. In his 1984 grammar of Wintu proper, Pitkin lists no past tense distinctions. But in his 2015 grammar of Patwin, Lawyer reports that Patwin distinguishes two past tense suffixes.

## (9) Patwin pasts

*-sa* Past Declarative

is attested in all Patwin dialects, and it signifies the past tense. . . . It is in semantic contrast with the remote past tense. (Lawyer 2015: 225)

*-n(i)* Remote Past

The prefinal suffix *-ni* Remote Past marks an event occurring in the remote past. It selects the hortative stem of the verb to which it attaches. In Hill Patwin, *-ni* is always followed by one additional inflectional suffix – either *-sa* PAST or *-ta-*, *-te-* Q.PAST [interrogative past]. (Lawyer 2015: 231)

Lawyer (2015: 232) further comments on variation in use of the distinction.

The temporal boundary between events marked with past *-sa* and events marked with remote past *-ni* varies by speaker, and maybe by dialect. Hill Patwin speaker Nelson Lowell uses *tilansa* (hila-n-sa) ‘shot (RPAST)’ for an event happening one year earlier; and *tila:sa* (hila-sa) ‘shot (PAST)’ for an event happening one day, one week, or one month earlier (Ultan notebook 6.41). Hill Patwin speaker Daisy Lorenzo describes a similar division, using the remote past for events occurring a ‘year, 2 years, or 2–3 months’ in the past (Ultan notebook 6.46). On the other hand, River Patwin speaker Sarah Gonzales uses *winni* (win-ni) ‘saw (RPAST)’ for an event happening w or 3 days earlier; and *wínsa* (wini-sa) ‘saw (PAST)’ for an event happening one day earlier (Ultan notebook 7.38).

He further notes differences across the languages.

A cognate of *-ni* ‘RPAST’ is attested in South Patwin, but is a general past tense marker rather than a remote past tense marker. . . . Shepherd (2005) does not mention the suffix *-ni*, and no obvious Wintu cognate presents itself. (Lawyer 2015: 232)

Southern Patwin has its own distinctions, however.

The suffix *-t* apparently marks the remote past in Southern Patwin, as in *Pisú-t* (pisu-t) ‘created’ and *Téhomat* (te\*homa-t) ‘gave birth’, in the contexts of the Christian creation and the birth of Jesus, respectively. . . .

Another Southern Patwin form denoting remote past tense is *-net*, probably a combination of the past tense suffixes *-ni* and *-t*. This suffix occurs three times in Vallejo’s translation of the Apostle’s Creed, referring to activities carried out in the remote past. (Lawyer 2015: 233)

Patwin and Southern Patwin territories, as noted, directly border Nisenan territory.

### 3.9 Utian

Directly to the south of Nisenan and Southern Patwin are the Sierra Miwok languages, members of the Miwok-Costanoan or Utian family. The northernmost of these, Northern Sierra Miwok, also contains multiple past tense verb suffixes.

(10) Northern Sierra Miwok: (Callaghan 1987)

<i>-ke-</i> ~ <i>-k:e-</i>	Past
<i>-ma-</i>	Immediate Past
<i>-ka-</i>	Recent Past
<i>-se-s:y-</i>	Remote Past

The Remote Past consists of two elements. The first, *-se-*, is a nominalizer. Callaghan notes (1987: 196) that forms with *-se-s:y-* appear with the possessive pronominal series. This suggests that the Remote Past is a later formation than the other pasts.

To the south of Northern Sierra Miwok is Central Sierra Miwok, described by Freeland (1951). The Present tense is based on the verb stem with no further marking, and Perfects are based on the verb stem with a suffix *-(n)ak*. Freeland (1951: 62) reports that “[i]n ordinary usage, both the present and the perfect apply to time of the present day, the present referring to incompleting action, the perfect to completed action”. But other tenses are based on nominal stems, usually agentive nominalizations.

(11) Central Sierra Miwok (Freeland 1951 60–61, 72; Broadbent 1964)

<i>-e-</i> , <i>-š:e-</i> , <i>-k'e-</i>	Recent Past
<i>-š:ĩ</i> :-	Distant Past (East Central, Southern Sierra Miwok)

Freeland notes that the three Recent Past forms alternate according to the preceding phonological context. She characterizes their meanings as follows.

When it is used predicatively, the recent past tense refers to events before the present day, from the day preceding to as far removed as several months or a year . . .

When it is used in subordinate constructions, the recent past tense does not necessarily refer to action of the previous day, or more remote, but is used for any action anterior in time to that of the main verb. This is an extremely important use of this tense, and most of the examples of the recent past in the text (since so little of the text deals with events in recent past time) are subordinate forms. (Freeland 1951: 66)

The Distant Past marker is a nominal suffix ‘past, erstwhile’: *hayá:po* ‘chief’, *hayá:po-š:u* ‘ex-chief’; *šiyéŋ-tu-* ‘comer from seeing; *šiyéŋ-tu-šu-* ‘past comer from seeing’ (Freeland 1951: 72).

To the west of Northern Sierra Miwok is Plains Miwok. For this language, Callaghan (1984) lists just a two-way distinction. (The symbol H indicates ‘length when canonically permissible’). The Plains Recent Past suffix matches the Immediate Past of Northern Sierra Miwok.

(12) Plains Miwok (Callaghan 1984)

*-Htu* ~ *-HtY-* Past  
*-ma-* Recent Past

Callaghan (1984: 142) does not list a Plains counterpart to the Northern Sierra Miwok Recent Past *-ka-*, though she does list a noun suffix *-s:y-* which she tentatively glosses ‘former?’.

Still further to the west, Lake Miwok contains no tense affixes. Temporal particles usually occur at the beginning of clauses, immediately after introductory particles, or at the end of clauses, before any clause markers. Callaghan lists a number of particles with tense and aspect meanings.

(13) Lake Miwok (Callaghan 1963: 247)

*há(a)li (ko)* ‘still, yet’  
*hójot, hojot* ‘quite a while ago’  
*húke, huke* ‘a while ago’  
*jóllejole* ‘often’  
*káşa* ‘still’  
*keláckelac* ‘long, long go’  
*máahuke* ‘a while ago’  
*máate* ‘soon, until’  
*maháli* ‘still’  
*née* ‘now’  
*níh* ‘now’  
*niháli ko* ‘still’

Thus no past tense suffixes can be reconstructed for Proto-Miwok-Costanoan, Proto-Miwok, or even Proto-Sierra-Miwok. The languages closest to Nisenan, that is, Plains, Northern, Central, and Southern Miwok have developed multiple past tense suffixes, in some cases with still clear etymologies, while the more distant Lake Miwok has not.

### 3.10 Yokuts

To the south and east of the Sierra Miwok are the Yokuts languages. Graded past tenses have been identified in just one of these: Chukchansi. In his 1968 grammar, Collord lists basic past, recent past, and remote past suffixes.

(14) Chukchansi past tenses

*-it* Simple Past

indicates that the event has happened very recently or is about to be completed. In a narration it may mark events which are recent in comparison to an action by the more remote past, *-t<sup>h</sup>a?*. . . . The simple past can also be used very generally for any past action that is not simply an event (where *-t<sup>h</sup>a?* or *-hil* would be used). (Collord 1968: 39)

*-hil* Aorist (Recent Past)

marks an event as having occurred recently, but usually not more than a few days ago. It is used, for example, with *lakyiw* ‘yesterday’. An event of the past few moments is indicated by the simple past tense *-it<sup>h</sup>*, and an event of more than a few hours or days past is referred to by the use of the past perfective *-t<sup>h</sup>a?*. These three tenses are, hence, relative to each other. Their use depends on whether the speaker thinks of the event as more remote or more immediate. (Collord 1968: 40)

*-t<sup>h</sup>a?* Past Perfective (Remote Past)

marks action as remote in past time and is the typical tense marker in narratives both formal and informal. (Collord 1968: 40)

Such distinctions do not appear in other Yokuts languages, as reported by Newman (1944: 121–122).

In its semantic pattern the category of tense in Yokuts is remarkably simple. There are only two fundamental tenses: the aorist, referring to a present or past time, and the future . . .

In a strictly semantic sense, the aorist suffix denotes present or past time. But even its wide semantic application does not fully suggest the almost unlimited scope of this suffix in actual usage, for it is employed about as frequently as all the other verbal suffixes of Yokuts combined. Undefined as to voice, aspect, or mode, and covering a wide temporal reference, the aorist serves as a kind of neutral (but non-future) suffix, fulfilling the formal requirement of completing the verb with a final suffix but acting as a relatively indefinite and ambiguous semantic element.

The cognate of the general Yokuts Aorist *-hin* in Chukchansi is the Recent Past *-hil*.

It should be noted that Chukchansi is located in a small area directly adjacent to Sierra Miwok, to its southeast, visible in Figure 1 as the small point of land between Miwok and Monache.

## 4 Implications and challenges

The languages indigenous to California show tremendous genealogical diversity, with around twenty distinct families and isolates. During the 20<sup>th</sup> century, efforts were made to uncover deeper relations among them, resulting in proposals of two major stocks: Hokan, consisting of Karuk, Shastan, Palaihnihan, Yana, Chimariko, Pomoan, Washo, Esselen, Salinan, Yuman, Cochimí, and Seri on the one hand, and Penutian, consisting of Takelma, Wintuan, Klamath-Modoc, Maiduan, Yokuts, Costanoan, and Miwok on the other, in addition to a few other recognized families including Algic, Na-Dene (Athabaskan), Uto-Aztecan, perhaps Yukian (Yuki-Wappo), and Chumash. The Hokan and Penutian proposals are not now generally accepted: perceived similarities were generally structural rather than substantive.

It is striking that groups of unrelated but neighboring languages show parallel elaboration of past tense categories. In the north, neighboring languages Shasta, Chimariko, Achumawi, Atsugewi, Yana, and Karuk all distinguish basic pasts from remote pasts. At the periphery of this area, Karuk has a basic past tense verb suffix, but has added Immediate and Remote categories as enclitics. Still further away, Yurok has no tense morphology, but it does have a rich array of tense/aspect particles. None of the forms of these markers are similar in form, however.

To the east, the isolate Washo contains a rich set of past tense markers, distinguishing Recent, Intermediate, Remembered, and Distant pasts. Its immediate neighbor to the west, the Maiduan language Nisenan, distinguishes Immediate or Recent, Intermediate, and Distant pasts, and its relative Konkow, another Maidun language immediately to the north of Nisenan but not adjacent to Washo, distinguishes a basic Past, Recent Past, and Remote Past. Their relative Maidu, to the north, makes no such distinctions at all. Immediately to the west of Nisenan, the Wintun languages Patwin and Southern Patwin distinguish a basic Past and Remote Past, but the other members of the Wintun family do not. Directly to the south of Nisenan, Northern Sierra Miwok distinguishes a basic Past, Immediate Past, Recent Past, and Remote Past. Both its sister language immediately to the west, Plains Miwok, and its sister to the south, Central Sierra Miwok, distinguish basic and Recent Past, but here the Recent Past constructions are based on nominal stems with a nominal past suffix ‘former’. A related language still further to the west, Lake Miwok, has no past tense morphology at all; time is indicated just with particles. South of the Sierra Miwok languages are languages of the Yokuts family. Just one of these languages, Chukchansi, distinguishes Simple, Recent, and Remote Past. Chukchansi territory is immediately adjacent to the Miwok territory. Here, too, similar distinctions are made throughout the area, but the forms of the markers do not match.

At the peripheries of these areas, the inventories are smaller, and beyond those no distinctions are found. On the basis of the current locations of the communi-

ties it is not entirely clear whether we might be dealing with one general contact area in Northern California, or two, the first perhaps encompassing Shasta, Chimariko, Achumawi, Atsugewi, and Yana, and the second Washo, Nisenan, Konkow, Patwin, Southern Patwin, Northern Sierra Miwok, Central Sierra Plains Miwok, and Chukchansi Yokuts. It should be noted that this second area contains a mixture of languages once hypothesized to be part of a “Hokan” stock (Washo, the Miwok languages) and languages once hypothesized to be part of a “Penutian” stock (the Maidun languages Nisenan and Konkow, the Wintun languages Patwin and Southern Patwin, and Chukchansi Yokuts). Konkow territory actually borders Yana territory to the north and Nisenan territory to the south.

The distribution of the distinctions strongly suggests contact effects, in which semantic distinctions were replicated across languages without substance. A reasonable pathway of development can be hypothesized. Where language mixing is traditionally avoided, bilingual speakers accustomed to making certain distinctions in one of their languages may seek to replicate those distinctions in their other language using material from that other language. They might use full lexical items, such as adverbials or various particles, or they might extend grammatical markers of that language semantically. Over a long period, the heightened frequency of use could result in the grammaticalization of new tense distinctions. The fact that Karuk, at the periphery of the first possible area, has developed enclitics to mark recent and remote pasts, and that Yurok, still further away, uses particles, would be in keeping with such a scenario. The fact that Central Sierra Miwok, further from the center of the second area, has formed a Remote Past marker from nominalized verb forms with a nominal Former Past marker is similarly suggestive.

Supporting this scenario is the fact that no tense distinctions are reconstructed for the ancestors of languages which have developed them. There are parallel developments among neighboring languages rather than genealogically related ones. No tense distinctions are reconstructed for Proto-Maiduan. The Maiduan language Nisenan, spoken in an area adjacent to Washo, shows elaboration of its past distinctions, but its relative Maidu does not. No tense distinctions are reconstructed for Proto-Yokuts. Chukchansi Yokuts, spoken in an area adjacent to Nisenan shows elaboration, but no other Yokuts languages do. No tense distinctions are reconstructed for Proto-Wintun, but the Wintun languages Patwin and Southern Patwin, also spoken in areas adjacent to Nisenan, do. Their Northern Wintun relatives do not. No tense distinctions are reconstructed for Proto-Miwok, but Northern Sierra Miwok, Central Sierra Miwok, and Plains Miwok, also spoken in areas near Nisenan and Patwin, show elaboration, while Lake Miwok, further to the west, does not.

Further suggestive of such a scenario is the fact that languages indigenous to the California linguistic area show additional extensive structural effects without transferred substance. A large proportion of them, for example, contain sets

of means/manner prefixes and locative/directional suffixes (Mithun 2007). The forms of the markers are not the same across the languages, and the inventories of markers vary as well. The prefixes and suffixes in languages more central to the area, presumably the origin of the patterns, are in general shorter, often just a single consonant, indicative of greater age, while those at the periphery are more substantial, and, in some cases, can be related to noun or verb roots that still survive in the languages.

The hypothesis of the development of graded tense systems via contact does face certain challenges. The first is the matter of chance. We do not know for certain how likely languages are to develop additional tense distinctions spontaneously, on their own. Second is the fact that the languages here have been in place for millennia. Details concerning the nature and intensity of contact among their speakers are of course undocumented. Finally, the majority of the languages are no longer spoken. Our understanding of the grammatical categories of each language is necessarily dependent on the existing records of them. In many cases documentation of connected speech is limited to elicited translations of sentences from English or Spanish, or, at best, narrative texts, most often myths and legends. We are fortunate that these do exist. Elicited sentences of course can tend to replicate the categories of the contact language; basic past tenses might predominate, and special recent past or remote past constructions might not appear at all. In myths and legends, remote past forms might predominate. In her work with Nisenan texts for example, Anderson (2024) noted that Distant Past forms dominate the corpus, occurring more than 3700 times, while the Recent and Intermediate Past tenses occur only in direct quotes from characters, with 22 Recent Pasts and 10 Intermediate Pasts.

We may never know for certain just which structures have been shaped by language contact, but as we learn more about more languages, and the situations in which they are and have been used, we should be able to sharpen our understanding of the possibilities.

## References

- Anderson, Skye. 2024. Revitalizing the graded past tense system of Southern Hill Nisenan. Paper read at the meeting of the Society for the Study of Indigenous Languages of the Americas, New York City.
- de Angulo, Jaime. 1930. The Achumawi language, with L.S. Freeland. *International Journal of American Linguistics* 6. 77–120.
- Bright, William. 1957. *The Karok language*. (University of California Publications in Linguistics 13). Berkeley: University of California Press.

- Broadbent, Sylvia M. 1964. *The Southern Sierra Miwok language*. (University of California Publications in Linguistics 38). Berkeley: University of California Press.
- Callaghan, Catherine. 1963. *A grammar of the Lake Miwok language*. Berkeley: University of California Ph.D. dissertation.
- Callaghan, Catherine. 1984. *Plains Miwok dictionary*. (University of California Publications in Linguistics 105). Berkeley: University of California Press.
- Callaghan, Catherine A. 1987. *Northern Sierra Miwok dictionary*. (University of California Publications in Linguistics 110). Berkeley: University of California Press.
- Collord, Thomas L. 1968. *Yokuts grammar: Chukchansi*. Berkeley: University of California Ph.D. dissertation.
- Freeland, L.S. 1951. *Language of the Sierra Miwok*. Indiana University Publications in Anthropology and Linguistics Memoir 6 of the *International Journal of American Linguistics* 17(1).
- Golla, Victor. 2011. *California Indian languages*. Berkeley: University of California Press.
- Harold, Laura, Jelmer Eerkens & Candice Ralston. 2016. Patrilocal post-marital residence and bride service in the Early Period. In Alexandra M. Greenwald & Gregory R. Burns (eds.), *Reconstructing lifeways in ancient California*, 33–44. Davis, CA: Center for Archaeological Research.
- Heizer, Robert F. (ed.). 1978. *Handbook of North American Indians: California*, vol 8. Washington, D.C.: Smithsonian Institution.
- Jacobsen, William H. 1964. *A grammar of the Washo language*. Berkeley: University of California Ph.D. dissertation.
- Jany, Carmen. 2009. *Chimariko grammar: Areal and typological perspectives*. (University of California Publications in Linguistics 142). Berkeley: University of California Press.
- Kniffen, Fred B. 1928. Achomawi geography. *University of California Publications in American Archaeology and Ethnology* 23(5). 297–332.
- Kroeber, Alfred. 1925. *Handbook of the Indians of California*. (Bureau of American Ethnology Bulletin 78). Washington, D.C.: Government Printing Office.
- Lawyer, Lewis. 2015. *A Grammar of Patwin*. Davis: University of California Ph.D. dissertation.
- Mithun, Marianne. Grammar, contact, and time. *Journal of Language Contact*. <https://brill.com/view/journals/jlc/1/1/jlc.1.issue-1.xml>
- Nevin, Bruce. 2019. Why Proto-Palaihnian is neither. [manuscript] DOI: 10.13140/RG.2.2.33917.20966
- Newman, Stanley S. 1944. *Yokuts language of California*. (Viking Fund Publications in Anthropology 2). New York: Viking Fund.
- Pitkin, Harvey. 1984. *Wintu grammar*. University of California Publications in Linguistics 94. Berkeley: University of California Press.
- Powers, Stephen. 1877. *Tribes of California*. Contributions to North American Ethnology 3. Washington, D.C.: U.S. Geographical and Geological Survey of the Rocky Mountain Region.
- Robins, R. H. 1958. *The Yurok language: Grammar, texts, lexicon*. (University of California Publications in Linguistics 15). Berkeley: University of California Press.
- Sapir, Edward. 1922. The fundamental elements of Northern Yana. *University of California Publications in American Archaeology and Ethnology* 13(6). 215–234.
- Shepherd, Alice. 2005. *Proto-Wintun*. Berkeley: University of California Press.
- Shipley, William F. 1964. *Maidu grammar*. (University of California Publications in Linguistics 41). Berkeley: University of California Press.
- Silver, Shirley. 1966. *The Shasta language*. Berkeley: University of California Ph.D. dissertation.
- Stewart, Omer C. 1943. Notes on Pomo ethnogeography. *University of California Publications in American Archaeology and Ethnology* 40(2). 29–62.

- Stolz, Thomas. 2002. General linguistic aspects of Spanish-Indigenous language contacts with special focus on Austronesia. *Bulletin of Hispanic Studies* 79. 133–158.
- Stolz, Thomas. 2005. Italianisierung in den alloglotten Sprachen Italiens. In Ermenegildo Bidese, James R. Dow & Thomas Stolz (eds.), *Das Zimbrische zwischen Germanisch und Romanisch*, 43–68. Bochum: Universitätsverlag Brockmeyer.
- Stolz, Thomas. 2006. Contact-induced typological change. In Dieter Stern & Christian Voss (eds.), *Marginal linguistic identities: Studies in Slavic contact and borderland varieties*, 13–30. Wiesbaden: Harrassowitz.
- Stolz, Thomas. 2008. Romancisation worldwide. In Thomas Stolz, Dik Bakker & Rosa Salas Palomo (eds.), *Aspects of language contact: New theoretical, methodological and empirical findings with special focus on romancisation processes*, 1–42. Berlin/ New York: Mouton de Gruyter.
- Stolz, Thomas. 2010. Don't mess with ergatives!: How the borrowing of the Spanish indefinite article affects the split-ergative system of Chamorro. *Language Typology and Universals – STUF* 63(1). 79–95.
- Stolz, Thomas. 2012. Survival in a niche: On gender-copy in Chamorro (and sundry languages). In Martine Vanhove, Thomas Stolz, Aina Urdze & Hitomi Otsuka (eds.), *Morphologies in contact*, 93–140. Berlin/Boston: De Gruyter Mouton.
- Stolz, Thomas. 2015. Adjective-noun agreement in language contact: Loss, realignment, and innovation. In Francesco Gardani, Peter Arkadiev & Nino Amiridze (eds.), *Borrowed Morphology*, 269–301. Berlin/Boston: De Gruyter Mouton.
- Uldall, Hans Jürgen & William Shipley. 1966. *Nisenan texts and dictionary*. University of California Publications in Linguistics 46. Berkeley: University of California Press.
- Ultan, Russell. 1967. *Konkow grammar*. Berkeley: University of California Ph.D. dissertation.



Eeva Sippola

# Causatives and contacts in Chabacano

**Abstract:** This study examines Hispanization processes in the Philippines, with a special focus on causative serial verb constructions in Ternate Chabacano, a Spanish-lexifier Creole spoken in a Tagalog-dominant context in the Philippines. Based on a corpus of natural conversations and elicited structures collected in fieldwork as well as written materials, I provide a description of different causatives and related constructions that make use of the main causative verb *hasé*. I discuss their properties in relation to the different data types and the influence of the languages present in the formation of Chabacano and its current language ecology.

**Keywords:** causative, light verb constructions, Chabacano, Spanish, Tagalog

## 1 Introduction

This study examines causative constructions in Chabacano, a contact variety that was born during the Spanish colonial period in the Philippines. These Creole varieties have Spanish as the lexifier and Philippine languages as the adstrates. The Chabacano varieties are for the most part mutually intelligible but differ in their sociohistorical formation as well as the background and current linguistic situations of the communities (Lesho and Sippola 2013, 2014). The speakers of these Creole varieties live in multilingual environments, often speaking Chabacano, Tagalog or other Philippine languages, and English. The data for this study comes from Ternate Chabacano, which is one of the Chabacano varieties still spoken today, by approximately 3000 people in the town of Ternate in the Manila Bay region. The speakers are bilingual in Tagalog, and the variety's vitality status is threatened but relatively stable (Lesho and Sippola 2013). Ternate Chabacano was most probably formed sometime in the 17<sup>th</sup> century, and the present-day speaker population has very little contact with the lexifier language, Spanish. Most of the contact today is with Tagalog, the main language of the region, and English, which holds a prestigious position throughout the Philippines. The community in Ternate has been more isolated from the main urban centers than Cavite and Zamboanga, where Chabacano varieties are spoken with more lexifier traits in their lexicon and structure.

---

Eeva Sippola, University of Helsinki, e-mail: eeva.sippola@helsinki.fi

As is to be expected in a Creole (Baxter 2009), the valency-changing operations in Ternate Chabacano are done by analytic means from serial verbs to periphrastic constructions. Causatives are formed with verb serialization with *hasé* ‘make’, *dáli*<sup>1</sup> ‘give’ or *mandá* ‘order’ as V1. These present an array of causative constructions adding a transitive subject (A) argument and causing the reassignment of the syntactic functions of the original subject (S) of an intransitive clause or the A and transitive object (O) of a transitive clause. In addition, both *hasé* and *dáli* function as full lexical verbs but are also frequently used to incorporate English and Tagalog loans into Ternate Chabacano transitive clauses. The study provides a description of causatives and related constructions and discusses their syntactic properties in relation to the different data types and the contact situation of the language, i.e. the contextual factors that may account for the strategies speakers adopt to express causation. Possible lexifier and adstrate influence will also be explored.

The analysis builds on language data from Sippola (2011) collected between 2003 and 2010 in Ternate. It features naturalistic conversations, sociolinguistic interviews and elicited structures from in total fifty-five participants aged 11–86. Additionally, selected written materials have been included (Nigoza 2007), as “complete ditransitive sentences are rare in non-elicited data types” (Sippola 2011: 32).

This paper is structured as follows. Section 2 provides an overview of the background on Hispanization in the Philippines and causatives in contact. Section 3 presents an array of causative constructions in Ternate Chabacano and compares them to the lexifier and adstrate languages, while Section 4 expands on related constructions. Sections 5 and 6 cover the discussion and the conclusions.

## 2 Background

### 2.1 Hispanization in the Philippines

The study is placed in the framework of studies focusing on Hispanization processes world-wide, and especially in Austronesia. Thomas Stolz and associates’ work in this area of research has been groundbreaking (cf. Stolz 1996, 2002, 2022; Stolz et al. 2008a,b; Stolz et al. 2021; Stolz and Levkovych 2022; Zimmermann and Stolz 2001). Various studies regarding Spanish contact situations around the world have explored questions of what is possible and what is exceptional or rare in language contact. The studies in this paradigm have added to our knowledge and understanding of language contact, as they allow us to compare the outcomes of contact

---

<sup>1</sup> *Dáli* is occasionally spelled as *dale*.

situations, identify persistent patterns and extrapolate on their basis, as well as formulate generalizations which go beyond the special case of Hispanization and are relevant for studies of language contact in general (Stolz et al. 2008b: vi).

The Philippines was under Spanish colonial rule from the late 1500s to 1898. Although the contact with Spain was never very intense and did not involve large Hispanic population groups, it led to different outcomes of Spanish in contact with local languages. This included the formation and later maintenance of a Philippine variety of Spanish (Quilis and Casado-Fresnillo 2008) and heavy borrowing from Spanish into local languages (e.g. Tagalog, Bisaya, Ilokano) (Bowen 1971; Stolz 1996; Wolff 2001). In addition, several Philippine-Spanish Creoles developed as a result of this contact, including in the Manila Bay region (in Manila, Cavite, and Ternate) and in Mindanao (in Zamboanga, Cotabato, and Davao) (cf. Sippola 2020). After the Spanish period, the Philippines were under American rule until the country gained independence in 1946. The influence of English has grown starting from the American period, replacing the Hispanic influence and creating yet another layer of contact for the Philippine linguistic reality. Another notable development is the creation and support of the national language, Filipino, which has been promoted in the education system throughout the independence (cf. Gonzalez 1998).

Chabacano varieties have historically been spoken in two Philippine regions, Manila Bay in the north (including Ternate, Cavite City, and Manila), and Mindanao in the south (including Zamboanga City and its surrounding areas, Cotabato, and Davao). There is no conclusive agreement on the history of the formation of these varieties (cf. Whinnom 1956; Fernández and Sippola 2018). According to Whinnom (1956), Chabacano developed out of a Portuguese-based contact variety in the island of Ternate in the Moluccas in Eastern Indonesia. From there, early Chabacano would have been transferred to the Philippines in the 17th century by *Mardikas*, local Christians who came to the Philippines with the Spanish. Whinnom's account has been challenged by both historical and linguistic evidence (Lipski 1988; Fernández 2012; Fernández and Sippola 2022). Instead, according to Fernández (2011, 2012), Chabacano could have crystallized and stabilized gradually from contact varieties in different locations of the Philippines. The differences between today's Chabacano varieties are connected with their local histories of formation and linguistic ecologies in different parts of the Philippines (cf. Sippola and Lesho 2020). As Sippola and Lesho (2020) report, Zamboanga Chabacano is a widely spoken language, but Cavite and Ternate Chabacano are both endangered (Lesho and Sippola 2013). The speakers of these varieties say that they can understand each other, but they consider their languages and communities to be distinct (Lesho and Sippola 2014).

Although the overall aim of the study of Hispanization is a comparative one, the Chabacano case study presented here will allow us to see how the linguistic

setting and social bilingualism are reflected in the causatives. This example provides some insight into how mutual compatibility, semantic and structural overlaps, as well as general tendencies in contact-induced change mold the causative constructions in Chabacano.

In line with many studies in the Hispanization paradigm, I take a general approach to contact-induced change and grammaticalization (or external vs. internal change): these are often framed as separate issues, but they often work jointly to trigger grammatical change (cf. Matras 2011). Furthermore, following Heine and Kuteva (2003: 561), I do not assume that language contact as observed in pidgins and creoles is qualitatively different from that observed in other languages. It is also known that universal processes of grammatical change are involved in the areal diffusion of linguistic features, which is especially relevant in social situations of large-scale bilingualism (Heine and Kuteva 2003), as in the Philippines, where Tagalog and other larger Philippine languages are in socially dominant positions. Contact-induced changes have communicative and/or sociolinguistic motivations, as the speakers aim to make the categories existing in the languages in contact mutually compatible and more readily intertranslatable (Heine and Kuteva 2003). The outcomes of contact are affected by the availability of resources for potential replication, which speakers as agents of change select and use.

## 2.2 Causatives in contact

Causative expressions denote a complex situation consisting of two component events (Comrie 1989: 165–166; Song 2001: 256–259):

- (i) the causing event, in which the causer does something or initiates something;  
and
- (ii) the caused event, in which the causee carries out an action, or undergoes a change of condition or state as a result of the causer's action.

Causation is typically viewed as a morphologically signaled process, which introduces an agent to the valency of verbs yielding verbs (and clauses) with  $n+1$  arguments, but periphrastic causative constructions are also widely attested (Kittilä 2009; Song 2013). In many Ibero-Asian Creoles, causatives are expressed in serial verb constructions (SVCs) (Baxter 2009).

Verb serializations are complex predicates that contain a sequence of verbs within the same predicate. SVCs do not constitute a unique category, and they are identified by way of structural and semantic criteria (Aikhenvald 2006; Baxter 2009: 66–67). Semantically, two basic classes have been established for SVCs: an asymmetrical class and a symmetric class (Aikhenvald 2006: 22). The first is con-

stituted by a closed-class verb and an open-class one and the structure expresses a single event represented by the open-class verb modified by the closed-class verb. In contrast to this, the symmetric SVC, often known as linking or chaining, comprises two or more open-class verbs, expressing a single overall event (Baxter 2009: 67). Due to their monoclausality, the lack of conjunctions, coordination, subordination or other syntactic dependencies as well as the lack of pauses within the structure, SVCs display strong tendencies of lexicalization and grammaticalization, thus blurring the limits of the V1s as full verbs (Aikhenvald 2006).

Our analysis will be concerned with asymmetrical SVCs that involve valency modification in the sense of increasing the number of arguments of the major verb expressing cause. The minor verbs in asymmetrical SVCs represent a wide range of semantic relations, which tend to be grammaticalized, their main functions including the expression of direction/orientation, aspect, modality, complementation, and valency modification (Aikhenvald 2006: 22–28, 34). Valency modification SVCs have two functions; first, valency increasing and identification functions, including causative (with verbs such as *do*, *give*, *leave*, *say*, *put*, *undertake*), benefactive, instrumental, and comitative functions, and second, valency reduction functions, such as the passive.

In previous studies of Ibero-Asian Creoles, two types of causative SVCs have been attested: the *direct causative* and the *facilitative causative* (Baxter 2009). According to Baxter (2009: 69) truly causative SVCs contain the semantic component of control [+ control] whereas the facilitative causatives permit a non-causative reading. The control situation imposed by the fully causative SVC implies the realization of the caused situation, while the facilitative does not entail this implication. The Indo-Portuguese varieties of Diu, Daman, and Sri Lanka, the Malayo-Portuguese contact varieties of Malacca, Batavia, Tugu and Macau, as well as the Philippine Chabacano varieties of Cavite, Ternate, and Zamboanga present causative verb serialization (Baxter 2009).

Previous research on Chabacano has identified Zamboanga Chabacano SVCs<sup>2</sup> with causative functions for *ase* ‘make, do’, *manda* ‘order’, *dali* ‘give’, *lyiba* ‘take, carry’ and *deha* ‘let, permit’, *ase* being the most frequent one of them (Forman 1972: 210–211; Frake 1980: 292; Riego de Dios 1989). *Ase*, *manda*, *deha* and *lyiba* express direct causative functions, while *dali* expresses facilitative causative (Baxter 2009). For Cavite Chabacano, Llamado (1972: 80–82) mentions *haci* ‘make’ and *dale* ‘give’; and for Ternate Chabacano, Sippola (2011: 256–259) presents *hasé* ‘make’, *mandá* ‘order’ and *dáli* ‘give’. The analysis of the causatives is mostly descriptive in these

---

<sup>2</sup> Forman (1972) calls the SVCs verbal chains, while Frake (1980) classifies them as verbal expressions.

works, but Baxter (2009: 75) explains their common occurrence based on the convergence of a number of semantic and structural precedents in the lexifier and the substrate languages, as well as, at least in the case of Kristang of Malacca, a tendency to form verbal complexes.

### 3 Causative constructions in Ternate Chabacano

#### 3.1 Causative serial verbs with *hasé*, *dáli* and *mandá* in Ternate Chabacano

The valency-changing operations in Ternate Chabacano are done by analytic means from serial verbs to periphrastic constructions. These present an array of causative constructions adding an A argument and causing the reassignment of the syntactic functions of the original S of an intransitive clause or the A and O of a transitive clause.

Causatives in Ternate Chabacano are formed with verb serialization with *hasé* ‘make’ (realized as *tasé* with the imperfective marker and *asé* without the initial *h*) (1)–(2), as V1.

- (1) Ternate Chabacano (Sippola fieldnotes)

*Tasé            lótru    prosesyón,    tasé            karakól,    tasé*  
 IPFV.make 3PL    procession IPFV.make    caracol    IPFV.make  
*beylá    kon    kel    santo    nínyo.*  
 dance    ACC    DEF    holy    child

‘They make a procession, they dance the caracol, they make the holy child dance.’

- (2) Ternate Chabacano (Sippola 2011: 106)

*Asé            rin    éli    pará    kása    riyál... .*  
 PFV.make    too    3SG    stand    casa    real  
 ‘He also built a casa real [= municipal hall]. . .’

Other verbs that are used in SVCs with causative and facilitative functions are *dáli* ‘give’ (3) or *mandá* ‘order’ (4). Also *dehá* ‘leave, allow’ and *yebá* ‘carry’ have been documented (cf. Forman 1972: 211).

- (3) Ternate Chabacano (Sippola 2011: 138)  
*Kabár dáli komé yo kon kel mi pábu.*  
 after give eat 1SG DAT DEF my turkey  
 ‘Then I feed my turkeys.’
- (4) Ternate Chabacano (Sippola 2011: 258)  
*A-mandá matá Milin kel guyabánus.*  
 PFV-order kill NAME DEF soursop.tree  
 ‘Milin ordered the soursop tree to be killed.’

The serial verb constructions have single specification for TAM, negation and evidentiality, but pronouns and other clitics can appear in second position after the V1 as in (2), where the Tagalog origin adverbial particle *rin* and the 3SG pronoun occur in second position. In the causative serial verb constructions, the actor of the V2 is presented as an undergoer and marked with the accusative marker *kon* for human and animate objects, while inanimates do not present overt marking. The statue of the Holy Child in (1) can be interpreted as personified and human, and therefore marked with *kon*, while in (2), *kása* ‘house’ is not marked.

Some remarks are in order when considering these examples from a formal point of view. The first strategy, the use of *hasé* ‘make’ presents some variation and reduction on the phonological level, as forms *asé* and *sé* are also observed. When the V1 in the verbal construction is combined with the TAM markers, the variation makes it challenging to distinguish between the basic form *asé* ‘make’ and the form *a-sé* [PFV-make] ‘made’ where the aspect is marked. As is common for some frequent expressions in the Chabacano varieties, such as locative existential constructions, in causative constructions the preverbal markers combine with the verb, providing the following forms:

ta+(h)asé [IPFV+make] > *tasé*  
 (y)a+(h)asé [PFV+make] > *(y)asé*  
 di+(h)asé [CTPL+make] > *disé*

The initial /h/ in the Ternate variety sets it apart from the varieties spoken in Cavite and Zamboanga, as well as Philippine Spanish, and it can be seen as a trace of older or variable Spanish influence at the period of formation of the Ternate variety. Different forms may occur within a single turn, showing evidence of an ongoing process of negotiation where the item is seen as a lexical verb and as a grammatical

element. It is also worth noting that in verb chains, V1 can be accentuated, which casts doubts on its nature as a grammatical item.

Although in elicited structures adding causers (+A) to both intransitive and transitive clauses is possible and all arguments can be stated, natural data show that clauses generally occur with a maximum of two arguments, and that in causative constructions, causees are mostly implied and unstated, as in (4) and (5).

(5) Ternate Chabacano (Sippola 2011: 256)

*Kabáandu misa, disé salí ya el prosesyón.*  
 after mass IRR.make go.out already DEF procession  
 ‘After mass, the procession will be taken out.’

According to Frake (1980: 293), *manda* in Zamboanga Chabacano is used in intransitives with an agentive subject as in (5), while Baxter (2009: 83) interprets *manda* as modifying a V2 that has a non-specific agent, or a specific experiencer, as in (4). This aspect of the unstated causees needs further explanation to which we will return later on in the discussion in Section 5.

### 3.2 Comparison to causatives in Spanish and Tagalog

In Spanish, causation can be expressed with lexical causatives, complex predicates or periphrastic constructions, as in Table 1.

**Table 1:** Causative Constructions in Spanish (after Zyzik 2014: 78).

Causation	Construction type	Example
direct	<i>lexical</i>	<i>quemar</i> ‘burn’, <i>matar</i> ‘kill’
indirect	<i>complex predicate</i>	<i>hacer</i> -infinitive: <i>hacer reír</i> ‘make laugh’
indirect	<i>periphrastic</i>	<i>hacer que</i> -subjunctive: <i>hacer que trabajen</i> ‘make that they work’

In Spanish, causative verbs are formed through morphological causative derivation of adjectives using several affixes listed in Table 2 (Gonzalez Vergara 2004: 64; Serrano Dolader 1999).

**Table 2:** Causative morphological derivation in Spanish.

affix	Spanish adjective	Spanish verb
-ar	<i>limpio</i> 'clean'	<i>limpiar</i> 'to clean'
-ear	<i>blanco</i> 'white'	<i>blanquear</i> 'to whiten'
-izar	<i>central</i> 'central'	<i>centralizer</i> 'to centralize'
-ificar	<i>auténtico</i> 'authentic'	<i>autenticar</i> 'to authenticate'
-ecer	<i>húmedo</i> 'wet'	<i>humedecer</i> 'to dampen'
a-...-ar	<i>fino</i> 'sharp', 'fine'	<i>afinar</i> 'to tune'
a-...-ecer	<i>claro</i> 'clear'	<i>aclarecer</i> 'to clear up'
en-...-ar	<i>borracho</i> 'drunk'	<i>emborrachar</i> 'to make drunk'
en-...-ecer	<i>bello</i> 'beautiful'	<i>embellecer</i> 'to beautify'

The verbs in (6a) combine to form a complex verb, as the cause argument cannot be placed between *hacer* and the infinitive, and it is grammatically encoded as an object. In (6b), a periphrastic construction with a subjunctive clause is used to express causative meanings.

(6) Spanish (Mourelle de Lema 1981: 20)

- a. *Juan hizo beber el café a Antonio.*  
 Juan made.3SG drink the coffee DAT NAME  
 'Juan made Antonio drink the coffee.'
- b. *Juan hizo que Antonio bebiera el café.*  
 Juan made.3SG that Antonio drank.SUBJ the coffee  
 'Juan made Antonio drink the coffee.'

In Chabacano, subjunctive mode does not form part of the grammatical system, and instead of these two options, the causative SVC is used. In sum, the expression of causatives follows fairly similar paths in both Spanish and Chabacano, with the major differences focusing on the nature of the verbal constructions.

In Tagalog, different causative expressions are possible. First, causative-focus verbs are predicates that select as topics nominals expressing the cause of the action, as in (7) and (8) (Schachter and Otanes 1972: 313–314). These can be derived from transitive and intransitive verbs (of the groups *-um-*, *ma-*, *mag-* and *mang-*), and they are prefixed according to the verb group with the prefixes *i-*, *ika-*, *i(ka)pag-*, or *i(ka)pang-*.

(7) Tagalog (Schachter and Otanes 1972: 313)

- Ikinaluha ni Nena ang usok.*  
 shed.tear.CF UA NAME T smoke  
 'The smoke made Nena shed tears.'

- (8) Tagalog (Sippola fieldwork data)

*Ikinaganda ni Jasmine ang talino niya.*  
 become\_beautiful.CF UA Jasmine T intelligence her  
 'Her intelligence made Jasmine beautiful.'

Another type of causative verbs in Tagalog is that of indirect action verbs (e.g. Schachter and Otnes 1972) which are formed with complex affixes that include the prefix *pa-*. The person(s) etc. designated by the actor topic or actor complement are not necessarily expressed as performing the action of the verb, but rather, as permitting or causing this action to be performed (McFarland 1984: 25; Schachter and Otnes 1972: 321). In total, there are eight different focus types of indirect action causative verbs and these verbs normally express actions whose occurrence is caused or permitted by some particular person. In sentences with causative indirect action verbs in predicate position, the sentence topic normally expresses the condition responsible for the occurrence of the action, while an indirect actor complement expresses the person responsible for it, as in (9) or (10). In other words, the causer is marked in the same way as the non-causative actor complement, and the causee is marked in the same way as the non-causative direction complement (McFarland 1984: 25).

- (9) Tagalog (Schachter and Otnes 1972: 321)

*Nagpaluto ako ng adobo kay Rosa.*  
 IAF.COOK 1SG.FA OBJ adobo UA Rosa  
 'I let/made/had Rosa cook some adobo.'

- (10) Tagalog (McFarland 1984: 25)

*Pinabigyan ko siya ng pera.*  
 cause.give 1SG.FA she OBJ money  
 'I had someone give her some money.'

From the point of deriving different indirect-action verbs there are several issues that we have to leave out of discussion at this point, but for the sake of clarity, it should be noted that when a causative-focus verb, as in (7) and (8), is transformed to an indirect-action verb, the topic of the causative-focus indirect-action verb expresses the cause of the action, as in (11):

- (11) Tagalog (Schachter and Otnes 1972: 330)

*Ikinapagpabunot ni Eddie ng ngipin sa dentista ang sakit.*  
 IAF.CF.pull UA Eddie OBJ tooth UA dentist T pain.  
 'The pain caused Eddie to have the dentist pull (his) tooth.'

Tagalog causative expressions are thus mainly formed by derivation on the verb, forming both causative direct-action verbs and indirect-action verbs expressed with verbal affixes. In comparison to Chabacano, the Tagalog system relies more on morphological marking on the verb in the expression of the causation, as well as the related marking of the core arguments.

## 4 *Hasé* in related constructions

In addition to SVCs, Ternate Chabacano has a number of constructions with *hasé* that can take a verb or a noun as the second component. Occasionally the same meaning can be expressed in either verbal or nominal component, as in *sé kaminá* [V1 make + V2 walk] or *asé pasyál* [V1 make + stroll] ‘go for a stroll’ (12).

(12) Ternate Chabacano (Sippola 2011: 168)

*Ta-andá mótru sé kaminá manga... ta-andá asé pasyál.*  
 IPFV-go 1PL make walk PL IPFV-go make stroll  
 ‘We go and walk [around]. . . [We] go for a stroll.’

The resulting constructions are generally transitive, even though intransitive constructions are also possible, as in (13) (see also Table 3). The expression *asé handá* is modelled in the Tagalog prefixed verb *ipaghandá* ‘to entertain’.

(13) Ternate Chabacano (Sippola 2011: 168)

*Kába kasá asé handá éli.*  
 After marry make party 3SG  
 ‘After getting married, he threw a party.’ [Past of context]

In these constructions, the second element can be of Spanish, Tagalog or English origin, as shown in (12), (13) and (14).

(14) Ternate Chabacano (Sippola fieldwork data)

*Nung sábi yo kung kwánda éli disé ritáy.*  
 NEG know 1SG if when 3SG CTPL.make retire  
 ‘I do not know when she will retire.’

A non-exhaustive list of formations of this type without a causative meaning is given in Table 3.

**Table 3:** *Hasé*-constructions in Chabacano.

<i>hasé</i> construction	transitivity	Etymological origin of the element containing meaning
<i>hasé kombersá</i> ‘talk’ > <i>platiká</i>	–	Sp. <i>conversar</i> ‘to talk’, ‘to converse’
<i>hasé tyénda</i> ‘sell’ > <i>bendé</i>	+	Sp. <i>tienda</i> ‘shop’
<i>hasé ritáyr</i> ‘retire’	–	Eng. <i>retire</i>
<i>hasé bakasyón</i> ‘go on holiday’	–	Sp. <i>vacación</i> / Eng. <i>vacation</i>
<i>hasé bíru</i> ‘to joke’	–	Tag. <i>birò</i> ‘joke’
<i>hasé akít</i> ‘persuade’ / ‘seduce’	+	Tag. <i>umakit</i> ‘seduce’
<i>hasé ingát</i> ‘take care’	–	Tag. <i>ingat</i> ‘care’
<i>hasé kwéntu</i> ‘tell stories’	+	Sp. <i>cuento</i> ‘story’
<i>hasé pála</i> ‘excavate’	+	Sp. / Tag. <i>pala</i> ‘shovel’
<i>hasé sepílyo</i> ‘brush’	+	Sp. / Tag. <i>cepillo</i> ‘brush’
<i>hasé supórt</i> ‘support’	+	Eng. <i>support</i>
<i>hasé text</i> ‘send an sms’	+	Eng. <i>text</i>
<i>hasé wítnes</i> ‘testify’	+	Eng. <i>witness</i>

From these examples, a complex picture emerges of Chabacano constructions including *hasé* with different options for analysis. First, *hasé* combines with a base that can be a noun or a verb. It can be analyzed as a light verb, as it has little or no semantic content of its own. It clearly forms part of the verb, as the preverbal markers attach to it. Second, it functions as a verbalizer that allows for the incorporation of items that do not conform to the canonical Chabacano verb form with an accentuated final vowel, as in *kantá* ‘sing’, *beylá* ‘dance’, etc. In this manner, verbs can also be borrowed from diverse languages including English into Chabacano. Thirdly, as already mentioned in Section 3.1, the verb is sometimes expressed in a contracted form, thus making it more affix-like. This might be connected to a diachronic cycle in which full verbs become auxiliary and light verb-like, then clitics and finally affixes.

When comparing these constructions to light verb constructions in the lexifier (Spanish *verbo ligero* or *verbo de apoyo*), it becomes clear that Spanish presents very similar patterns. Spanish light verb constructions include a construction with *hacer* ‘to make’ among other verbs such as *dar* ‘give’,  *echar* ‘throw’, *tener* ‘have’ and *tomar* ‘take’ that generally takes an abstract noun as a complement (Piera and Varela Ortega 1999: 4415–4418). *Hacer* is used with action verbs in general (Sanromán Vilas 2011), and it forms both transitive and intransitive constructions, as in (15):

## (15) Spanish (own competence)

<i>hacer mención</i>	<i>mencionar</i>	‘mention’
<i>hacer falta</i>	<i>ser necesario</i>	‘to be necessary’
<i>hacer presente</i>	<i>comunicar</i>	‘communicate’
<i>hacer un descanso</i>	<i>descansar</i>	‘take a rest’
<i>hacer uso</i>	<i>usar</i>	‘use’

In Chabacano, there seem to be some differences with the Spanish construction, because the complementing verb or noun is often not abstract (see Table 3).

If a Chabacano light verb construction (16a) is translated into Tagalog, the resulting construction is expressed in (16b) with the verb *pinagtawanan* ‘laughed at’ in the undergoer voice.

## (16) Ternate Chabacano (Modified from Vibar and Navarro 2006: 15)

- a. *Yasé risáda kel prinsipál kung kel maéstra.*  
 PFV.make laughter the principal OBJ the teacher
- b. *Pinagtawanan ng prinsipal ang maestra.*  
 Laughed UA principal T teacher.  
 ‘The teacher was laughed at by the principal.’

## 5 Discussion

### 5.1 Causes and consequences of contacts

From the examples presented in the previous sections, it is possible to see some semantic and structural overlaps between Chabacano and its main lexifier, Spanish. The use of *hasé* and other causative verbs in serial verb constructions is a clear case of matter replication from Spanish, as the form of the causativizing verbs is clearly derived from their Spanish counterparts: *hacer*, *dar*, *mandar*, etc. However, although the form is derived from Spanish, it is restructured in Chabacano serial verbs, as in Spanish causative expressions can be bi-clausal. Another clear example of the Spanish matter replication can be seen in the related constructions with *hasé* + V or N, which express a variety of meanings close to Spanish light verb constructions. The case of causatives and related expressions making use of *hasé* thus provides a clear case of lexifier influence that has been adapted to the Chabacano grammatical structure. However, as Table 4 shows, Chabacano has not derived its causative verbs directly from Spanish in all cases but uses SVCs to express many

meanings that are expressed with derivational affixes in Spanish. The resulting semantic and structural overlaps are not completely compatible, as the independent developments and general structural tendencies of the languages compared have influenced the outcome.

**Table 4:** Causative morphological derivation in Spanish and the corresponding Chabacano causatives.

affix	Spanish adjective	Spanish verb	Chabacano verb
-ar	<i>limpio</i>	<i>limpiar</i> ‘to clean’	<i>limpyá</i>
-ear	<i>blanco</i>	<i>blanquear</i> ‘to whiten’	<i>hasé blánku</i>
-izar	<i>central</i>	<i>centralizer</i> ‘to centralize’	-
-ificar	<i>auténtico</i>	<i>autentificar</i> ‘to authenticate’	-
-ecer	<i>húmedo</i>	<i>humedecer</i> ‘to dampen’	<i>hasé muháw</i>
a-...-ar	<i>fino</i>	<i>afinar</i> ‘to tune’	<i>apiná</i>
a-...-ecer	<i>claro</i>	<i>aclarecer</i> ‘to clear up’	<i>hasé kláru</i>
en-...-ar	<i>borracho</i>	<i>emborrachar</i> ‘to make drunk’	<i>embohrachá</i>
en-...-ecer	<i>bello</i>	<i>embellecer</i> ‘to beautify’	<i>hasé koryósu</i> <i>hasé áyus</i>

For its part, Tagalog (or Philippine) influence is perhaps more indirect. Naturally, semantic overlaps can be found, but as Tagalog makes use of morphological voice marking to express causatives in causative-focus and indirect-action-focus verbs, its influence is not as directly comparable as the Spanish one. But when examining cases where *hasé* is used in its contracted form *se*, seemingly losing its status as a lexical verb and turning more clitic- or affix-like, the morphological precedent of the Tagalog patterns becomes evident. For bilingual speakers of Chabacano and Tagalog (or other Philippine languages), it would not be surprising to reanalyze the first verb of the serial verb to an affix following the patterns of the dominant language in their surrounding environment. This is further supported by the fact that *hasé* is used as a verbalizer to incorporate borrowings from other languages that do not follow the canonical Chabacano verb stress pattern. Overall, these uses might point towards a grammaticalization process from full verbs and V1s in serial verbs to general verbalizing affixes with parallels in Tagalog prefixation, as in Tagalog *ipinag-away* [CF-fight] ‘cause to fight’. Furthermore, universal tendencies can be observed in the close association between transitivity and causativization that is seen in different uses of *hasé*.

The causatives and related constructions have provided some insight into how semantic and structural overlaps, as well as universal tendencies, mold the causative constructions in Chabacano.

## 5.2 Unsolved cases: Unstated causees and social motivations

Another interesting phenomenon is presented by unstated causees. When *hasé* or *mandá* occur with transitive verbs, as in *hasé kambyá* ‘make change’ or *asé kitá* ‘make take away’, it modifies generally a V2 that has a non-specific agent or a specific experiencer, as in (1) (see also Baxter 2009: 83). This can be found also in lexical causatives in (17):

(17) Ternate Chabacano (Sippola 2011: 264)

*Kayá lang a-matá kun-éli.*

Therefore just PFV-kill OBJ-3SG

‘That is why he was killed.’

It is common that the causee is left to inference in causativized transitive clauses, and in fact, there are very few or no examples found in naturalistic data and none in the spoken natural corpora collected in fieldwork. In elicited structures and in written data, causees can also occur when transitive sentences are modified, as in (18):

(18) Ternate Chabacano (Sippola fieldwork notes)

*Su mádri a mandá kuneli kasá kun quel ómbri de*  
her mother PFV order 3SG.OBJ marry OBJ DEF man of  
*bisínu pwéblu.*

neighbour village

‘Her mother made her marry the man from the neighbouring village.’

This requires further study. One possible reason could be the tendency for subject omission in general in spoken discourse. Sippola (2011: 262) mentions that subject omission is common in Ternate Chabacano, although the identification of the element depends on the context, such as the previous discourse or shared information (see also Lipski 1994). For example, Ternate Chabacano allows null subjects in sequential verbal structures when the subject referent is available in the immediate discourse context or is of shared knowledge, as in (19).

(19) Ternate Chabacano (Sippola 2011: 263)

*Yo namán. . . lebantándo na durmí, de-labá kára, disé*  
1SG only waking PREP sleep CTPL-wash face CTPL.make  
*totbrash. . . kabándo, di-prepará akél para komé na almwésu.*  
toothbrush after CTPL-prepare that for eat PREP lunch

‘Well, I, when getting up, I will wash my face, brush my teeth, after, I will prepare that [food] for lunch.’

Another motivation could be found in the influence of the Philippine voice system. In Philippine languages, accusative clauses are intransitive, and transitive clauses are in patient voice. Also, the syntactic alignment of the Philippine languages is accusative, oriented towards the agent, when the theme or the patient is indeterminate, and ergative, oriented towards the patient, when the theme or the patient is determinate (Fernández 2007, 2009: 427; Payne 1982; Nolasco 2005). As we notice in the previous examples, the patients seem to be determinate, pointing towards the fact that the orientation is towards the patient in these clauses. However, the Chabacano verb shows no inflection that would allow for the expression of the voice, while in the Philippine languages the verb expresses the focused element and the voice of the predicate. In addition, argument marking seems to be influenced by the adstrate influence of oblique marking to different degrees in the Chabacano varieties (*kan*, *kang*, *kay* and *sa*) (Fernández 2007, Forman 1993). For reasons of space and problems with the current data, these are only initial observations that might explain the case of unstated causes for Chabacano. Again, an external reason for this would be the bilingualism of the speakers who could resort to the Philippine system in the expression of the Chabacano causative sentences. Future research should compare the frequency of unstated causees with subject omission in other types of clauses.

In addition to the general statements of societal bilingualism facilitating convergence and replication of grammatical matter, it is difficult to pinpoint specific social motivations for the expression of causatives in Chabacano. Sippola and Lesko (2020) connect differences in grammaticalization processes of Chabacano varieties to sociolinguistic reasons, highlighting fossilization of Spanish forms and the adstrate influence from Philippine languages that cause more Tagalog-like patterns. Language endangerment and the shift to Tagalog are noticeable, as is the historical remoteness of Ternate Chabacano from Spanish centers of power in the Philippines. However, the expression of causatives and related constructions, the affix-like behavior of *se-*, and the unstated causees as a result of Philippine influence remains at most conjectural for the time being.

## 6 Conclusions

In this article, I have presented an overview of causative constructions and related expressions in Chabacano in order to shed light on the valency changing operations in the Hispanization process in the Philippines. The overview has shown

that matter and pattern replication can be observed to different degrees between Spanish and Chabacano, as well as between Tagalog and Chabacano. Chabacano causative constructions draw on Spanish expressions with the verbs *hasé*, *dáli* and *mandá* (among others), although the Spanish derivational mechanisms have been somewhat lost in contact. This is not surprising as we know that much of the morphological productivity is not carried on to the new contact varieties formed at language contact.

Ternate Chabacano shows both matter and pattern replication of the Spanish elements. Causative markers like V1 express direct causation and their occurrence as a light verb in complex predicates with nouns and verbs is an example of this. It is not all lexifier influence, though, as the variation in the occurrence of the form *hasé*, *asé* and *se* would indicate. The semantic parallels with causative-focus and causative indirect-action-focus verbs in Tagalog seem to set a morphological precedent for the transformation of the accented verb *hasé* towards a prefix *se-*.

Universal tendencies can be observed in the close association between transitivity and causativization, although the use of *asé* as a verbalizer is an example of the typologically non-prototypical functions that are not confined to the causative prototype and agent introduction. Other internal developments due to Philippine influence such as the relationship between subject omission, unstated causees and Philippine-type voice and argument systems remain to be explored in future studies.

## Abbreviations

1/2/3	first/second/third person
A	transitive agent/subject
ACC	accusative
CAUS	causative
CF	causative focus
CTPL	contemplative
DAT	dative
DEF	definite
Eng.	English
FA	focs actor
IAF	indirect action focus
IPFV	imperfective
IRR	irrealis
N	noun
NEG	negation
O	transitive patient/object

OBJ	object
PFV	perfective
PL	plural
PREP	preposition
S	intransitive subject
SG	singular
SUBJ	subjunctive
Sp	Spanish
SVC	serial verb construction
T	topic
TAM	time, aspect, mode
UA	undergoer actor
V	verb

## References

- Aikhenvald, Alexandra. 2006. Serial Verb Constructions in Typological Perspective. In Alexandra Y. Aikhenvald & R.M.W. Dixon (eds.), *Serial verb constructions: a cross-linguistic typology*, 1–68. Oxford University Press.
- Baxter, Alan. 2009. Causative and facilitative serial verbs in Asian Ibero-romance Creoles – a convergence of substrate and superstrate systems?. *Journal of Portuguese Linguistics* 8(2). 65–90. <https://doi.org/10.5334/jpl.119>
- Bowen, J. Donald. 1971. Hispanic languages and influence on Oceania. *Current Trends in Linguistics* 8. 938–952.
- Comrie, Bernard. 1989 [1981]. *Language universals and linguistic typology*. Chicago: University of Chicago Press.
- Fernández, Mauro. 2007. Sobre el origen de *con* en chabacano. In Laura Morgenthaler García & Martina Schrader-Kniffki (eds.), *Romania en interacción: entre historia, contacto y política. Ensayos en homenaje a Klaus Zimmermann*, 457–478. Frankfurt am Main: Vervuert.
- Fernández, Mauro A. 2009. La partícula *con* y la organización de la transitividad en chabacano. In Montserrat Veyrat Rigat & Enrique Serra Alegre (eds.), *La lingüística como reto epistemológico y como acción social. Estudios dedicados al profesor Ángel López García con ocasión de su sexagésimo aniversario. Volumen 1*, 423–436. Madrid: Arco Libros SL.
- Fernández, Mauro. 2011. Chabacano en Tayabas: implicaciones para la historia de los criollos hispano-filipinos. *Revista internacional de Lingüística Iberoamericana* 17. 189–218.
- Fernández, Mauro. 2012. Leyenda e historia del chabacano de Ermita (Manila). *UniverSOS* 9. 9–46; 65–70.
- Fernández, Mauro & Eeva Sippola. 2018. On the chronology of the formation of the Chabacano varieties: A reply to Parkvall & Jacobs. *Journal of Ibero-Romance Creoles* 8. 38–56.
- Fernández, Mauro & Eeva Sippola. 2022. The lexicon and creole formation: A reply to Jacobs & Parkvall (2020). *Journal of Pidgin and Creole Languages* 37(2). 263–284.
- Forman, Michael L. 1993. Verb Serialization, Word Order Typology, and Zamboangueño: A Comparative Approach. *Oceanic Linguistics* 32. 163–182.

- Forman, Michael. 1972. *Zamboangueno texts with grammatical analysis. A study of Philippine Creole Spanish*. Ithaca, NY: Cornell University dissertation.
- Frake, Charles O. 1980. Zamboangueno verb expressions. In Charles O. Frake & Anwar S. Dil (eds.), *Language and cultural description: Essays by Charles O. Frake; selected and introduced by Anwar S. Dil*, 277–310. Stanford: Stanford University Press.
- González Vergara, Carlos. 2004. Sobre la formación de verbos causativos deadjetivales. Algunas regularidades semánticas. *Onomázein* 10(2). 57–92.
- Gonzalez, Andrew. 1998. The language planning situation in the Philippines. *Journal of Multilingual and Multicultural Development* 19(5). 487–525.
- Heine, Bernd & Tania Kuteva. 2003. On contact-induced grammaticalization. *Studies in Language* 27(3). 529–572.
- Kittilä, Seppo. 2009. Causative morphemes as non-valency increasing devices. *Folia Linguistica* 43(1). 67–94. <https://doi.org/10.1515/FLIN.2009.002>
- Lesho, Marivic & Eeva Sippola, 2013. The sociolinguistic situation of the Manila Bay Chabacano-speaking communities. *Language Documentation & Conservation* 7. 1–30.
- Lesho, Marivic & Eeva Sippola. 2014. Folk perceptions of variation among the Chabacano creoles. *Revista de Crioulos de Base Lexical Portuguesa e Espanhola* 5. 1–47.
- Lipski, John M. 1988. Philippine creole Spanish: Assessing the Portuguese element. *Zeitschrift für romanische Philologie* 104. 25–45.
- Lipski, John. 1994. The evolution of null subjects in Philippine Creole Spanish. 1994 *Mid-America Linguistics Conference Papers*, Volume II, 387–401. Lawrence, Kansas: University of Kansas Linguistics Department.
- Llamado, Librada C. 1972. The Phrase structures of Cavite Chabacano. *Philippine Journal of Linguistics* 3. 67–96.
- Matras, Yaron, 2011. Grammaticalization and language contact. In Heiko Narrog & Bernd Heine (eds.), *The Oxford Handbook of Grammaticalization*, 279–290. Oxford: Oxford University Press.
- McFarland, Curtis D. 1984. Tagalog Causative Verbs. *Philippine Journal of Linguistics* 15. 19–47.
- Mourelle de Lema, Manuel. 1981. Los verbos causativos en español. *Thesaurus* 36(1), 14–22.
- Nigoza, Evangelino. 2007. *Bahra. Manga historia, alamat, custumbre y tradiciong di Bahra: The history, legends, customs and traditions of Ternate, Cavite*. Dasmariñas: Cavite Historical Society.
- Nolasco, Ricardo 2005. The Chabacano challenge to Philippine ergativity. In Danilo T. Dayag & Stephen J. Quakenbush (eds), *Linguistics and language education in the Philippines and beyond: A festschrift in honor of Ma. Lourdes S. Bautista*, 401–433. Manila: De La Salle University Press.
- Payne, Thomas E. 1982. Role and Reference Related Subject Properties and Ergativity in Yup'ik Eskimo and Tagalog. *Studies in Language* 6(1). 75–106. <https://doi.org/10.1075/sl.6.1.05pay>
- Piera, Carlos & Soledad Varela Ortega. 1999. Relaciones entre morfología y sintaxis. In Ignacio Bosque & Violeta Demonte (eds), *Gramática descriptiva de la lengua española* 3. *Entre la oración y el discurso*, 4367–4422. Madrid: Espasa.
- Quilis, Antonio & Celia Casado-Fresnillo 2008. *La Lengua Española en Filipinas: Historia; Situación Actual; el Chabacano; Antología de textos*. Madrid: CSIC.
- Riego de Dios, Maria Isabelita. 1989. A Composite Dictionary of Philippine Creole Spanish. *Studies in Philippine Linguistics* 7(2). 1–210.
- Sanromán Vilas, Begoña. 2011. En torno a la vaciedad semántica de los verbos de apoyo. *Actas del XII Simposio Internacional de Comunicación Social*, 59–63. Santiago de Cuba.
- Schachter, Paul & Fe T. Otanes. 1972. *Tagalog reference grammar*. Berkeley: University of California Press.

- Serrano Dolader, David. 1999. La derivación verbal y la parasíntesis. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*. 3. *Entre la oración y el discurso*, 4683–4755. Madrid: Espasa Calpe.
- Sippola, Eeva. 2011. *Una gramática descriptiva del chabacano de Ternate*. Helsinki: Unigrafia.
- Sippola, Eeva. 2020. Contact and Spanish in the Pacific. In Raymond Hickey (ed.), *The Handbook of Language Contact*, 453–468. Hoboken: Wiley-Blackwell.
- Sippola, Eeva & Marivic Lesho. 2020. Contact-induced grammatical change and independent development in the Chabacano creoles. *Bulletin of Hispanic Studies* 97(1). 103–121.
- Song, Jae Jung. 2001. *Linguistic Typology*. Routledge.
- Song, Jae Jung. 2013. Periphrastic Causative Constructions. In Matthew S. Dryer & Martin Haspelmath (eds.), *WALS Online* (v2020.4) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13950591>
- Stolz, Thomas 1996. Grammatical Hispanisms in Amerindian and Austronesian languages. *Amerindia* 21. 137–160.
- Stolz, Thomas. 2002. General linguistic aspects of Spanish-Indigenous language contacts with special focus on Austronesia. *Bulletin of Hispanic Studies* 79. 133–158.
- Stolz, Thomas. 2022. Voy – para – siempre: Three Spanish-derived function words and the chamorro irrealis. In Nataliya Levkovich (ed.), *Susceptibility vs. Resistance: Case studies on different structural categories in language-contact situations*, 91–181. Berlin/Boston: De Gruyter.
- Stolz, Thomas, Deborah Arbes & Christel Stolz. 2021. Pero – Champion of Hispanization? On the challenges of documenting function word borrowing in Mesoamerican languages. In Danae Perez & Eeva Sippola (eds.), *Postcolonial language varieties in the Americas*, 17–54. Berlin/Boston: De Gruyter.
- Stolz, Thomas, Dik Bakker & Rosa Salas Palomo (eds.). 2008a. *Aspects of Language Contact: New Theoretical, Methodological and Empirical Findings with Special Focus on Romancisation Processes*. Berlin/Boston: Mouton de Gruyter.
- Stolz, Thomas, Dik Bakker & Rosa Salas Palomo (eds.). 2008b. *Hispanisation: The impact of Spanish on the lexicon and grammar of the indigenous languages of Austronesia and the Americas*. Berlin/Boston: Mouton de Gruyter.
- Stolz, Thomas & Nataliya Levkovich. 2022. Parallel Romancization: Chamorro and Tetun Dili – two heavy borrowers compared. In Nataliya Levkovich (ed.), *Susceptibility vs. Resistance: Case Studies on Different Structural Categories in Language-Contact Situations*, 393–466. Berlin/Boston: De Gruyter.
- Vibar, Arwin & Toribio Navarro. 2006. Update on Ternate Chabacano. Presentation in 9<sup>th</sup> Philippine Linguistics Congress (25–27 January 2006). Quezon city: University of the Philippines.
- Whinnom, Keith. 1956. *Spanish contact vernaculars in the Philippine Islands*. Hong Kong: Hong Kong University Press.
- Wolff, John 2001. The influence of Spanish on Tagalog. In Klaus Zimmermann & Thomas Stolz (eds.), *Lo propio y lo ajeno en las lenguas austronésicas y amerindias: procesos interculturales en el contacto de lenguas indígenas con el español en el Pacífico e Hispanoamérica*, 233–253. Madrid: Iberoamericana.
- Zimmermann, Klaus & Thomas Stolz. 2001. *Lo propio y lo ajeno en las lenguas austronésicas y amerindias: procesos interculturales en el contacto de lenguas indígenas con el español en el Pacífico e Hispanoamérica*. Madrid: Iberoamericana.
- Zyzik, Eve. 2014. Causative verbs in the grammar of Spanish heritage speakers. *Linguistic Approaches to Bilingualism* 4(1). 1–33. <https://doi.org/10.1075/lab.4.1.01ziz>

Christoph Schroeder

# What is heritage language research good for? A critical discussion, with examples from Turkish language contacts

**Abstract:** The term “heritage language” describes a sociolinguistically specific constellation of language contact – a non-dominant language in close contact with another language, which is the socially dominant language of the respective society. Three examples from recent research on clause combining in Turkish as a language in contact with German and English are discussed. They show that acknowledging the specific sociolinguistic constellation of (non)-dominance in heritage languages contributes to and benefits from language contact research. It is argued that the term “heritage” obscures more than it helps to make this contribution visible.

**Keywords:** heritage language, bilingualism, language contact, Turkish, clause combining

## 1 Introduction

The following discussion of the term “heritage language” aims to clear some fog and make it visible that much of the research that goes by this name contributes to contact linguistics and that the term obscures more than it helps to clarify the intersection between bilingualism research and contact linguistics. I begin with an examination of the term and then discuss three examples from research on the development of Turkish as a non-dominant language in Germany and the USA. They are intended to strengthen my argument.<sup>1</sup>

---

<sup>1</sup> This contribution is dedicated to my dear colleague Thomas Stolz. Thomas and I have been acquainted for a very long time (roughly 35 years, after all!) and share a common interest in contact linguistic research. I am much more limited in this area than Thomas; unlike him, I don't dare to venture into new waters as much as he does. Another area in which I find myself in common with Thomas' work is my mistrust of terminology. I see this in Thomas' work on colonial linguistics.

I benefit immensely from the co-operation with the following colleagues, without who the Turkish-related research and analysis, which I base Part 3 on, would have been impossible: Kateryna Iefremenko, Onur Özsoy, Cem Keskin and Jaklin Kornfilt. Furthermore, I am grateful to the editors of this volume for helpful comments.

---

**Christoph Schroeder**, University of Potsdam, e-mail: christoph.schroeder.iii@uni-potsdam.de

## 2 Heritage language, heritage language speakers

### 2.1 Definition

A common, and generally accepted definition of “heritage language” (HL) goes like this: “A language qualifies as a HL if it is a language spoken at home or otherwise readily available to young children, and crucially this language is not a dominant language of the larger (national) society” (Rothman 2009: 156).

The concept of a heritage language thus appears as a sociolinguistically specific constellation of language contact: A heritage language is a sociolinguistically non-dominant language that exists in close contact with the dominant language of society, which is another language. Speakers of this language become bilingual at the latest when they enter the societal institutions where the dominant language of the majority is used (e.g., kindergartens, schools).

### 2.2 The object of study in current research practice

When reviewing the academic literature on heritage languages and heritage language speakers, it becomes evident that with few exceptions, languages identified as “heritage languages” are allochthonous languages in countries where nation-building has historically been closely tied to the establishment of a national language, and where the languages identified as “heritage languages” have established themselves in the course of migration.

This is the first peculiarity of the term: In linguistic research (though not in education), the term is almost exclusively applied to allochthonous minority languages, not to autochthonous ones:

- In North American linguistic research, no publication applies the term to indigenous languages. For these languages, different terms exist, despite the evident (non-)dominance relationship.
- Similarly, research (in English) on heritage languages in Germany does not apply the term to recognized minority languages such as Sorbian, Frisian, or Romani.
- Studies on “German abroad” sometimes use the term “heritage language” (e.g. Shah et al. 2024 on German in South Africa), but usually they do not.
- In Franco-Romance studies, an interesting distinction exists between (i) *langue d’héritage*, the direct translation equivalent sometimes used for allochthonous minority languages, and (ii) *langue héréditaire*, which refers to French as an autochthonous language outside France (e.g., in Switzerland, Belgium).

- In Hispano-Romance studies, the equivalent term *lengua de herencia* seems to be well-established for contexts where Spanish is an allochthonous minority language; however, this does not (yet?) apply to indigenous languages in South America (see Mulík et al. 2022).<sup>2</sup>

Occasionally, translating “heritage language” into another language seems to challenge linguistic intuition. In linguistic publications in German, the equivalent term *Erbssprache* has sometimes been used (e.g., by Tracy 2014 and Gagarina 2014), but this use is marginal. Typically, the term *Herkunftssprache* is used, which conveys “origin” rather than “inheritance” or “transmission.” Similarly, the Turkish translation *miras dili* is occasionally used in linguistic studies published in Turkish, but only in order to refer to heritage languages in Europe or the USA (and not to minority languages in Turkey), and the term mainly provokes confusion.

## 2.3 Conceptual critique

The second peculiarity of the term “heritage language,” as well as “heritage speaker,” lies in the first part of the compound, “heritage.” Without much debate, it is generally assumed that “heritage” here means something akin to “something transmitted by or acquired from a predecessor” (Merriam-Webster Dictionary)<sup>3</sup>. This brings us to UNESCO, which with its *World Heritage Convention* seeks to protect tangible and intangible values: “Heritage is our legacy from the past, what we live with today, and what we pass on to future generations”.<sup>4</sup>

Interestingly, however, the *World Heritage Convention*, which is the UNESCO’s key document on world heritage and so far ratified by 196 states,<sup>5</sup> does not mention language at all. Although some member states include language in their working definitions of intangible heritage (see United Nations Educational, Scientific and Cultural Organization 2001), these references are exclusive to autochthonous or indigenous languages, with no mention of allochthonous languages.

Regardless of whether allochthonous or autochthonous languages are being discussed, the backward-looking perspective implied by the term “heritage” remains problematic (McCarty 2008, cited in Ennser-Kananen and King 2013: 1). “Heritage” conveys something static and isolable, potentially meant to be preserved as is – an

---

2 I am grateful to Annette Gerstenberg and Melanie Uth (both Potsdam University) for sharing with me their inside in the “heritage”-terminology concerning Romance languages.

3 Retrieved from: <https://www.merriam-webster.com/dictionary/HERITAGES>, accessed 2024-12-20.

4 Retrieved from <https://www.unesco.org/en/world-heritage>, accessed 2024-11-29.

5 Retrieved from <https://whc.unesco.org/en/conventiontext/>, accessed 2024-11-29.

understanding that contradicts the dynamic concept of language held by modern linguistics, which views language as arising from communicative and social interaction (see Ennser-Kananen and King 2013: 1).

## 2.4 Historical Background of the term (and more critique)

How, then, did the term “heritage language” come to be used in linguistics? Its establishment is relatively recent, with its explicit use beginning in the Canadian educational sector in the 1970s through the *Ontario Heritage Language Program* (Cummins 2005). Initially, the term encompassed “languages spoken by indigenous people or immigrants” (Cummins 1991: 601–602, cited in Ennser-Kananen and King 2013: 1). Until the year 2000, the term remained primarily within the educational sector, focusing on “heritage learners” and the schooling of migrant or indigenous minorities in North America, particularly in the USA and Canada. Around the early 2000s, a significant shift occurred: linguists began to focus on the “heritage language speaker”. This was likely influenced by the founding of the *Heritage Language Journal* in 2003. The journal became the primary venue for publishing research on the linguistic development of “heritage language speakers.” Researchers who had previously studied speakers of Russian, Korean, Spanish, etc. in the USA under the label of “American/Korean/Spanish/etc in the USA”, “minority languages”, “allochthonous languages”, or even “reduced languages” (Polinsky 1995) adopted the term “heritage” for their linguistic research agenda. This research explored issues such as the acquisition of non-dominant languages, the role of input in language acquisition, and linguistic phenomena like attrition, incomplete acquisition, erosion, or simplification – all of which sparked significant theoretical debate.

A key concept that has always been central to heritage language research is the “baseline”. This term was first explicitly mentioned in Polinsky and Pereltsvaig (2003: 136):

The language which constitutes the main input for heritage learners is the baseline. This baseline may be the same as the standard language promoted by the literature, media, or religion, but it can also be vastly different from the standard. Determining the baseline spoken by the original immigrants is crucial for our understanding of the lexicon and structure of a heritage language. Once the baseline is established, we are faced with the question of what factors in the history of a particular émigré community helped preserve the community’s language and what factors moved them away from the original language.

The baseline thus became the reference point for investigating how heritage speakers develop linguistic competence in their heritage language, and the influence of extralinguistic factors, such as input intensity and dominance relations between languages, on this development.

Now, of course, the exact determination of the baseline becomes a methodologically complex, if not almost impossible, endeavour: There is nowhere near enough data available to determine what exactly the Russian or Turkish etc. looked like, which provided the initial linguistic input for what then developed into the linguistic practice of the heritage language speaker. In order to escape this data gap, heritage language research helped itself by simply comparing monolingual speakers of the respective language (in the country where the language is a majority language) with the linguistic productions of the heritage speakers, and interpreting systematic deviations in the linguistic practice of the heritage language speakers from the linguistic practices of the monolingual speakers in the context of input factors and language loss.

This, I suppose, is one reason why the heritage language research programme focused exclusively on the languages of immigration and largely ignored autochthonous minority languages: monolingual speakers of autochthonous minority languages are difficult to find, and so the reference point for linguistic change, crucial for the research agenda of heritage language research, could not be identified. A second reason may have to do with who has the power over labelling: Immigrants and their descendants in the super-diverse societies of the Global North have less of a tendency to see themselves as a homogeneous group and to represent themselves to the outside world as such; indigenous minorities, on the other hand, do. They often have a lobby and therefore the power to demand their own labels and defend themselves against external labelling. Cummins (2005: 591) points this out early on: “Canadian First Nations communities generally do not see their languages as heritage languages and prefer to use terms such as indigenous or aboriginal languages”.

Thus, there are two good reasons why linguistic heritage language research has focussed almost exclusively on the languages of immigration: firstly, practical reasons, namely the need to determine the baseline, which could be better achieved by studying non-dominant languages that had not only multilingual but also monolingual speakers. And secondly, these are hegemonic issues: In the case of allochthonous languages, the labelling power lies unchallenged with the investigators and not with the investigated, unlike in the case of autochthonous languages.

## 2.5 New developments in heritage language research

More recently, however, there have been two interesting new strands in heritage language research, and the reception of these new approaches raises the question of whether the term as a whole does not obscure more than advance multilingualism and language contact research.

On the one hand, voices are becoming louder that fundamentally question the meaningfulness of the comparison between monolingual and bilingual speakers. In principle, this message is already contained in the famous title “Neurolinguists, beware! The bilingual is not two monolinguals in one person”, where Grosjean (1989) argues that the complementary distribution of linguistic repertoires in multilinguals between their languages makes competence comparisons between monolinguals and bilinguals impossible, in principle. At the same time, corpus linguistic studies make it clear that the individual speaker variation in monolinguals is no less large than that in bilinguals (Shadrova et al. 2021), and this makes the idea of a monolingual “control group” problematic – and thus also that of the “baseline” (see above). Rothman et al. (2022, 2023) accordingly argue that heritage language research should leave the monolingual speaker aside and instead focus more on subgroups of bilingual speakers in order to work out the significance of individual factors (input, access to literacy, etc.) for multilingual language development.<sup>6</sup>

Another new approach to heritage language research moves away from the deficit-orientated approach that has dominated heritage language research for some time with terms such as attrition, incomplete acquisition, erosion or simplification. Instead, this approach understands systematic phenomena in heritage languages, that differ from the monolingual variety of the respective languages where they are dominant languages, as instances of language change through language contact. These can and should be analysed using the ‘toolbox’ of language contact research – albeit with an important extension that takes up the special sociolinguistic situation of heritage languages as non-dominant languages in the larger society: It is assumed that the varieties that arise in the language of heritage speakers have an important source in the limited access of heritage speakers to the structures of the written norm of the respective language and to the registers of formality in

---

<sup>6</sup> This approach of detailed investigation of input factors is also interestingly extended by recent studies on bilingual speakers of different ages, one of their languages being a non-dominant language, who immigrate to the country where this language is the dominant language. We are dealing with a (re)activation of language knowledge, the investigation of which can be an important contribution to multilingualism research, especially with regard to the central question in second language acquisition research about the role of age in the acquisition of another language (cf. Antonova-Unlu and Bayram 2023).

this language. As Wiese et al. (2022) argue, this can lead to a “levelling” between the linguistic practices of the heritage language used in different communicative situations (Wiese et al. 2022).

## 2.6 Do we (still) need the term “heritage language”?

Considering what was said above, the term “heritage language” appears as problematic. For one thing, it is problematic in itself, as the semantics of “heritage” go in the wrong direction. Secondly, the term initially implies that all languages are meant that are in a non-dominant sociolinguistic constellation with the respective majority language. In research practice, however, we see that specific non-dominant languages are meant, namely those that have come into this sociolinguistic situation through migration. For the linguistic research questions associated with heritage language research, however, such a distinction cannot be constitutive.

Especially if we consider the more recent trends in heritage language research mentioned above, it becomes clear that heritage language research, with its specific sociolinguistic orientation, contributes to two fields of research. One is cognitively orientated bi- and multilingualism research, and the other is language contact research. It sensitises both fields of research to the challenge of investigating dominance relationships between languages. The term “heritage language”, however, implies an independent field of research – but this is not the case. Rather, it is counterproductive to stick with this term, as it isolates heritage language research and thus blocks the view of the enrichment that research on the development of non-dominant languages brings or could bring for the aforementioned fields of research.

In the following, I will use the term “non-dominant language” instead of “heritage language”, in order to stress the societal status of these languages, which appears to be the crucial feature with regard to the contribution research on these languages makes to the wider field of language contact research. The attributes which constitute the heritage language definition beside the non-dominance remain untouched (nativeness of the language, early bilingualism of the speakers).

## 3 Heritage language research enriching language contact research

Now that we have (hopefully convincingly) successfully “deconstructed” the term heritage language, I would like to give three examples from research on Turkish as a non-dominant language in Germany and the USA and put them in context with

other language contact literature on Turkish. In this way, I hope to show that and how this research, freed from the “straitjacket” of the heritage concept, contributes to language contact research.

### 3.1 The data

The research reported in the following comes from a large-scale, cross-linguistic investigation carried out within the context of the Research Unit “Emerging Grammars in Language-Contact-Situations: A Comparative View” (short ‘RUEG’; <https://www.linguistik.hu-berlin.de/en/rueg>)<sup>7</sup>. In the investigation, monolingual speakers of English, German, Russian, Greek and Turkish in those countries, where these languages are the dominant languages (U.S., Germany, Russia, Greece and Turkey, respectively) and bilingual speakers of the respective languages in Germany and the U.S., as well as Turkish-Kurdish (Kurmancî) bilingual speakers from Turkey and Turkish-Kurdish-German trilingual speakers in Germany and bilingual German-English speakers in the U.S. were investigated. Bilinguals and monolinguals were approached as two speaker groups to be investigated, rather than experimental vs. control group. All bilingual speakers were born in the country of the respective dominant language or had arrived there at an early age.<sup>8</sup> In all categories, two age groups were covered: adolescents (14–18 years), and adults (22–35 years), and roughly 30 speakers from each monolingual and bilingual age group participated.<sup>9</sup>

RUEG based its research upon production data and in order to be able to capture broader repertoires of the speakers, RUEG used the “Language Situations” (“LangSit”) set-up for elicitations (Wiese 2020). In this elicitation method, participants are familiarized with a fictional event (here: a film narrating a minor car accident) and asked to imagine themselves as a witness to this event, and then act out telling different interlocutors about it in different communicative situations, here:

---

7 Funded by Deutsche Forschungsgemeinschaft, DFG (FOR 2537). Speaker: Heike Wiese; further PIs: Artemis Alexiadou, Shanley Allen, Oliver Bunk, Natalia Gagarina, Mareike Keller, Anke Lüdeling, Judith Purkarthofer, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Sabine Zerbian; postdoc: Kalliopi Katsika; PhDs: Katerina Iefremenko, Esther Jahns, Martin Klotz, Thomas Krause, Annika Labrenz, Maria Martynova, Katrin Neuhaus, Tatiana Pashkova, Vicky Rizou, Wintai Tsehaye, Yulia Zuban.

8 In general before the age of 36 months, although in some cases this was extended (up to 48 months) where otherwise it would not have been possible to recruit enough speakers.

9 Kurdish (Kurmancî) speakers (bilingual and trilingual) only belonged to the adult group.

- 1) leave a voice message for a friend, via instant messenger (informal-spoken);
- 2) write a message to a friend, via instant messenger (informal-written);
- 3) leave a voice message on a police “witness line” (formal-spoken);
- 4) write a witness report for the police (formal-written).

At the end of data elicitation, participants were asked to fill in a sociolinguistic questionnaire on biographical data including language use and personality traits.

Bilingual speakers were recorded twice, in their non-dominant language and in the majority language, with the two sessions at least three days apart. Monolingual speakers were recorded once, in the majority language.<sup>10</sup>

For Turkish, then, just as for the other languages, the elicitations thus yielded matched elicited, semi-spontaneous data across registers, contact-linguistic settings, and bilingual and monolingual speaker groups in two age groups.

Research on Turkish in the framework of RUEG concentrated on two grammatical domains, namely on the postverbal position, and on clause combining phenomena. We concentrate here on the latter, not only, but also, because it contributes to Thomas Stolz’ research.<sup>11</sup>

## 3.2 Aspects of clause combining in Turkish as a non-dominant language in the U.S. and in Germany

### 3.2.1 Example 1: Frequency changes in clause combining

The first example I discuss is that of general frequencies of finite and non-finite hypotactic vs. paratactic clause combining. As is well known, the SOV-language Turkish is a language which strongly relies on non-finite subordination, where a clause is headed by a clause-final verbal form which combines with a subordinator in the form of a suffix. The language has three main strategies of nonfinite subordination: (i) complement clauses which are clausal nominalizations; (ii) preposed relative clauses headed by participles; and (iii) adverbial clauses headed by *converbs* (Kornfilt 1997; Göksel and Kerslake 2005) or nominalizations in combination with postpositions.

---

<sup>10</sup> All data is accessible through the online corpora Lüdeling et al. (2024) and Iefremenko, Klotz and Schroeder (2024).

<sup>11</sup> Allow me to stress that all the research I address here comes from collaborate publications, where the author of this chapter is involved. I am grateful to my colleagues for allowing me to use the examples and discuss the joint findings.

Previous studies on Turkish as a non-dominant language in Germany and the Netherlands produce a clear picture, at first appearance: In comparison with the monolingual setting, finite means of clause combining are more frequent in Turkish as a non-dominant language in Germany or the Netherlands, be these subordinating and/or paratactic combinations, while non-finite means of clause combining are less frequent (Dollnick 2013, Bayram 2013, Treffers-Daller et al. 2006). The named studies discuss this tendency as “incomplete acquisition”, in a pair with contact-induced convergence. That is, the dominance of the majority language minimalizes the input of Turkish, respectively, and the acquisition of the language remains incomplete. In addition, the contact language (German and Dutch, respectively), where non-finite subordination plays but a minor role and finite subordination is the preferred means in hypotaxis, triggers Turkish to “steer away”, so to speak, from non-finite subordination.

In a recent publication from the RUEG-group, Özsoy et al. (2022) as well as Iefremenko et al. (2024) partly confirm the frequency relations: In terms of overall frequencies, native speakers of Turkish in Germany and in the U.S. indeed use fewer non-finite clausal structures compared to monolinguals in Turkey, but instead prefer paratactic clause combining strategies. However, this does not seem to account for converbs in the language use of native speakers of Turkish in the U.S., which show similar frequencies like the monolingual speakers in Turkey. Also, the authors show that the frequency differences between bilingual and monolingual varieties of Turkish are observed only in formal communicative situations: When bilingual speakers are asked to perform in the formal communicative situation of reporting the traffic accident they have witnessed to the police, they use significantly less non-finite subordinating constructions than the monolingual speakers from Turkey in the same communicative situation. On the other hand, the difference between monolingual and bilingual speakers dissolves when it comes to informal communicative situations. Here, also monolingual speakers produce significantly less non-finite structures than in the formal communicative situations. And the frequencies of non-finite subordinations are more or less alike between monolingual and bilingual speakers, in the informal communicative situations.

Non-finite clause combining thus reveals itself as the preferred means of clause combining in the formal register(s) of monolingual Turkish in Turkey (see also Schroeder 2002), while in the informal register(s), its role within the domain of clause combining is less prominent. The reduced frequency of non-finite subordination, which is attested in all investigations of Turkish as a non-dominant language in contact with German, Dutch and English, thus reveals itself neither as an instance of “incomplete acquisition” nor as triggered by contact with the majority language(s). It rather is an instance of “internal change”: Informal features of a language, that are present in the monolingual variety, spread and consolidate in the

specific contact situation of the same language when it is a non-dominant language (see also Wiese et al. 2022 as well as Schroeder et al. accepted and Iefremenko and Schroeder 2024).

### 3.2.2 Example 2: Frequency and qualitative changes in converbs

As mentioned above, adverbial subordination in Turkish is mainly carried out by means of clauses headed by converbs. Following Johanson (1995) the canonical system of Turkish converbs can be described along three parameters, namely

- (i) the aspectual relation between the converb clause and the superordinate clause,
- (ii) the modificational relation which the converb clause exerts on the superordinate clause, and
- (iii) the subject relation between the converb and the superordinate clause.

With respect to (i), the aspectual relation can be

- either intra-terminal, that is it expresses the event occurring within the limits of the event in the superordinate clause,
- or it can be post-terminal, where the event is expressed as having begun before the event in the main clause unfolds,
- or it can be terminal, where the event is expressed as coming to an end before the event in the superordinate clause has begun.

However, as Johanson (1995: 319) stresses, it is “not unusual for aspectual units to combine or to vacillate between post-terminality and terminality”.

As for (ii), the converb clause can

- either modify the superordinate clause in the sense that it provides further information about the event, its purpose or cause, the conditions under which it occurred, the degree, manner and means of realizations,
- or it can be non-modifying, i.e. it simply expresses a sequentiality of events.

Finally, with respect to (iii), there are some converb forms which clearly prefer same-subject relations between the converb clause and the superordinate clause, allowing for exceptions only when the subject of the latter is in a possessor-posses-sum-relationship with the subject of the former, while some other converbs freely allow varying subject relations between the converb and the superordinate clause. Iefremenko et al. (2021: 135) summarize these relations with the following table (Table 1).

**Table 1:** Features of converbs in canonical Turkish. Parentheses indicate a marginal and more restricted use.

Converb-form	Aspectual relation		Modificational relation		Subject relation	
	Intra-terminal	(Post-)terminal	Modifying	Non-modifying	Same-subject	Varying subject
-(y)Ip		x		x	x	(x)
-(y)IncA		x	x	x		x
-(y)ArAk	x	(x)	x	(x)	x	(x)
-(y)ken	x	(x)	x	x		x
-DIğIndA	x		x	x		x
-mAdAn		x	x	x		x

As for converbs in Turkish(es) in a non-dominant sociolinguistic constellation, a study by Turan et al. (2020) on Turkish in Germany supports converbs to be in line with the more general results for nonfinite subordination in Turkish in Germany as given in the previous section. The study concludes that bilingual speakers of Turkish in Germany use significantly fewer converbs than monolinguals (in Turkey) and instead use a higher number of finite forms. Turan et al. (2020) also note what they call “unconventional uses” of the converb *-(y)Ip* by bilingual speakers, and they base this on the assessment of monolingual judges, which were asked to consider appropriateness. These uses pertain to issues of coreferentiality, aspectuality and coordination. That is, the bilingual speakers in their study used different subjects with the verb in the main clause and with the converb; moreover, they used this converb in aspectual functions which differed from monolingual Turkish, or they connected the converb to the main clause by means of the coordinate conjunct *ve* ‘and’.

In the RUEG-context, Iefremenko et al. (2021) present four findings which partly confirm Turan et al. (2020), but also go beyond: First, as already stated above, lower frequencies are only found with bilingual speakers in Germany, but not so in the U.S. Second, the converbs *-(y)Ip*, *-(y)ArAk* and *-(y)ken* in bilingual Turkish appear to extend in functionality. To be more specific, these converbs can also be modifying (post-)terminal and express succession of events with the emphasis on the causality of the action, i.e. the function that in canonical Turkish is expressed exclusively by the converb *-(y)IncA*. The third finding concerns the phenomenon of coreferentiality: Like Turan et al. (2020), Iefremenko et al. (2021) find examples, where bilingual speakers use the converbs *-(y)Ip* and *-(y)ArAk* with different subjects in superordinate and converb, and the two subjects are not in a possessor-possessum relationship. However, these are always examples, where these forms are used outside of their canonical aspectual and (non-)modifying relation. The

fourth finding concerns systematicity across groups: While these non-canonical converb uses rather appear to be matters of individual variation with the speakers in the U.S. and the adult speakers in Germany, they appear to be more systematic with adolescent speakers in Germany.

Turan et al. (2020) are rather ‘helpless’, one may say, when it comes to the interpretation of the phenomena they found. They consider the decrease of frequency of converbs in bilingual Turkish in Germany as a process of “grammaticalization” (2020: 9), without being able to say in which direction this goes, and a “vulnerability” of syntax to “foreign Influence” (2020: 10) concerning non-canonical uses, again remaining nebulous with regard to the details of this “foreign influence”.

To me, this helplessness is another indicator of heritage language research remaining very much within its own limits and not situating their research into a wider perspective of studies on language contact and language change, and not considering the specific sociolinguistic situation of non-dominant languages within this.

For once, we can refer back to what was said in the sub-section above about frequency relations: These are strongly related to issues of register and, thus, rather an instance of internal change (as for the direction of this change, more will be said in the following sub-section). In addition, the wider perspective of Iefremenko et al. (2021), which not only look at bilingual Turkish in Germany, but also in the U.S., allows to be clearer with regard to “foreign influence”: With the highly productive *-ing* form, English has a converb with a much higher frequency and broader functionality than comparable forms in German (Iefremenko et al. 2021: 138 based on König and Gast 2018: 71–72). Thus, it is reasonable to assume that the presence of the English converb also stabilises the use of converbs in Turkish among Turkish-English bilinguals, hence the higher frequency of converbs in Turkish in the U.S. compared to Turkish in Germany.

Furthermore, both Turan et al. (2020) as well as Iefremenko et al. (2021) find “unconventional uses” of converbs in terms of aspectual, modifying and subject relations. However, Turan et al. (2020) miss out that these uses are interdependent, namely that they are an outcome of a general broadening of functions of converbs in Turkish in the non-dominant constellation. “Unconventional” different-subject relations with *-(y)Ip* and *-(y)ArAk*-converbs only occur where these forms are used outside of their canonical aspectual and (non-)modifying relation. Furthermore, these uses are not a completely new development, but they are an extension of a limited freedom of varying subject relations with these converbs when the subjects are in a possessor-possessum-relation already in canonical (Turkey’s) Turkish, and in this, these subject relations adhere to a continuum noted for converbs in Turkic languages in general (Bárány and Nikolaeva 2020).

Lastly, Iefremenko et al. (2021) show that the “unconventional uses” of the *converbs* in non-dominant native Turkish are most systematic in the group of adolescent speakers in Germany. Here, sociolinguistic considerations are in order. As Iefremenko et al. (2021: 133) point out, Turkish is much more vital in Germany than it is in the U.S. The Turkish community in Germany is large and sociologically comparatively tightly-knit, especially in urban areas. In contrast, the Turkish community in the U.S. is not only much smaller in number, but is also scattered throughout the much bigger country. As a consequence of this demoscopic, sociological and geographical situation, Turkish is used in more social domains in Germany than in the U.S., i.e. people use it in informal public domains (markets, cafes, shops, during leisure times, etc.), and it is a vital language in adolescent peer groups, in Germany. In variationist sociolinguistics, adolescents are identified as a central group for language change (Tagliamonte 2016) – given, of course, a situation, where the language at issue is actually used in the peer groups. This is arguably much more likely the case in Germany than in the U.S., hence the stronger systematicity of non-canonical uses in the group of bilingual Turkish-German adolescents in Germany may well be linked to a sociolinguistic dynamic.

### 3.2.3 Example 3: Qualitative changes in clause combining

The third example I want to discuss has as a starting point the paratactic clause combining strategy which Turkish varieties both in Germany as well as in the U.S. seem to prefer. Above, I identified this phenomenon as internal dynamics, where informal features of a language, that are present in the monolingual canonical variety, spread and consolidate in the specific contact situation of the same language when it is a *nnpn*-dominant language.

Paratactic clause combining brings with it a higher functional load on connectors, which organize the semantic and pragmatic relationship between clauses, so to speak ‘in the place of’ the various means of hierarchic syntactic organization which characterize hypotactic structuring (Givon 1979; Miller and Weinert 2009). As Schroeder (2002), Kerslake (2007) and Schroeder (2016) show, in the informal registers of (Turkey’s) Turkish, the use of paratactic connectors not only increases in comparison to formal language use, but there exist also particular forms which belong almost exclusively to this register. Özsoy et al. (2022) as well as Iefremenko et al. (2024) show how this tendency continues in non-dominant Turkish in Germany and the U.S. The bilingual speakers not only use more paratactic connectors than monolingual speakers, but it appears that “the general shift to more paratactic structuring seems to trigger an expansion of the function of specific connectors (. . .) and temporal adverbs (. . .) into more general clause-connecting devices” (Özsoy

et al. 2022: 17). Thus, the higher functional load on paratactic clause combining in bilingual Turkish in Germany and the U.S., as opposed to monolingual Turkish, trigger an expansion of the part of speech of connectors in Turkish.

However, the story does not end here, and its continuation brings us to one core term in language contact, namely, convergence.

Studies on Turkic languages in contact with Indo-European show how and that Turkic languages incorporate Indo-European-type subordination strategies, that is, finite clauses with a clause-initial subordinating connector, which is a free morpheme. Balkan Turkic and Karaim in contact with Slavic languages and Khalaj, Uzbek, and Azeri in contact with Persian contact can be cited as examples, see Johanson (2021: Secs. 55.2.6, 55.3.8, 903–904, 913–916, 923–924) and further literature cited within.

In a detailed investigation of Balkan Turkic varieties Keskin (2023) shows how in contact with the Indo-European subordination strategies, Balkan Turkic varieties develop all kinds of different subordination strategies, which belong neither to the “Turkic type” nor to the “Indo-European type” of subordination, but are hybrid structures in-between. See the following three examples of such hybrid clauses from Rumelian Turkic: In (1), the relative clause is preposed, but it has a clause-initial free connector and is a finite clause. In (2), the finite relative clause is postposed and has a free subordinator, which is, however, clause-internal. And in (3), the adverbial (consecutive) subordination has an initial free subordinating connector, but the clause is non-finite (all examples cited from Keskin 2023: 161–162):

- (1) Rumelian Turkic (Razgrad, Bulgaria; Murtaza 2016: 81)

*[ani sırala-dı-m] to urba-lar*  
 CONN tell-PST-1PL DIST clothes-PL  
 ‘the clothes that I told you about’

- (2) Rumelian Turkic (Mireshe, Kosovo; Sulçevsi 2019: 255)

*o sene [bu cade ne yap-ıl-di]*  
 DIST year PROX road CONN make-PASS-PST.3SG  
 ‘the year that this road was built’

- (3) Rumelian Turkic (Razgrad, Bulgaria; Haliloğlu 2017: 214)

*Çuval doqū-du-lā onlā-dan [ani zāre-ler-i quy-mā].*  
 sack weave-PST-3PL 3PL-ABL CONN grain-PL-ACC put-NOM.DAT  
 ‘They weaved sacks from them to put the grains in.’

Keskin (2023) argues that these structures should be understood as instable “transition types” in the transition from the Turkic to the Indo-European type of subordination.

Now, similar structures have been noted also in non-dominant Turkish(es) in Europe, however, with the authors not really knowing what to do with them. Bohnacker and Karakoç (2020: 182) as well as Şan (2023: 13), for example, cite “non-standard” examples, where Swedish-Turkish bilingual children (in Bohnacker and Karakoç 2020) and the German-Turkish bilingual children (in Şan 2023) combine the (paratactic) causal coordinator *çünkü* with a non-finite adverbial clause. Also Turan et al. (2020: 8) consider as “unconventional use”, where bilingual speakers of Turkish in Germany connect the converb to the main clause by means of a coordinate conjunct.

In a study using the RUEG corpus, Keskin et al. (2024) concentrate on such structures and find an abundance of them in the texts from German-Turkish and English-Turkish bilinguals and adults from Germany and the U.S. Most frequent are the constructions already noted by Bohnacker and Karakoç (2020) and Şan (2023), where a paratactic connector introduces a non-finite clause, see (4).<sup>12</sup>

- (4) Turkish in Germany, bilingual speaker 05FT, informal-spoken  
*gerek-iyor-muş* [*çünkü* *bunlar-a* *şahit ol-duğ-um için*]  
 be necessary-PROG-EVID because these-DAT witness-NOM-1SG for  
 ‘It was necessary because I witnessed these.’

However, Keskin et al. (2024) also find (with Labrenz et al. (accepted)), an emergence of the use of discourse markers as subordinating connectors in the language use of the bilingual speakers, sometimes in combination with finite clauses, as in (5), sometimes with non-finite clauses, as in (6):

- (5) Turkish in Germany, bilingual speaker 18MT, informal-spoken  
*anla-dı-n* [*yani* *bi çift* *var-dı* *top-lu*]  
 understand-PST-2SG CONN/DM a couple exist-PST.3SG ball-COM  
 ‘You understood {that/well} there was a couple with a ball.’

- (6) Turkish, in the U.S., bilingual speaker 74FT, formal-spoken  
*top-la* [*işte* *gid-er-ken*] *oynu-yo-du*  
 ball=with CONN/DM go-AOR-CVB play-PROG-PST.3SG  
 ‘S/he was playing with the ball as he went.’

<sup>12</sup> All examples from the RUEG Corpus (Lüdeling et al. 2024).

Thus, in Turkish as a non-dominant language in Germany and in the U.S. we also find examples of “transient structures”, hybrid structures between the Turkish and the Indo-European type of subordination, instable as they may be. By way of relating structures of Turkish as a non-dominant language in contact in Germany and the U.S. to structures of Turkic languages in other contact scenarios, Keskin et al. (2024) succeed in reinterpreting a phenomenon within the framework of contact-induced dynamics. In the sole focus on the “heritage language”, these structures are either overlooked or somehow set aside as “non-standard” or “unconventional”.

## 4 Conclusion

The term “heritage language” describes a sociolinguistically specific constellation of language contact – a non-dominant language that exists in close contact with another language, which is the socially dominant language of the respective society. Speakers of this language are native speakers which become bilingual at the latest when they enter the societal institutions, where the dominant language prevails. The term, however is problematic, not only because of the semantics of the word “heritage” and the application of the term “heritage language” exclusively to migrant languages, but also because it creates the impression of some independent area of research. The term blurs the fact that research on these languages and their speakers contributes substantially to bilingualism research, and to language contact research. I propose to use the term “non-dominant language” instead, in order to highlight the particular sociolinguistic relation, which is an issue here (with the other attributes of the non-dominant language – native language, early bilingualism of the speakers prevailing). I have discussed three examples from our research on Turkish as a non-dominant language in contact with German and English. I argue that the examples show how acknowledging the specific sociolinguistic constellation of the non-dominant language Turkish in Western societies contributes to and benefits from language contact research.

## Abbreviations

ABL	ablative
ACC	accusative
AOR	aorist
COM	comitative
CONN	connector

CVB	converb
DAT	dative
DIST	deictic form (distal)
DM	discourse marker
EVID	evidential
NOM	nominalizer
PASS	passive
PL	plural
PROG	progressive
PROX	deictic form (proximal)
PST	past tense (preterite)
SG	singular

## References

- Antonova-Unlu, Elena, Fatih Bayram. 2023. The role of external factors on the reactivation of the heritage language of Turkish-German returnees. *Frontiers in Psychology* 14. 1156779. DOI: 10.3389/fpsyg.2023.1156779.
- Bárány, András, Irina Nikolaeva. 2020. Possessive and non-identity relations in Turkic switch-reference. *Studies in Language* 44 (3). 606–658. DOI: 10.1075/sl.19061.bar.
- Bayram, Fatih. 2013. *Acquisition of Turkish by heritage speakers. A Processability Approach*. PhD Dissertation. Newcastle University. URI: <http://hdl.handle.net/10443/1905>.
- Bohnacker, Ute & Birsel Karakoç. 2020. Subordination in children acquiring Turkish as a heritage language in Sweden. In Fatih Bayram (ed.), *Studies in Turkish as a Heritage Language*, 155–204. (Studies in Bilingualism 60). Amsterdam: Benjamins.
- Cummins, Jim. 1991. Introduction. *Canadian Modern Language Review* 47 (4). 601–605.
- Cummins, Jim. 2005. A proposal for action. Strategies for recognizing heritage language competence as a learning resource within the mainstream classroom. *Modern Language Journal* 89 (4). 585–592.
- Dollnick, Meral. 2013. *Konnektoren in türkischen und deutschen Texten bilingualer Schüler*. Frankfurt a.M.: Peter Lang.
- Ennsner-Kananen, Johanna & Kendall A. King. 2013. Heritage languages and language policy. In Carol A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*, 1–6. Wiley. DOI: 10.1002/9781405198431.wbeal0500.pub2.
- Gagarina, Natalia. 2014. Die Erstsprache bei Mehrsprachigen im Migrationskontext. In Solveig Chilla & Stefanie Haberzettl (eds.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen. Mehrsprachigkeit. 1. Auflage*, 19–37. München: Urban & Fischer.
- Givón, Talmy. 1979. *On understanding grammar* (Perspectives in neurolinguistics and psycholinguistics). New York: Academic Press.
- Göksel, Asli & Celia Kerslake. 2005. *Turkish. A comprehensive grammar* (Routledge comprehensive grammars). London: Routledge.
- Grosjean, François. 1989. Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language* 36 (1). 3–15. DOI: 10.1016/0093-934X(89)90048-5.

- Haliloğlu, Neşe. 2017. *Zavet köyü Türk ağzı* [The Turkish dialect of Zavet village]. Ahi Evran University, Kırşehir. MA thesis.
- Iefremenko, Kateryna, Martin Klotz & Christoph Schroeder. 2024. *RUEG subcorpus of Kurmanji Kurdish and Turkish data*. (0.1). Data set. Zenodo. DOI: 10.5281/zenodo.10810768
- Iefremenko, Kateryna & Christoph Schroeder. 2024. Language Contact. Preprint. *Reference Module in Social Sciences*. DOI: 10.1016/B978-0-323-95504-1.00134-4.
- Iefremenko, Kateryna, Onur Özsoy & Christoph Schroeder. 2024. Paratactic clause combining in heritage Turkish in Germany and the U.S. In Fatma Hülya, Özcan Önder, Tuncay Karalık, Bayram Çibik, İlnur Civan Biçer & Samet Deniz (eds.), *Selected Essays on Turkish Linguistics. The Anadolu Meeting. Proceedings of the 20th International Conference on Turkish Linguistics*, 303–321. (Turcologica 127). Wiesbaden: Harrassowitz.
- Iefremenko, Kateryna, Christoph Schroeder & Jaklin Kornfilt. 2021. Converbs in heritage Turkish: A contrastive approach. *Nordic Journal of Linguistics* 44(2). 130–154. DOI: 10.1017/S0332586521000160.
- Johanson, Lars. 1995. On Turkish converb clauses. In Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective. Structure and meaning of adverbial verb forms-- adverbial participles, gerunds*, 313–348. (Empirical Approaches to Language Typology 13). Berlin/New York: Mouton de Gruyter.
- Johanson, Lars. 2021. *Turkic*. (Cambridge Language Surveys). Cambridge/New York: Cambridge University Press.
- Kerslake, Celia. 2007. Alternative subordination strategies in Turkish. In Jochen Rehbein, Christiane Hohenstein & Lukas Pietsch (eds.), *Connectivity in Grammar and Discourse*, 231–258. (Hamburg Studies on Multilingualism 5). Amsterdam/Philadelphia: Benjamins.
- Keskin, Cem. 2023. Transient subordinate clauses in Balkan Turkic in its shift to Standard Average European subordination. Dialectal and historical evidence. *Folia Linguistica* 57(s44-s1). 155–197. DOI: 10.1515/flin-2023-2001.
- Keskin, Cem, Kateryna Iefremenko, Jaklin Kornfilt & Christoph Schroeder. 2024. Hybrid clause combining strategies in Turkish language contacts. *Studia Linguistica*, early view, 1–44. DOI: 10.1111/stul.12256.
- König, Ekkehard & Volker Gast. 2018. *Understanding English-German contrasts*, 4th edn. (ESV Basics 29). Berlin: Erich Schmidt Verlag.
- Kornfilt, Jaklin. 1997. *Turkish*. (Descriptive Grammars). London: Routledge.
- Labrenz, Annika, Kateryna Iefremenko, Kalliopi Katsika, Christoph Schroeder, Heike Wiese. accepted. Dynamics of discourse markers in language contact. In Mareike Keller, Shanley Allen, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers*. (Current Issues in Bilingualism). Berlin: Language Science Press.
- Lüdeling, Anke; Artemis Alexiadou, Shanley Allen, Oliver Bunk, Natalia Gagarina, Sofia Grigoriadou, Rahel Gajaneh Hartz, Kateryna Iefremenko, Esther Jahns, Kalliopi Katsika, Mareike Keller, Martin Klotz, Thomas Krause, Annika Labrenz; Maria Martynova, Onur Özsoy, Tatiana Pashkova, Maria Pohle, Judith Purkharthofer, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehaye, Heike Wiese, Sabine Zerbian, Yulia Zuban & Nadine Zürn. 2024. *RUEG Corpus*. Data set. Zenodo. DOI: 10.5281/zenodo.11234583.
- McCarty, Teresa L. 2008. Native American languages as heritage mother tongues. *Language, Culture and Curriculum* 21 (3). 201–225. DOI: 10.1080/07908310802385881.
- Miller, James Edward & Regina Weinert. 2009. *Spontaneous spoken language. Syntax and discourse*. Oxford/New York: Oxford University Press.

- Mulík, Stanislav, Mark Amengual, Ricardo Maldonado & Haydée Carrasco-Ortíz. 2022. Hablantes de herencia: ¿una noción aplicable para los indígenas de México? *ELA* 73. 7–37. DOI: 10.22201/enallt.01852647p.2021.73.970.
- Murtaza, Durgyyul Metin. 2016. *Razgrad ili Kubrat ilçesi Alevi-Bektaşlı köyleri ağızı* [The dialect of the Alevi-Bektashi villages in Razgrad province Kubrat county]. Trakya University, Edirne. MA thesis.
- Özsoy, Onur, Kateryna Iefremenko & Christoph Schroeder. 2022. Shifting and expanding clause combining strategies in heritage Turkish varieties. *Languages* 7(3). 242. DOI: 10.3390/languages7030242.
- Polinsky, Maria. 1995. Cross-linguistic parallels in language loss. *Southwest Journal of Linguistics* 14. 88–123.
- Polinsky, Maria, Asya Pereltsvaig. 2003. Book Review: E.A. Zemskaja (Ed.) *Jazyk russkogo zarubež'ja* [The Language of the Russian Diaspora]. Moscow/Vienna: Wiener slawistischer Almanach, 2001. (Wiener slawistischer Almanach 53.). *Heritage Language Journal* 1(1). 135–143. DOI: 10.46538/hlj.1.1.7.
- Rothman, Jason. 2009. Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism* 13(2). 155–163. DOI: 10.1177/1367006909339814.
- Rothman, Jason, Fatih Bayram, Vincent DeLuca, Grazia Di Pisa, Jon Andoni Duñabeitia, Khadij Gharibi, Jiuzhou Hao, Nadine Kolb, Maki Kubota, Tanja Kupisch, Tim Laméris, Alicia Luque, Brechje van Osch, Sergio Miguel Pereira Soares, Yanina Prystauka, Deniz Tat, Aleksandra Tomić, Tom Voits & Stefanie Wulff. 2022. Monolingual comparative normativity in bilingualism research is out of “control”: Arguments and alternatives. *Applied Psycholinguistics* 44(3). 1–14. DOI: 10.1017/S0142716422000315.
- Rothman, Jason, Fatih Bayram, Vincent DeLuca, Jorge González Alonso, Maki Kubota & Eloi Puig-Mayenco. 2023. Defining bilingualism as a continuum. In Gigi Luk, John A. E. Anderson & John G. Grundy (eds.), *Understanding language and cognition through bilingualism. In honor of Ellen Bialystok*, 38–67. (Studies in Bilingualism, Volume 64). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Şan, Nebiye Hilal. 2023. Subordination in Turkish heritage children with and without developmental language impairment. *Languages* 8(4). 239. DOI: 10.3390/languages8040239.
- Schroeder, Christoph. 2002. On the structure of spoken Turkish. *Essener Linguistische Skripte– elektronisch* 2. 73–90. [https://www.uni-due.de/imperia/md/content/elise/ausgabe\\_1\\_2002\\_schroeder.pdf](https://www.uni-due.de/imperia/md/content/elise/ausgabe_1_2002_schroeder.pdf) (accessed 6th March 2025)
- Schroeder, Christoph. 2016. Clause combining in Turkish as a minority language in Germany. In Didar Akar, Mine Güven, Meltem Keleşir & Balkız Öztürk (eds.), *Exploring the Turkish linguistic landscape. Essays in honor of Eser Erguvanlı-Taylan*, 81–102. (Studies in language companion series 175). Amsterdam/Philadelphia: Benjamins.
- Schroeder, Christoph, Kateryna Iefremenko, Kalliopi Katsika, Annika Labrenz & Shanley Allen. accepted. Clause combining in narrative discourse: A contrastive approach. In Mareike Keller, Shanley Allen, Artemis Alexiadou & Heike Wiese (eds.): *Linguistic dynamics in heritage speakers*. (Current Issues in Bilingualism). Berlin: Language Science Press.
- Shadrova, Anna, Pia Linscheid, Julia Lukassek, Anke Lüdeling & Sarah Schneider. 2021. A Challenge for contrastive L1/L2 corpus studies. Large inter- and intra-individual variation across morphological, but not global syntactic categories in task-based corpus data of a homogeneous L1 German group. *Frontiers in Psychology* 12. 716485. DOI: 10.3389/fpsyg.2021.716485.
- Shah, Sheena, Theresa Biberauer & Erika Herrmann. 2024. *Sixth-generation contact German in South Africa. The case of Kroondal German*. London: EL Publishing.

- Sulçevsi, İsa. 2019. *Kosova Türkçesi ve fiil yapıları* [Kosovar Turkish and its verb forms]. Istanbul University. PhD diss.
- Tagliamonte, Sali. 2016. *Teen talk: The language of adolescents*. Cambridge: Cambridge University Press.
- Tracy, Rosemarie. 2014. Mehrsprachigkeit: Vom Störfall zum Glücksfall. In Manfred Krifka, Joanna Błaszczak, Annette Leßmöllmann, André Meinunger, Barbara Stiebels, Rosemarie Tracy & Hubert Truckenbrodt (eds.), *Das mehrsprachige Klassenzimmer*, 13–33. Berlin/Heidelberg: Springer.
- Treffers-Daller, Jeanine, Sumru A. Özsoy & Roeland van Hout. 2006. Oral language proficiency of Turkish-German bilinguals in Germany and Turkey: An analysis of complex embeddings in Turkish picture descriptions. In Hendrik E. Boeschoten & Lars Johanson (eds.), with the editorial assistance of Sevgi Ağacagül and Vildan Milani. *Turkic languages in contact*, 203–219. (Turcologica 61). Wiesbaden: Harrassowitz.
- Turan, Dilek, Elena Antonova-Ünlü, Çiğdem Sağın-Şimşek, Mehmet Akkuş. 2020. Looking for contact-induced language change: Converbs in heritage Turkish. *International Journal of Bilingualism* 24 (5–6). 1035–1048. DOI: 10.1177/1367006920926263.
- United Nations Educational, Scientific and Cultural Organization. 2001. *International Round Table “Intangible Cultural Heritage” – Working definitions. Piedmont, Italy, 14 to 17 March 2001. Consultation with member states on “Identification (inventory) and documentation of folklore”*. UNESCO. New York. <https://ich.unesco.org/doc/src/05303.pdf> (accessed 29th November 2024)
- Wiese, Heike. 2020. Language situations. A method for capturing variation within speakers’ repertoires. In Yoshiyuki Asahi (ed.), *Proceedings of Methods XVI. Papers from the Sixteenth International Conference on Methods in Dialectology 2017*, 105–117. (Bamberger Beiträge Zur Englischen Sprachwissenschaft / Bamberg Studies in English Linguistics 59). Frankfurt a.M.: Peter Lang.
- Wiese, Heike, Artemis Alexiadou, Shanley Allen, Oliver Bunk, Natalia Gagarina, Kateryna Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehay, Sabine Zerbian & Yulia Zuban. 2022. Heritage speakers as part of the native language continuum. *Frontiers in Psychology* 12. 717973. DOI: 10.3389/fpsyg.2021.717973.



Livio Gaeta

# Remodeling inflectional classes: Strong and weak verbs in Walser German

**Abstract:** In Titsch, a variety of Walser German spoken in Gressoney, a linguistic island of northern Italy, the Germanic strong / weak verb classes were remodeled in a very peculiar way. While the weak classes have been partially generalized in compliance with a trend which is generally observed throughout the whole Germanic family, the strong verbs have been reorganized according to the morphosyntactic environment in which they are used. When the latter require contextual agreement, the strong form is selected, while the weak form appears when no agreement is required by the context.

**Keywords:** inflection, language change, corpus linguistics, linguistic islands, endangered languages

## 1 Introduction

The label ‘Walser German’ (= WG) identifies a group of dialects belonging to the Highest Alemannic branch of Upper German, originally spoken in the most south-western province of Switzerland, the *Wallis* ‘the (Rhone) Valley’, whence *Wal(li)ser*. At the outset of the last Millennium groups of settlers left the Wallis and migrated south- and eastwards in search of better conditions for life. They founded villages on the higher segments of the alpine valleys characterized by a common architectural landscape – hallmarked by the *Städl*, the typical Walser house made of wood and stone – and close contacts with the native homeland (cf. Rizzi 1993). A number of villages were also founded on the south side of the Monte Rosa massif, in which they were in contact with the local Romance-speaking population. In particular, the village of Gressoney – consisting of the two settlements of La Trinité and Saint Jean – lies immediately below the Monte Rosa on the upper side of the valley traversed by the river Lys, immediately followed by the Romance-speaking village of Gaby. After centuries of close contacts with the homeland as well as with the rest of Switzerland and of southern Germany, the Italian unification and the First World War meant for Gressoney sinking into a condition of linguistic island which reached its extreme point during the repressive policy of the Fascist regime. This is an important premise for understanding the present situation, because this segre-

---

Livio Gaeta, University of Turin, e-mail: livio.gaeta@unito.it

gation characterized the last generation who grew up in a thorough German-speaking environment but neatly separated from any contact with the ancestral homeland and lies behind the striking language changes which will be discussed below. After 1945, the status of linguistic minority could not stop its assimilation to the Italian surroundings also because of enhanced migrations and mixed marriages. In contrast to the other German minorities of north-eastern Italy, the population was never educated in the Standard German variety, which is not mutually intelligible with the local variety. This stands in neat contrast with the situation found in the Swiss homeland, where Standard German – or its Swiss version, the so-called *Schweizerdeutsch* or *Schriftdeutsch* – serves as written code. Nowadays, Gressoney as well as the other Walser islands in north-western Italy are losing their linguistic identity with the last speakers of the Walser German variety mostly using the other varieties of their repertoire, namely Piedmontese, Standard Italian and French, although the process of language shift is not yet completed. A number of projects have been funded to rescue this identity, to collect written and oral documents and to make it available for the future generations as well as for the research. Thanks to these projects, the data presented in this and in the following sections could be collected into an archive and carefully analyzed.

## 2 Strong and weak verbs in Walser German

Lying at the southern edge of the Upper German area, Walser German dialects are traditionally known for their conservative character typical of such marginal areas (cf. Bohnenberger 1913; Russ 1990: 367), and this holds true also for the islands found in Italy (cf. Zürner 2011; Eufe and Mader 2018). This is for instance reflected in the preservation of distinct weak inflectional classes (= ICs) as they were witnessed in Old High German (= OHG) in neat contrast to all other dialectal varieties found in Germany and Switzerland (see Russ 1990: 383 for a survey). As we will also see in the case of the preservation of adjective agreement, the role of language contact with the surrounding Romance languages should not be forgotten which surely acted as a relevant factor for certain conservative processes like for instance the retention of final unstressed vowels (cf. Moulton 1941: 39; Zürner 2011: 105). Clearly, such a retention lies at the heart of the distinction of three ICs in the verbal morphology of the WG variety spoken in Gressoney called Titsch which will be investigated in the rest of the paper, see Table 1.<sup>1</sup>

---

<sup>1</sup> I adopt here and in the rest of the paper the grammatical classification normally used in the dictionary edited by the Walser Kulturzentrum (cf. GrW 1998), which results from the speakers' collec-

**Table 1:** ICs of Titsch verbs.

	<b>Examples (infinitive / past participle)</b>
IC I	<i>bisse</i> ‘to itch, bite’ / <i>béssét</i> <i>fénne</i> ‘to find’ / <i>gfönnet</i> <i>abschliesse</i> ‘to close’ / <i>abgeschlossen</i> <i>schmelze</i> ‘to melt’ / <i>gschmolzen</i> <i>lache</i> ‘to laugh’ / <i>glachet</i>
IC II	<i>teile</i> ‘to divide’ / <i>teilt</i> <i>läbe</i> ‘to live’ / <i>gläbt</i> <i>chläbe</i> ‘to stick’ / <i>kläbt</i>
IC III	<i>moalò</i> ‘to paint’ / <i>gmoalòt</i> <i>rächnò</i> ‘to calculate’ / <i>grächnòt</i>

As a general feature, notice that as a consequence of the so-called *Präteritum-schwund* (‘preterite loss’) commonly found in southern German varieties (cf. Russ 1990: 377), all that is left of the root-vowel alternations typical of the Germanic Ablaut (= AB) class is only retained in the past participle (see Zürrer 1982: 94 for details). As will be further discussed below, because of the preterite loss the past participle has acquired a central role in Titsch for the expression of past tense by means of the perfect construction.

IC I hosts verbs belonging to the original strong macro-class (cf. OHG *bīzan* / *gibīzan*,<sup>2</sup> *findan* / *fundan*, *slīozan* / *gislozan* and *smelzan* / *gismolzan*) while the weak IC II and IC III go back etymologically to the OHG weak classes displaying respectively the thematic vowels (= ThVs) *-ja-* (already merged in OHG into *-e-*, cf. OHG *teilen* / *giteilit*) or *-ē-* (cf. OHG *lebēn* / *gilebēt* and *klebēn* / *giklebēt*) and *-ō-* (cf. OHG *mālōn* / *gimālōt*, *rehhēnōn* / *girehhēnōt*) and no AB alternation. In addition, IC I also hosts verbs going back to the etymological weak *-ē-* class (cf. OHG *lahhēn* / *gilahhēt*) which select the dental suffix *-et* to form the past participle reflecting the OHG suffix *-ēt* and is neatly distinct from the suffix *-t* selected by IC II which goes back to the OHG suffix *-it*. Note that verbs of the etymological *-ē-* class migrated to

---

tive enterprise and is established as the reference work in the Gressoney community and adopted also in Gaeta et al. (2024). By the same token, I also adopt the orthographic norms recommended therein in which <é>, <ä> and <ò> roughly correspond respectively to [i], [æ] and [u] while vowel sequences like <ie>, <éé>, etc. correspond to true (falling) diphthongs: [iɛ], [iɛ̃], etc. For a general description of the orthographic norms adopted in the community, see Gaeta et al. (2024). It must be added that the texts of our data-base do not always follow these orthographic norms, also because to a large extent they have been written before their adoption.

2 The OHG sign <ʒ(ʒ)> refers to a voiceless strident fricative [ʃ] normally resulting from a Germanic voiceless dental occlusive due to the so-called Second Consonant Shift (cf. Braune / Reiffenstein 2004: 167).

IC II display the suffix *-t* in the past participle as shown by *gläbt* and *kläbt*.<sup>3</sup> On the other hand, the etymological strong verbs of IC I swing between the older nasal suffix of the strong macro-class as in *abgeschlossen* and *gשמולזן*<sup>4</sup> and the suffix *-et* of the etymological *-ē-* class found in *bésset* and *gfönnet*. This variation is fairly widespread: *schribe* ‘to write’ / *gשמרעבן* or *gשמרעבט*, *usstéerbe* ‘to die out’ / *usgstörben* or *usgstörbet*, *vergässe* ‘to forget’ / *vergässen* or *vergässet*, *verliere* ‘to lose’ / *verlören* or *verlòret*, etc.

In this connection, it has to be stressed that the tendency towards the extension of the weak suffix *-et* to the etymological strong verbs does not imply the levelling of the AB pattern, which is generally preserved, apart from some marginal examples where the levelled alternative is found as in *verschwénde* ‘to disappear’ / *verschwòndet* or *verschwéndet*, *wäbe* ‘to weave’ / *gwòbet* or *gwäbt*, etc. We record at least four AB types within IC I, regardless of the weak or strong suffix selected, see Table 2.<sup>5</sup>

**Table 2:** AB types in IC I.

AB types	Examples
AB-1: <i>i / é</i>	<i>blibe</i> ‘to remain’ / <i>blébet</i> or <i>blében</i> , <i>ritte</i> ‘to ride’ / <i>gréttet</i>
AB-2: <i>ie / o</i>	<i>biete</i> ‘to offer’ / <i>bottet</i> , <i>siede</i> ‘to seethe’ / <i>gsotten</i> or <i>gsottet</i>
AB-3: <i>é / ò</i>	<i>bénne</i> ‘to bind’ / <i>bònnet</i> , <i>stéerbe</i> ‘to die’ / <i>gstörbet</i> or <i>gstörben</i>
AB-4: <i>ä / o</i>	<i>bräche</i> ‘to break’ / <i>brochet</i> or <i>brochen</i> , <i>träffe</i> ‘to strike’ / <i>troffet</i>

This dissociation between the AB and the suffix selection whereby the etymologically weak suffix *-et* is largely preserved also in concomitance of AB alternations is

<sup>3</sup> Similarly to what is observed in MSG, the past participle is formed by the simultaneous affixation of a prefix *g-* and one of the suffixes *-en*, *-et* or *-t*. Accordingly, we can speak of a circumfix. However, in contrast to the distribution found in MSG, in Titsch the prefix *g-* does not appear before an initial stop (cf. Titsch *bésset* vs. MSG *gebissen*), while it is unvoiced if the verbal stem begins with a vowel (cf. *oalte* ‘to age’ / *koaltet* vs. MSG *altern* ‘to age’ / *gealtert*) and unvoiced and merged with an initial posterior fricative (cf. [x] *läbe* ‘to stick’ / *kläbt* and [h] *efte* ‘to stick’ / *kaft*). Finally, it is also unvoiced and merged with the initial voiced stop in the three verbs *gä* ‘to give’ / *kät*, *gé* ‘to take’ / *két* and *goa* ‘to go’ / *kanget* (but not elsewhere: *grieze* ‘to greet’ / *griezt*).

<sup>4</sup> Actually, since the nasal in final position is normally dropped (cf. Gaeta 2023), we should assume the forms *abgeschlosse(n)* and *gשמולזן*. In fact, the nasal shows up only when it is followed by another segment, as in the inflected forms *abgeschlossne*, *abgeschlossens*, etc. For the sake of simplicity, I will refer to the basic form ending with the nasal.

<sup>5</sup> Further minor types are attested which display other allomorphies like *loufe* ‘to run’ / *gloffet*, *schloa* ‘to beat’ / *gשמלאגט*, *stoa* ‘to stand’ / *gstannet*, *troage* ‘to carry’ / *treit*, etc. Most of them can be assumed to form a class containing highly frequent verbs including auxiliaries and modals, the so-called *kurzformige Verben* (‘short-formed verbs’, see Angster and Gaeta 2018).

arguably supported by a mirror-image phenomenon which affected the weak verbs of the etymological *-ja-* class, namely so-called *Rückumlaut* (= RU) ('lit. backwards metaphony', cf. Vennemann 1986; Fertig 2020: 209). Because of the deletion of the thematic vowel *-i-* in the preterite of verbs displaying heavy syllables or more than one syllable, the OHG *i*-umlaut gave rise to a root-vowel alternation in the etymological *-ja-* class which runs in the opposite direction compared to the normal effect of the umlaut, from the marked category (the preterite) displaying the back vowel to the unmarked (the infinitive or the present) displaying the front vowel: OHG *h[ø:]ren* 'to hear' / *hōrta*, *w[æ:]nen* 'to mean' / *wānta*, *stellen* 'to put' / *stalta*, etc. This phenomenon was further expanded in Titsch (cf. Hotzenköcherle 1956; Zürner 1982: 92) to the extent that we can identify at least four different RU types which go well beyond the original alternations, see Table 3.

**Table 3:** RU types in IC II.

RU types	Examples
RU-1: <i>e / a</i>	<i>decke</i> 'to set' / <i>dackt</i> , <i>henge</i> 'to hang' / <i>kangt</i> , <i>verderpe</i> 'to corrupt' / <i>verdarpt</i>
RU-2: <i>ie / ue</i>	<i>ergriene</i> 'to green' / <i>ergruenet</i> , <i>rieme</i> 'to praise' / <i>gruemt</i> , <i>stiere</i> 'to lean' / <i>gstuert</i>
RU-3: <i>é / ò</i>	<i>chnéffe</i> 'to tie' / <i>knöpf</i> , <i>drécke</i> 'to print' / <i>dröckt</i> , <i>réschte</i> 'to roast' / <i>grösch</i>
RU-4: <i>é / oa</i>	<i>bréme</i> 'to soot' / <i>broamt</i> , <i>féerbe</i> 'to color' / <i>gfoarbt</i> , <i>méche</i> 'to mow' / <i>gmoat</i>

Note that in several cases the form displaying the RU alternates with its levelled-out correspondent as in *erblente* 'to be blended' / *erblant* or *erblent*, *stiere* 'to lean' / *gstuert* or *gstiert*, etc. Thus, the weak IC II converges with the strong IC I in the adoption of a non-concatenative technique in contrast to their original clear-cut distinction (see Dammel (2011: 251) for a further discussion of the parallel between AB and RU verbs).

As a consequence of the extension of the weak dental suffix in IC I, the outcomes of the OHG *-ē-* class like *lache* 'to laugh' / *glachet* merged with the outcomes of the strong class, especially when no root-vowel alternations occurred as in verbs like *ässe* 'to eat' / *kässet*, *falle* 'to fall' / *gfallet* (cf. OHG *e3zan* / *gie3zan*, *fallan* / *gifallan*), etc. In few cases, etymological strong verbs where no AB alternations occur migrated to the IC II as in *grabe* 'to dig' / *grabt*, *pflagen* 'to take care' / *pflegt* (cf. OHG *graban* / *gigraban*, *phlegan* / *giphlegan*). On the other hand, the distinction between the suffix of the past participle of IC I and of IC II is blurred when the verb stem ends in a coronal stop. In these cases a schwa vowel is mostly inserted to the effect that the verbs going back to the OHG *-ja-* class turn out to display a suffix *-et*: *erette* 'to save' / *erettet*, *leite* 'to lead' / *gleitet*, *anditte* 'to hint' / *andittet*, *spreite* 'to spread' / *gspreitet*, *verdörschte* 'to be parched' / *verdörschtet* (cf. OHG *retten* / *giretit*, *leiten* /

*gileitit, diuten / gidiutit, spreiten / gispreitit, dursten / gidurstit*), etc.<sup>6</sup> Furthermore, epenthesis is normally observed when the stem ends with an obstruent followed by a coronal sonorant like *entwéckle* ‘to develop’ / *entwécklet*, *zeichne* ‘to mark’ / *gzeichnet*, *ändre* ‘to change’ / *kändret*, etc. In other few cases, the participle – independently of the etymological class – swings between the form *-et* or *-t* respectively of IC I and IC II, possibly when one and the same verbal stem stands in combination with different prefixes: *blieche* ‘to blossom’ / *bliechet* or *bliecht*, *sörge* ‘to worry’ / *sörget* vs *versörge* ‘to provide’ / *versörget*, *verlange* ‘to demand’ / *verlanget* or *verlangt* vs *erlange* ‘to attain’ / *erlanget* (cf. OHG *bluoen / gibluoit, sorgēn / gisorgēt, langēn / gilangēt*), etc.

In sum, apart from IC III which is well carved out and only contains weak verbs that show no root-vowel alternations, the etymological distinction between the strong and the weak macro-class is blurred for several reasons: the OHG *-ē-* class is split across IC I and IC II, and the strong verbs take over the *-et* suffix which replaces the etymological *-en* in a number of cases. As a consequence, the etymological strong verbs where no root-vowel alternation of the four AB types seen in Tab. 2 above occurs like *lade* ‘to load’ / *gladet* or *gladen* (cf. OHG *ladan / giladan*) cannot be distinguished from etymological weak verbs like *nachte* ‘to get dark’ / *gnachtet* (cf. OHG *nachtēn / ginachtēt*) except for the nasal suffix *-en* which appears as an alternative to *-et*. Thus, the occurrence of the (strong) nasal suffix as a possible alternative provides a criterion for IC assignment to the extent that this latter alternative is justified in etymological terms. Accordingly, I will assume as a null hypothesis that the presence of the nasal suffix *-en* reflects its origin rather than being an innovation, and I will assign to IC I all verbs displaying the nasal suffix, independently of the absence of any AB alternation as in *broate* ‘to roast’ / *broatet* or *broaten* and *scheide* ‘to choose’ / *gscheidet* or *gscheiden*, etc. (cf. OHG *brātan / gibrātan, skeidan / giskeidan*) or of the occurrence of a suffix *-t* instead of the expected *-et* in the past participle as for instance in *läse* ‘to read’ / *gläst* or *gläsen* (cf. OHG *lesan / gilesan*). In the same vein, the few etymological weak verbs which display the nasal suffix as a possible alternative in the past participle as for instance *bégleite* ‘to accompany’ / *bégleitet* or *bégleiten* and *stére* ‘to stir’ / *gstért* or *gstört* or *gstören* (cf. OHG *gileiten / gileitit, stiuren / gistiurit*), etc. will be included into IC I. On the other hand, since verbs whose stem ends with an obstruent-sonorant cluster such as *ändre* ‘to

<sup>6</sup> In few cases like *schétte* ‘to shake’ / *gschétt*, and *hefte* ‘to stick’ / *kaft*, etc. (cf. OHG *scutten / giscutit, heften / giheftit*) no epenthesis is found and the suffix *-t* is fused with the stem. Moreover, it has to be added that a few verbs of the etymological *-ō-* class which migrated to IC II also belong here as their stem ends with a coronal stop: *bélde* ‘to form’ / *béldet*, *anméle* ‘to announce’ / *angméldet* (cf. OHG *bilidōn / gibilidōt, meldōn / gimeldōt*).

change' / *kändret* never display the nasal suffix as a possible alternative, they will be assigned to IC II.

Since the AB alternations characterizing the strong verbs are paralleled by the root-vowel alternations resulting from the RU of the weak verbs, we are not surprised to observe cases like *schrie* 'to scream' which used to be an etymological strong verb belonging to the AB-1 class like *blibe* and shifted to the RU-2 class as shown by its past participle *gschruet*. In addition, in at least one case we happen to find exactly the same root-vowel alternation as shown by the AB-3 and the RU-3 class, where only the suffix allows us to distinguish the strong *bénne* 'to bind' / *bònnet* from the weak *drécke* 'to print' / *dròckt*. In practice, in Titsch we are left with four additional micro-classes found within ICs I and II which are characterized by different root-vowel alternations accompanying the different participial suffixes, as summarized in Table 4.

**Table 4:** The first two verbal ICs in Titsch.

IC I	examples	IC II	examples
IC Ia	<i>vergässe / vergässen or -et</i>	IC IIa	<i>chläbe / kläbt</i>
IC Ib-AB-1	<i>blibe / blében or -et</i>	IC IIb-RU-1	<i>decke / dackt</i>
IC Ic-AB-2	<i>siede / gsotten or -et</i>	IC IIc-RU-2	<i>rieme / gruemt</i>
IC Id-AB-3	<i>bénne / bònnen or -et</i>	IC IId-RU-3	<i>drécke / dròckt</i>
IC Ie-AB-4	<i>bräche / brochen or -et</i>	IC IIe-RU-4	<i>bréme / broamt</i>

Moreover, as observed above, IC I displayed the occurrence of the strong and the weak participial suffixes, respectively *-en* or *-et*. In this regard, Table 4 emphasizes the dissociation between the root-vowel alternation and the type of suffix selected, insofar as the former is found both in IC I and in IC II and in some cases points to a convergence between them.<sup>7</sup> Finally, no micro-classes occur within IC III, where only sparse allomorphies are found like RU-1 in *stéckò* 'to dismember' / *gstackt*.

<sup>7</sup> It must be added that – similarly to what we have seen above for *lade* 'to load' / *gladet* – when the verbal stem ends with a coronal stop a complete neutralization between the IC Id and the IC Iic takes place due to schwa insertion, as in the case of the etymological strong verb *verschwénde* 'to disappear' / *verschwòndet* vs. the etymological weak verb *grénde* 'to found' / *gròndet*.

### 3 Modeling variation in the Titsch verbal classes: A quantitative view

In the previous section, we have seen that a certain degree of variation is found with regard to the inflectional properties of the verbs in Titsch. Thanks to the research projects mentioned above, I will try to model this variation by making use of the CLiMALp archive which allows us to provide a quantitative evaluation of the verbal system of Titsch (cf. Gaeta et al. 2022). The CLiMALp archive, which is freely accessible online at the website [www.climalp.org](http://www.climalp.org), consists of two parts. First, a dictionary resulting from the community's collective engagement (cf. GrW 1998) has been stored and filed, and is now consultable in Titsch, Italian and German. This dictionary can be credited as a true thesaurus which is normally used as a reference work by the speakers in their everyday conversations (see also Gaeta et al. 2024). Second, the dictionary is interfaced (and thereby enriched) with a corpus which collects written texts produced in the local variety in the last century relating to disparate subjects, from church bulletins to cookbooks and further ethnological material. Being interfaced with the dictionary, the corpus is fully annotated and manually checked. Similar archives are already partially available for the other Walser German varieties spoken in Italy as well. The possibility of relying on a corpus besides a dictionary is particularly important in the light of the strong processes of decay which characterize Walser varieties in Italy and especially Titsch, with only a few dozen speakers left who are really able to master it at a mother-tongue level alongside Italian, Piedmontese and partially French, which constitute the dominant varieties of their repertoire (cf. Zürner 1999, 2009; Dal Negro 2004 for discussion). For this reason, in what follows I will mainly rely on the data extracted from the corpus and I will assign the verbs to their respective ICs solely on the basis of what is witnessed in the corpus, independently of what is recorded in the dictionary. This is also consistent with the fact that inflected forms of the past participles – which, as we will see, are of crucial importance for our investigation – can only be found in the corpus. By maximizing the usage-based perspective in this way, I also aim at reducing the possible effects of normalization and/or analogization typically found with lexicographic collections, which have particularly dramatic consequences in the case of a language undergoing a strong process of decay like Titsch.

At present, for Titsch the archive roughly features 12,315 lexical entries and 93,052 corpus tokens. The data relating to the verbs provided by the archive are reported in Table 5.<sup>8</sup>

---

<sup>8</sup> Since the CLiMALp archive is constantly being revised and corrected, the data presented here can slightly vary with respect to those found in the platform. However, this variation is not likely to alter the results presented in the rest of the paper.

**Table 5:** Verbs in CLiMAIp.

Verbs not attested in the corpus	1,799	61.0%
Verbs without attested participle	401	13.6%
Verbs displaying weak participles	674	22.9%
Verbs displaying strong participles	28	1.0%
Verbs displaying strong and weak past participles	45	1.5%
Total verbs	2,947	100.0%

As one may note, one quarter of the lexical entries of the dictionary consists of verbs, and about forty percent are also attested in the corpus, although only one quarter of the verbs has a corpus attestation for the past participle, which is the relevant object of investigation here. On the other hand, the verbs whose participle is attested constitute two-thirds of all verbal lexemes occurring in the corpus, which – in spite of its limited size – is large enough to warrant a significant empirical contribution to model the variation observed above and to justify the usage-based perspective adopted here. In this light, nine out of ten verbs whose past participle is attested in the corpus select the weak suffix, independently of the occurrence of any root-vowel alternation, as shown in Table 6.

**Table 6:** Verbs displaying past participles in CLiMAIp.

Verbs displaying weak past participles	674 / 90.2%
Verbs displaying strong past participles	28 / 3.7%
Verbs displaying strong and weak past participles	45 / 6.0%
Total	747 / 100.0%

This clearly shows the tendency towards the extension of the weak participles to the detriment of the strong participles, which is commonly found throughout all Germanic languages. On the other hand, there is a small but significant number of verbs which only display the strong form and a slightly larger number of verbs displaying double forms. Thus, the weak suffix is clearly dominant, but the strong suffix remains robust insofar as it is found with about one tenth of the verbs.

Note that the process of extension of the weak suffix does not imply the shift to IC II, the etymological weak class. In fact, if we investigate the verbs whose past participles is found in the corpus in dependence of their IC, the following figures obtain, see Table 7.

**Table 7:** Verbal ICs in CLiMALp.

IC I	241 / 32.3%
IC II	332 / 44.4%
IC III	174 / 23.3%
Total	747 / 100.0%

Table 7 shows that IC II is larger than the other two, but also that IC I is fairly robust and includes one third of all verbs whose past participle occurs in the corpus. Furthermore, even if in both IC I and IC II the pattern without root-vowel alternation is dominant, the other micro-classes are well-attested as shown by Table 8.<sup>9</sup>

**Table 8:** Allomorphies in the first two verbal ICs in CLiMALp.

IC I		IC II	
IC Ia (AB-Ø)	94 / 39.2%	IC IIa-(RU-Ø)	274 / 82.3%
IC Ib (AB-1: <i>i/ê</i> )	27 / 11.3%	IC IIb (RU-1: <i>e/a</i> )	27 / 8.1%
IC Ic (AB-2: <i>ie/o</i> )	30 / 12.5%	IC IIc (RU-2: <i>ie/ue</i> )	8 / 2.4%
IC Id (AB-3: <i>é/ò</i> )	31 / 12.9%	IC IId (RU-3: <i>é/ò</i> )	8 / 2.4%
IC Ie (AB-4: <i>ä/o</i> )	17 / 7.1%	IC IIe (RU-4: <i>é/oa</i> )	2 / 0.6%
IC Ir (residue)	42 / 17.5%	IC IIr (residue)	13 / 3.9%
Total	241 / 100.0%	Total	332 / 100.0%

Especially in IC I all different AB types are fairly well attested, while the RU alternations of IC II are far less frequent and micro-class IIa without alternation is the clearly dominant type. It must be added that the skewness of this finding is partially due to the choice of including into IC II all verbs ending with a coronal stop or sonorant where no strong participles are found, although theoretically they might also be assigned to IC I. In this latter IC we observe quite a significant amount of the residual micro-class Ir in which a root-vowel alternation – possibly accompanied by other allomorphies – of a different type is found. This is due to the group of *kurzformige Verben* hinted at in footnote 5 above, which mostly go back to etymological strong verbs or to other highly frequent verbs coming from other classes and display the suffix *-et* as for instance *foa* ‘to catch’ / *gfanget*, *schloa* ‘to hit’ / *gschlaget*, etc. or further idiosyncratic allomorphies as in *troage* ‘to carry’ / *treit*, *tue* ‘to do’ / *toat*, *goa* ‘to go’ / *kanget*, etc.

<sup>9</sup> In IC II, there are around twenty verbs displaying forms which may present a root-vowel alternation or not. In order to maximize variation, they have been assigned to their respective alternation classes.

Focusing solely on IC I, in Table 9 the distribution of the strong suffix *-en* and of the weak suffix *-et* across its micro-classes is reported.

**Table 9:** Strong and weak inflection in the IC I in CLiMAIp.

IC I	Strong suffix	Double suffix	Weak suffix	Total
IC Ia (AB-Ø)	7 / 7.4%	19 / 20.2%	68 / 72.3%	94 / 100.0%
IC Ib (AB-1: <i>i/é</i> )	8 / 29.6%	4 / 14.8%	15 / 55.6%	27 / 100.0%
IC Ic (AB-2: <i>ie/o</i> )	8 / 12.5%	5 / 16.7%	17 / 56.7%	30 / 100.0%
IC Id (AB-3: <i>é/ò</i> )	4 / 12.9%	6 / 19.4%	21 / 67.7%	31 / 100.0%
IC Ie (AB-4: <i>ä/o</i> )	1 / 7.1%	3 / 17.6%	13 / 76.5%	17 / 100.0%
IC Ir (residue)	- / -	8 / 19.0%	34 / 81.0%	42 / 100.0%

As can be gathered from the Table 9, verbs displaying weak inflection are scattered across all the micro-classes, although micro-class Ia, where no root-vowel alternation occurs, is robustly attested. These figures clearly support the idea that a neat dissociation between the root-vowel alternation and the suffix selection of the strong or weak suffix has to be assumed. Accordingly, the selection of the strong / weak suffix and the occurrence of AB alternations must be considered two independent issues which have to be investigated separately. The latter issue relating to the AB alternations will be completely ignored in the rest of the analysis.

## 4 The role of agreement in the Titsch past participles

Apparently, the distribution of the strong and the weak suffix across the verbs of IC I seems to be a classic example of (morpho-)lexically determined ICs showing a reduction of lexical complexity through the generalization of the most recent pattern. As commonly found in the Germanic languages, the strong pattern is apparently ceding to the more recent weak model at different speeds in dependence of the specific micro-class. Thus, in Table 9 the micro-classes Ib and Ic seem to be slightly more resistant to the shift in favor of the weak class than the micro-classes Id and Ie where the process appears more advanced.

However, a different explanation can be suggested for the variation observed in Table 9. This is because in Titsch – as well as in other southern German varieties, cf. Fleischer (2007) – participles and adjectives show agreement in predicative position in addition to its presence in attributive position in sharp contrast to MSG and

the rest of the West-Germanic varieties where agreement in predicative position was lost fairly early in history (cf. Gaeta 2018 for discussion).

Accordingly, in Titsch adjective agreement is observed in the copulative constructions containing the verbs *si* ‘to be’ (1a), *chéeme* ‘to come’ (1b) and *blibe* ‘to remain’ (1c), and the same holds for agreement on past participles in the BE-perfect construction (1d–e).<sup>10</sup>

- (1) a. *d’hannetò moss nid déck-é ò nid*  
 DEF=mush[F] must NEG thick-F.SG or not  
*dénn-é si*  
 thin-F.SG be.INF  
 ‘The mush has to be neither thick nor thin’.
- b. *well de were der ange*  
 because then were DEF.M.SG butter[M]  
*rék-e kiem-et*  
 rancid-M.SG come-PST.PTCP  
 ‘because the butter would then have become rancid’.
- c. *ésch z’ganz hus woarem-s*  
 is DEF.N.SG=whole house[N] warm-N.SG  
*bléb-et*  
 remain-PST.PTCP  
 ‘The whole house has remained warm’.
- d. *Hilde òn Cristina sinn dr-obèr gsatz-t-é*  
 Hilde and Cristina be.3PL there-on seat-PST.PTCP-F.PL  
 ‘Hilde and Cristina are seated thereon’.
- e. *de Jònker-Ronker ésch rächt*  
 DEF.M.SG Jònker-Ronker is right  
*antschloaf-n-e gsid*  
 fall.asleep-PST.PTCP-M.SG been  
 ‘the Jònker-Ronker had immediately fallen asleep’.

It has to be noted that the past participle of the copulative verbs (see respectively *kiemet* and *blébet* in (1b–c)) as well as of the auxiliaries (see *gsid* in (1e)) do not display agreement. Moreover, while the BE-perfect is found with unaccusative

<sup>10</sup> Titsch examples are extracted from the corpus and given in the orthographic form found in the original text archived in the CLiMALp archive. For brevity, no exact reference of the examples in the archive is provided, but they can easily be found accessing the free dictionary at [www.climalp.org/index.php/dizionari/](http://www.climalp.org/index.php/dizionari/).

verbs, unergative verbs normally display HAVE-perfect where agreement – as expected – is not found, see (2).

- (2) a. *wenn all-e hein hert gschloaf-et*  
 when all-PL have.3PL hard sleep-PST.PTCP  
 ‘When all were sleeping hard’.  
 b. *aber héibèr véll glach-et*  
 but have.1PL much sleep-PST.PTCP  
 ‘But we slept hard’.

However, uninflected forms are commonly found also with the BE-perfect when weak past participles occur, see (3).<sup>11</sup>

- (3) a. *Laura òn Augusta sinn vòn ussna kchéem-et*  
 Laura and Augusta be.3PL of outside come-PST.PTCP  
*als jòng-é techtr-e*  
 as young-F.PL daughter[F]-PL  
 ‘Laura and Augusta have come from abroad as young daughters’.

**11** Moreover, unergative verbs swing between the HAVE- and the BE-perfect as shown by the example (i).

- (i) *de Jònker-Ronker esch òf de réck gschloaf-et*  
 DEF.M.SG Jònker-Ronker[M] is on DEF back[M] sleep-PST.PTCP  
 ‘The Jònker-Ronker has slept on his back’.

On the other hand, unaccusative verbs can also display the selection of the HAVE-perfect, as in (ii).

- (ii) *heintsche de meischter òn d’préftò am*  
 have.3PL DEF owner.PL and DEF=back.staff[F] at.DEF  
*tésch gsetz-t*  
 table[M] seat-PST.PTCP  
 ‘The owners and the back staff have seated at the table’.

Notice that this variation cannot be related to the influence of the contact varieties (Piedmontese and Italian), where respectively the HAVE- and the BE-perfect are used. Further problems are caused by the unaccusative usage of transitive verbs like *bräche* ‘to break’ as in the following sentence in which the BE-perfect is used in the absence of any agreement, see (iii).

- (iii) *Débel tieber scribe ésch d’chelte broch-et*  
 while do.1PL write.INF is DEF=cold[F] break-PST.PTCP  
 ‘While we are writing, the cold has given in’.

At any rate, more research is required here which goes far beyond the limits of this contribution.

- b. *Noa deer gwess-o zit ésch d'sònnò*  
 after DEF.F.SG.DAT certain-F.SG time[F] is DEF=sun[F]  
*emòm erschéen-et*  
 again shine-PST.PTCP  
 'After a certain time the sun has (lit. is) shined again'.

Much more consistent in the written sources is the expression of agreement in the several passive constructions found in Titsch which select different auxiliaries, namely the BE-passive (4a), the COME-passive (4b) and the GO-passive (4c) as well as the small clauses involving past participles (4d) (see Gaeta 2018, 2020; Gaeta et al. in press for the discussion of the passive construction in Titsch).

- (4) a. *Os der Schreft chammo erchenne*  
 from DEF.F.SG.DAT writing[F] can.one recognize.INF  
*wette d'Beldong von de Greschoneyer-Walser*  
 how DEF=education[F] of DEF.M.SG Greschoneyer-Walser  
 ... *entwecklo-t-e gsid escht*  
 develop-PST.PTCP-F.SG been is  
 'From writing one can recognize how the culture of the Greschoneyer-Walser has been developed'.
- b. *was hie setter en par joar*  
 what here since INDEF.N.SG pair year.PL  
*chén-t publizier-z*  
 come-s publish-PST.PTCP.N.SG  
 'what is being published here for a couple of years'
- c. *De toufnoamn-a sin of franzesésch*  
 DEF.M.PL forename[M]-PL be.3PL on French  
*abkändre-t-e kanget*  
 change-PST.PTCP-M.PL gone  
 'The forenames have been changed into French'.
- d. *Fenn-é où ufgschréb-en-z of enz*  
 find-1SG also write.down-PST.PTCP-N.SG on our  
*notizbuech dass wier hättéber sollò*  
 notebook that we have.SUBJ.PST.1PL should.INF  
*noasieché*  
 investigate.INF  
 'I also find written down on our notebook that we should investigate'.

Note that in the small clause construction (4d) the past participle displays agreement with the object clause which selects neuter and singular as default values. Finally,

past participles also agree with their nominal heads, both in the attributive prenominal position (5a) and in the appositive postnominal position (5b).

- (5) a. *en dé vergang-n-é zitte*  
       in DEF.F.SG pass-PST.PTCP-F.SG time[F]  
       ‘in the past time’
- b. *mé dem eigen-e noame*  
       with DEF.M.SG.DAT own-M.SG name[M]  
       *dr-ònder gschréb-n-e*  
       there-under write-PST.PTCP-M.SG  
       ‘with one’s own name written thereunder’

To account for the widespread occurrence of adjective agreement, a continuity explanation has traditionally been suggested which assumes a direct preservation of the OHG situation where agreement is normally found. Alternatively, the effects of a more or less recent contact with Romance varieties have been emphasized where agreement shows up throughout all contexts seen above (see Fleischer 2007 for a detailed discussion and further references). However, while the effects of the constant contact with Italian and with the other Romance varieties – in particular Piedmontese – are undeniable, especially with regard to the preservation of the unstressed final vowels mentioned above, the actual state-of-affairs cannot be said to go immediately back to the contact varieties. Apart from specific Titsch constructions which are unknown in the contact varieties such as the GO-passive restricted to the past tense (see (4c) above), it has to be stressed that the restriction on the agreement of the past participle of auxiliaries is peculiar of Titsch, while both Italian and the other Romance varieties display full agreement throughout the whole verbal complex: Italian *la ragazza è stata accompagnata dal fratello* ‘the girl has been accompanied by her brother’.<sup>12</sup> Thus, the contact with Romance has surely played a crucial role in supporting the preservation of agreement, but its peculiar properties in Titsch cannot be explained away as a calque based on Romance varieties.

Given the wide range of contexts in which Titsch past participles display agreement, here is where the CLiMAlp archive really becomes fruitful insofar as it allows us to extract the data showing participle agreement found in the corpus. They are reported in Table 10 distributed across the three ICs:

<sup>12</sup> French – at least in its written variety – provides a partial exception because in verbal complexes containing the auxiliary BE the latter remains uninflected while the participle of the full verb must display agreement: *la fille à été / \*étée accompagnée / \*accompagné par son frère* ‘the girl has been accompanied by her brother’.

**Table 10:** Inflectional behavior of past participles in CLiMAIp scattered across the ICs.

	IC I	IC II	IC III
Verbs with uninflected weak participles	158 65.6%	228 68.7%	116 66.7%
Verbs with inflected weak participles	10 4.1%	104 31.3%	58 33.3%
Verbs with inflected strong participles	28 11.6%	–	–
Verbs with inflected strong and uninflected weak participles	38 15.8%	–	–
Verbs with inflected strong and weak participles	7 2.9%	–	–
Total	241 100.0%	332 100.0%	174 100.0%

While Table 10 confirms that the presence of strong participles only affects IC I verbs, it also shows that for each IC two-thirds of the verbs display past participles with a weak suffix which is uninflected. The remaining third largely consists of verbs displaying inflected weak participles in IC II and IC III, while in IC I we find verbs in which the presence of strong inflection is dominant, possibly accompanied by uninflected weak inflection: they make up a total of more than one quarter of IC I verbs. Finally, it must be stressed that no verbs displaying uninflected strong past participles are found in the corpus. Thus, while in all ICs the same proportion of verbs displaying uninflected weak participles is observed, in IC I the group of verbs displaying inflected weak past participles is clearly outranked by those displaying inflected strong participles. On the basis of these findings, we can concentrate on IC I where the strong / weak alternation is found.

## 5 A new feature for the strong / weak alternation

Let us now consider in detail the distribution of the strong and weak suffixes found in IC I verbs with respect to the relevant morphosyntactic environment. In the following two tables the data presented for IC I in Table 10 above are recast (respectively in types and tokens) introducing the morphosyntactic environments in which past participles are found, that is, the HAVE-perfect (H-Pe), the BE-perfect (B-Pe), the BE-passive (B-Pa), the COME-passive (C-Pa), the GO-passive (G-Pa), the attributive position (At), the appositive position (Ap) and the small clause construction (SC):

**Table 11:** Past participles of the IC I scattered across their morphosyntactic environment (types).

	H-Pe	B-Pe	C-Pa	SC	G-Pa	Ap	B-Pa	At	Total
a. [+ w, - i] <sub>PP</sub>	142 99.3%	68 85.0%	1 33.3%	–	2 9.5%	2 7.4%	2 6.7%	–	217 63.5%
b. [+ w, + i] <sub>PP</sub>	1 0.7%	–	1 33.3%	–	2 9.5%	7 25.9%	–	18.8% 6	17 5.0%
c. [(+ s, + i) & (+ w, + i)] <sub>PP</sub>	–	–	–	–	2 9.5%	–	–	–	2 0.6%
d. [(+ s, + i) & (+ w, - i)] <sub>PP</sub>	–	4 5.0%	–	–	1 4.8%	–	2 6.7%	–	7 2.0%
e. [+ s, + i] <sub>PP</sub>	–	8 10.0%	1 33.3%	6 100.0%	14 66.7%	18 66.7%	26 86.7%	26 81.3%	99 28.9%
Total (distribution regarding the environments)	143 41.8%	80 23.4%	3 0.9%	6 1.8%	21 6.1%	27 7.9%	30 8.8%	32 9.4%	342 100.0%

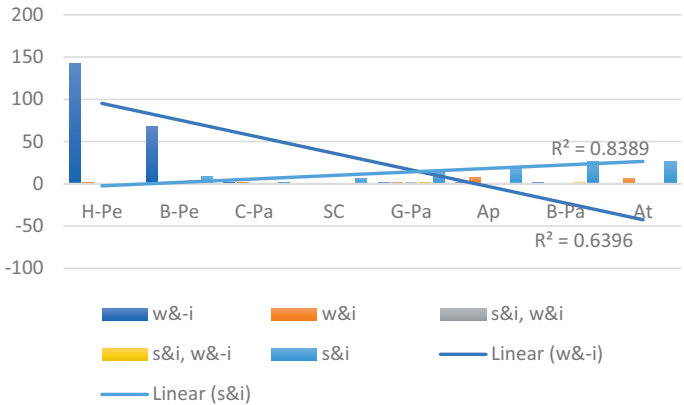
**Table 12:** Past participles of the IC I scattered across their morphosyntactic environment (tokens).

	H-Pe	B-Pe	C-Pa	SC	Ap	B-Pa	G-Pa	At	Total
a. [+ w, - i] <sub>PP</sub>	719 99.9%	564 92.9%	1 20.0%	–	2 5.6%	5 10.2%	2 3.5%	–	1293 83.6%
b. [+ w, + i] <sub>PP</sub>	1 0.1%	–	2 40.0%	–	9 25.0%	–	5 8.8%	6 9.1%	23 1.5%
c. [(+ s, + i) & (+ w, + i)] <sub>PP</sub>	–	–	–	–	–	–	6 10.5%	–	6 0.4%
d. [(+ s, + i) & (+ w, - i)] <sub>PP</sub>	–	19 3.1%	–	–	–	9 18.4%	3 5.3%	–	31 2.0%
e. [+ s, + i] <sub>PP</sub>	–	24 4.0%	2 40.0%	7 100.0%	25 69.4%	35 78.4%	41 71.9%	60 90.9%	194 12.5%
Total (distribution regarding the environments)	720 46.5%	607 39.2%	5 0.3%	7 0.5%	36 2.3%	49 3.2%	57 3.7%	66 4.3%	1547 100%

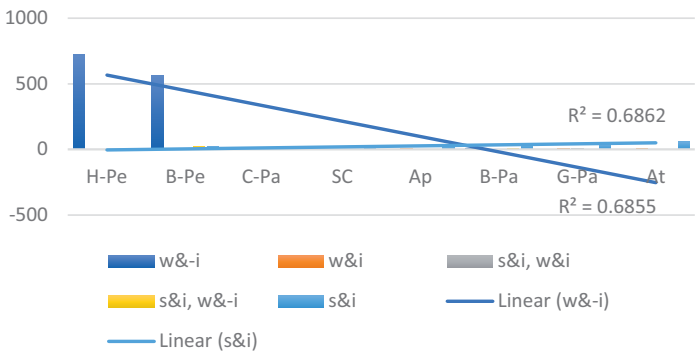
Past participles can display the weak (+ w) or the strong (+ s) suffix and can be inflected (+ i) or not (– i) in dependence of the morphosyntactic environment. Recall that the only case where inflection is not expected to occur is in the HAVE-perfect, where in fact only one isolated case is found which does not disturb the general picture (see the type (b) under H-Pe). The data displayed in Tables 11 and 12 are clearly polarized towards the two opposite morphosyntactic environments: on the

one pole, the uninflected HAVE-perfect (type (a)), where verbs select the weak suffix, and on the other all other cases where inflected participles are required, where the strong suffix is selected (type (e)).

Figure 1 and 2 chart the data of Table 11 and 12, and – in spite of the distribution skewed in favor of the perfect – show with the help of a linear regression trendline the polarized distribution of the weak and the strong participles in dependence of the respective morphosyntactic environments in which inflection is required:



**Figure 1:** Past participles of the IC I scattered across their morphosyntactic environment (types).



**Figure 2:** Past participles of the IC I scattered across their morphosyntactic environment (tokens).

In order to better visualize the polarized distribution, the morphosyntactic environments are aligned according to their type/token frequency shown in the last row of Tables 11 and 12, considered respectively from left to right for the two weak

uninflected contexts and from right to left for the inflected strong contexts. Note that the values are clearly more solid for the type distribution, while among the tokens a robust quantitative prevalence of the perfect constructions flattens the results. However, such a skewed distribution is not surprising in the light of the relevance of the perfect construction as the only way to express anteriority in Titsch, and the comparatively low frequency of the passive constructions normally observed cross-linguistically.

Between the poles sparse examples are found, where either only verbs selecting weak suffixes are found (type (b) in Tables 11 and 12), or uninflected weak participles are accompanied by their inflected strong correspondents within one and the same verb (type (d) in Tables 11 and 12). The only case in which the correlation between the morphosyntactic environment and the distribution of the strong / weak suffix crashes is represented by the type (c) in Tables 11 and 12, but this only regards the two verbs *ertue* ‘to open’ (6a–b) and *läse* ‘to read’ (6c–d), where strong and weak suffixes are found with inflected participles.

- (6) a. *En heiò éscht e nu-e teil vòm*  
 in July is INDEF new-M.SG part[M] von.DEF.N.SG  
*ekomuseum Walser ertò-a-n-s kanget*  
 eco-museum[N] Walser open-PST.PTCP-N.SG gone  
 ‘In July a new part of the Walser eco-museum has been opened’.
- b. *Desch-e Sommer éscht ertoa-t-s kanget*  
 DEM-M.SG summer[M] is open-PST.PTCP-N.SG gone  
 ... *z Hus fer Sport*  
 DEF.N.SG house[N] for sport  
 ‘This summer the house for sport has been opened’.
- c. *bés nid langorsch éscht em hitteg-e tag*  
 until NEG longest is in.DEF.M.SG present-M.SG day[M]  
*en friemäsch kanget gläs-n-e*  
 INDEF.F.SG High.Mass[F] gone read-PST.PTCP-F.SG  
 ‘Until not long ago a High Mass has been celebrated in the present day’.
- d. *dernoa gséchéber de eppes geit*  
 thereafter see.1PL then something[N] goes  
*ou gläs-z*  
 also read-PST.PTCP.N.SG  
 ‘Thereafter let’s see then, something has also to be read’.

Note that *ertue* is one of those short-formed verbs which display a high number of morphological idiosyncrasies and in fact has been grouped with the IC 1 only for convenience, while *läse* only occurs in the two tokens reported above and is

therefore difficult to judge. It must be added that – as expected on the basis of what has been discussed with respect to (2) and (3) above – the BE-perfect, which typically characterizes unaccusative verbs, normally follows the same pattern of the HAVE-perfect insofar as an uninflected weak participle is used, while only in few cases the inflected strong model prevails (see the types (d) and (e) under B-Pe in Tables 11 and 12).

A global pattern emerges from the polarized distribution reported in Table 11, which is made evident by the comparison of the total number of types found in the last column of Table 11 with the total number of the verbs of the IC I reported in Table 10. The increment of hundred types (viz. 342 vs. 241) is apparently due to the total number of types displaying an inflected strong participle. This, however, only means that for most inflected strong participles occurring in the appropriate morphosyntactic environment an uninflected weak participle in the HAVE-perfect is found. In other words, besides verbs where only the weak inflection is found such as *fénne* ‘to find’, *gwénne* ‘to win’, *schnide* ‘to cut’ and few others, the strong / weak alternation has been completely remodeled according to the construction type in which the verb is involved. Therefore, most verbs of the IC I display two past participles which are selected in dependence of the morphosyntactic environment, as exemplified below by the verbs *schrìbe* ‘to write’ and *wäsche* ‘to wash’. When agreement is not required by the context as in the HAVE-perfect, the weak uninflected form is found (7a–b), while the strong inflected form shows up when agreement is contextually required as in the passive (7c–d).

- (7) a. *wenn Benito Leopold Curtaz ... hät éndsch*  
 when Benito Leopold Curtaz has 1PL.OBL  
*gschréb-et / \*gschréb-n-e*  
 write-PST.PTCP / write-PST.PTCP-M.SG  
 ‘When Benito Leopold Curtaz . . . has written to us’.
- b. *heintsch d’anket-e entfläcko-t*  
 have.3PL DEF=skirt[F]-PL remove.stain-PST.PTCP  
*òn d’hus-gspònt-o woll-schtrangn-a*  
 and DEF=house-woven-PL wool-skein-PL  
*gwässch-et / \*gwässch-n-e*  
 wash-PST.PTCP / wash-PST.PTCP-PL  
 ‘They have removed the stains from the skirts and washed their home-made wool-skeins’.

- c. *al-z*      *éscht*    *kanget*  
 all-N.SG    is            gone  
*gschréb-en-z*            / \**gschréb-et-s*  
 write-PST.PTCP-N.SG / write-PST.PTCP-N.SG  
 ‘Everything has been written’.
- d. *òn*    *d’gröss-ò*    *lougò*      *ésch*  
 and    DEF=big-F.SG    laundry[F]    is  
*gwässch-n-e*            / \**gwässch-et-e*      *kanget*  
 wash-PST.PTCP-F.SG / wash-PST.PTCP-F.SG    gone  
 ‘And the big laundry has been washed’.

Note that this remodeling is also found with a few originally weak verbs where the strong suffix is not expected etymologically, such as for instance *spreite* ‘to spread’ in (8).

- (8) a. *En*    *de*      *sòmmermaned-a*    *hämmò*    *schén*      *déck*  
 in    DEF.PL    summer.month-PL    has.one    beautiful    thick  
*òn*    *géere*    *d’wäsch*            *òf*  
 and    gladly    DEF=laundry[F]    on  
*d’husmattò*                    *gspreit-et*  
 DEF=house.meadow    spread-PST.PTCP  
 ‘In summer months one has quite often and gladly spread the laundry on the meadow in front of the house’.
- b. *d’Mann-a*    *hein*      *de*      *Trosso*            *gmach-t:*  
 DEF=man-PL    have.3PL    DEF.M.SG    hay.bundle[M]    make-PST.PTCP  
*dri*    *Seil-ene*    *uf*    *em*            *Bode*    *gschpreit-n-e,*  
 three    rope[N]-PL    on    in.DEF.M.SG    ground    spread-PST.PTCP-PL  
*dä*    *Tregi-e*                    *em*            *Bode*      *engschtackt-e*  
 DEF.PL    packing.wood-PL    in.DEF.M.SG    ground[M]    pouch.PST.PTCP-PL  
*un*    *druf*            *Oarfal*    *uf*    *Oarfal*  
 and    thereupon    armful    on    armful  
 ‘The men have made the bundle of hay: three ropes spread out on the ground, the packing wood pouched in the ground, and thereupon armful on armful of hay’.

Note that even in the cases assumed to display only the strong or the weak suffix in previous investigations, we observe in the corpus the scattered distribution of Tables 11 and 12. For instance, Zürrer (1982: 91), based on fieldwork carried out in the Seventies, deems as “strong verbs” *sénge* ‘to sing’ for which the strong / weak alternation is documented in the corpus (see (9a–b) below), while other verbs

labelled as “strong verbs with weak ending of the past participle” such as *bache* ‘to bake’ and *zie* ‘to pull’ also display the alternation (see (9c–d)).

- (9) a. *Ende der Mäsch heintsch met däm*  
 end DEF.F.SG.GEN Mass[F] have.3PL with the.M.SG.DAT  
*Pfoaher z’Stabat Mater gsung-et*  
 priest[M] DEF.N.SG=Stabat Mater sing-PST.PTCP  
 ‘At the end of the Mass they have sung the Stabat Mater with the priest’.
- b. *Of jede fall éscht z’lied*  
 on any-M.SG fall[M] is DEF.N.SG=song[N]  
*gsong-en-z kanget*  
 sing-PST.PTCP-N.SG gone  
 ‘In any case the song has been sung’.
- c. *de ma wò hannensch noch hientoa*  
 DEF.PL men REL have.3PL.1PL.OBL yet occasionally  
*z’Noversch schwarz brot bach-et*  
 to=Noversch black bread bake-PST.PTCP  
 ‘The persons who still have occasionally baked us brown bread in Noversch’.
- d. *Ennéra halb stòn òngefer ésch*  
 in.F.SG.DAT half hour[F] about is  
*z’bròt bach-en-z gsid*  
 DEF.N.SG=bread[N] bake-PST.PTCP-N.SG been  
 ‘Within about half an hour the bread was baked’.

However, independently of the diachronic aspects, the synchronic treatment of this remodeling raises a number of questions centering on the key factors which have to be made responsible for the occurring alternation. This issue will be pursued in the next section.

## 6 Remodeling verbal ICs in Titsch

To sum up, the verbal ICs in Titsch are organized around different properties. First, it is necessary to consider (morpho-)lexical information. While IC III verbs are clearly individuated on the basis of ThV -ò found in the infinitive and in the participle, this is not straightforward for the other two ICs because in both classes the infinitive ending is -e. For IC I verbs the vowel -e can be held to be a true ThV because it is found both in the infinitive and in the participle. On the other hand,

IC II verbs are exclusively individuated by the ending of the past participle *-t* in the absence of any ThV. Accordingly, the vowel *-e* found in their infinitive might be taken to be a default vowel required by syllabic well-formedness: recall that in the past participles of IC II verbs ending with a coronal stop or sonorant the same vowel *-e* is inserted as in *leite* ‘to lead’ / *gleitet* and *entwéckle* ‘to develop’ / *entwécklet*, blurring in this way the contrast with IC I verbs.

Second, once the verbs of the respective ICs have been distinguished on a (morpho-)lexical basis, then we need to know – again on a lexical basis – which verbs belonging to IC I only show the weak suffix. This is a significant subset which consists of verbs displaying stem vowel alternations (those going back to etymological strong verbs) and of verbs not displaying any alternation (mostly going back to etymological weak verbs). We will comprise this first subset under the label of IC I<sub>W</sub>. In addition, we find the verbs where both strong and weak suffixes are found. The latter represent the majority of IC I verbs and will be comprised under the label of IC I<sub>S</sub>. Here is where (morpho-)syntactic information is required. In fact, in clear contrast to IC I<sub>W</sub> verbs where only the weak suffix is adopted IC I<sub>S</sub> verbs select either suffix in dependence of the (morpho-)syntactic context in which they occur.

How can we account for the peculiar situation found in Titsch? The result of our corpus-based investigation can be summarized by means of the following Strong Verb Rule (= SVR):

**SVR:** In verbs belonging to IC I<sub>S</sub> the strong suffix is used when the past participle is found in the context of morphosyntactic agreement; otherwise, the weak suffix is used.

Clearly, morphosyntactic agreement needs to be specified with regard to the contexts seen above, namely (i) the passive construction and (ii) the BE-perfect in a partial way for the adverbial position, and (iii) the attributive, (iv) the appositive and (v) the small clause construction for the adnominal position. Note that as an expression of contextual inflection the agreement markers are placed after the past participle markers expressing inherent inflection, namely tense (see Booij 1993 for the distinction). In other words, inherent inflection, i.e. the selection of the strong or the weak suffix to encode tense, is sensitive to contextual inflection, i.e. type of syntactic environment in which the participle is employed. This peculiar property displayed by all past participles independently of their ICs, as shown in Table 10 above, is the crucial factor for determining the selection of the strong suffix. Thus, it is not that the strong suffix is sensitive to agreement, because the latter is also found with the weak suffix.

In addition, the weak suffix clearly serves as the default suffix because it appears when no contextual inflection, i.e. agreement, occurs. Accordingly, the verbal paradigm of a strong verb must be provided with two slots for the past par-

ticiples which stand in competition and in fact in complementary distribution, in accordance with the SVR.

This state-of-affairs might also be taken as an instance of syncretism insofar as the marker of the  $I_W$ -class is spread to a slot of the  $I_S$ -class, amounting to a lexical split – intended as a case of segmentation of a paradigm into parts determined on a (morpho-)lexical basis – of the kind advocated by Corbett (2015).

**Table 13:** Syncretism in the Titsch past participles.

		Types	[+AGR]	[-AGR]
IC I [ThV = -e]	$I_S$ -class	73	(-e)-n	(-e)-t
	$I_W$ -class	10		(-e)-t
IC II		105		-t
IC III [ThV = -ò]		58		(-ò)-t

In Table 13, the figures refer to verbs for which both inflected and uninflected past participles are found in the corpus. This excludes all verbs whose participles are not found in agreement environments in the corpus.

In sum, the presence of agreement as an instance of contextual inflection, which is a conservative property of Titsch as seen above, has been recast as a crucial feature for deciding about IC membership. Thanks to the functionalization of the new feature [AGR] for IC membership – an  $I_S$ -class can now be distinguished from an  $I_W$ -class. Viewed from a diachronic perspective, the functionalization of the feature [AGR] guarantees the survival of the strong class. In the light of the systematic distribution observed in Table 13 and of the examples of extension to etymologically weak verbs, this functionalization can be treated as an exaptation, i.e. the reuse of morphological junk devoid of any evident function (cf. Lass 1990; Gaeta 2016 for discussion).

Such a reuse can also be observed in other parts of verbal ICs, namely in the RU verbs discussed in Table 8 above. Recall that the latter are etymologically weak verbs that happened to display a root-vowel alternations resembling the AB classes of IC I. The class of RU verbs has functionalized the root-vowel alternation in connection with the occurrence of agreement in the past participle, as shown by Table 14.

**Table 14:** RU verbs and past participle agreement.

IC II	+RU, +AGR	+RU, -AGR	-RU, +AGR	-RU, -AGR
IC IIb (RU-1: <i>e/a</i> )	22	1	–	3
IC IIc (RU-2: <i>ie/ue</i> )	4	3	2	–
IC II d (RU-3: <i>é/ò</i> )	4	1	1	–
IC II e (RU-4: <i>é/oa</i> )	2	–	–	–
Tot.	32	5	3	3

The first left column (+RU, +AGR) refers to the RU alternations described in Table 4 above. In the second column (+RU, –AGR) the figures refer to the verbs found in the corpus displaying RU alternations in concomitance of a past participle displaying agreement, while in the absence of agreement no alternation is found, as in the following examples in (10).

- (10) a. *darnoa hämmo d'Meschtre pteckt*  
 afterwards has.one DEF=manure[F] cover.PST.PTCP  
 'Afterwards, one has covered up the manure'.  
 b. *dass z'ganz-a sigge ptack-z*  
 that DEF=whole-N.SG be.SUBJ.3SG cover-N.SG  
 'that the whole thing be covered up'.

In the other columns, deviant cases are found in which a participle shows RU alternation but no agreement in the expected context (+RU, –AGR), or no RU alternation but agreement in the expected context (–RU, +AGR), or finally no RU alternation and no agreement in the expected context (–RU, –AGR). Apart from these few deviations, the figures speak clearly in favor of a context-sensitive distribution of the RU alternation, which however follows the same pattern covered by the SVR seen above for IC I<sub>5</sub> verbs. Namely, when the environment requires agreement, the RU alternation shows up; elsewhere the infinitive root-vowel occurs. The functionalization of the root-vowel alternation in concomitance with the occurrence of the feature [+AGR] resembles another interesting change observed in another WG variety that is now extinct, namely Saley (see Dammal 2011: 248–259 for data and discussion). In this variety, past participles displaying agreement are used to express resultative value in contrast to the uninflected participles which simply carry a past value (cf. Dammal 2011: 256).

- (11) a. [dɪft̪ muiɾu ɪft̪ tserfallæ fægæm vættær]  
 DEM.F.SG wall[F] is collapse.PST.PTCP because.of whether  
 ‘This wall has collapsed because of the thunderstorm’.
- b. [dɪft̪ muiɾu ɪft̪ tserfalln-ɪ]  
 DEM.F.SG wall[F] is collapse.PST.PTCP-F.SG  
 ‘This wall has collapsed’.
- c. [nʊŋ heft̪s tsærdræt, ts mæsser –  
 now have.2SG=it twist.PST.PTCP DEF.N.SG knife[N]  
 dʊmna: makft̪s tsærdra:t-s hæ]  
 afterwards can.2SG=it twist.PST.PTCP-N.SG have.INF  
 ‘Now you’ve twisted it, the knife – afterwards you can have twisted it’.

In contrast to the resultative value of inflected past participles, the construction containing the uninflected participle focuses on the process leading to the resultant state. Note that the past participle displaying agreement also displays root-vowel alternation (see [tsærdra:t-s] in (11c)), while this is not the case when no agreement occurs (see [tsærdræt]). In other words, the class of RU verbs has functionalized the root-vowel alternation in connection with the occurrence of agreement in the past participle, triggered in its turn by the feature [resultative].

## 7 Conclusion

To conclude, far from merely being the result of the contact of German plus Italian, WG varieties display significant examples of the rise of secondary inflectional features, that exploit in an original way old pieces of junk coming from the Germanic repertoire in combination with contextual information available in morphosyntactic constructions in which past participles display agreement. In particular, I must stress the relevance of such a small language such as Titsch for discovering developments unprecedented within the Germanic family. This was also possible thanks to the text corpus which could be collected and elaborated in a data-base. Combining these methods with traditional fieldwork is a promising research perspective which allows us to document, preserve, and make available for research the invaluable cultural heritage represented by language islands. I am confident that these tiny phenomena will be of interest to the Festschriftee whose works has largely contributed to a better understanding of the intriguing factors involved in language contact, grammaticalization. . . and junks: *Ad multos annos!*

## Abbreviations

AB	Ablaut
AGR	agreement
Ap	appositive position
At	attributive position
B-Pa	BE-passive
B-Pe	BE-perfect
C-Pa	COME-passive
DAT	dative
DEF	definite
DEM	demonstrative
F	feminine
GEN	genitive
G-Pa	GO-passive
H-Pe	HAVE-perfect
IC	inflectional class
IND	indicative
INDEF	indefinite
INF	infinitive
M	masculine
MSG	Modern Standard German
N	neuter
NEG	negation
OBL	oblique
OHG	Old High German
PASS	passive
PERF	perfect
PL	plural
PP	past participle
PS	person
PST	past
PTCP	participle
REL	relative
RU	<i>Rückumlaut</i> (backwards metaphony)
S	strong
SC	small clause construction
SG	singular
SUBJ	subjunctive
SVR	Strong Verb Rule
ThV	thematic vowel
W	weak
WG	Walser German

## References

- Angster, Marco & Livio Gaeta. 2018. Wie kurz sind die Kurzverben? Morphologische Merkmale in Gressoney und Issime. *Germanistische Linguistik* 239–240. 211–237.
- Bohnenberger, Karl. 1913. *Die Mundart der deutschen Walliser im Heimattal und in den Außenorten*. Frauenfeld: Huber.
- Booij, Geert. 1993. Against split morphology. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1993*, 27–49. Dordrecht: Kluwer.
- Braune, Wilhelm & Ingo Reiffenstein. 2004. *Althochdeutsche Grammatik I. Laut- und Formenlehre*. 15<sup>th</sup> ed. by Ingo Reiffenstein. Tübingen: Niemeyer.
- Corbett, Greville G. 2015. Morphosyntactic complexity: A typology of lexical splits. *Language* 91. 145–193.
- Dal Negro, Silvia. 2004. *The Decay of a language: The case of a German dialect in the Italian Alps*. Bern: Peter Lang.
- Dammel, Antje. 2011. *Konjugationsklassenwandel. Prinzipien des Ab-, Um- und Ausbaus verbalflexivischer Allomorphie in germanischen Sprachen*. Berlin: De Gruyter.
- Eufe, Rembert & Mader, Anna. 2018. Das Walserdeutsche im deutschen und italienischen Sprachgebiet. In Nicole Eller-Wildfeuer, Paul Rössler & Alfred Wildfeuer (eds.), *Alpindeutsch*, 113–138. Regensburg: vulpes.
- Fertig, David. 2020. Verbal inflectional morphology in Germanic. In Michael T. Putnam & Richard B. Page (eds.), *The Cambridge Handbook of Germanic Linguistics*, 193–213. Cambridge: Cambridge University Press.
- Fleischer, Jürg. 2007. Zur Herkunft des flektierten prädikativen Adjektivs im Höchstalemannischen. *Zeitschrift für Dialektologie und Linguistik* 74. 196–240.
- Gaeta, Livio. 2016. Co-opting exaptation in a theory of language change. In Muriel Norde & Freek Van de Velde (eds.), *Exaptation in language change*, 57–92. Amsterdam / Philadelphia: John Benjamins.
- Gaeta, Livio. 2018. Im Passiv sprechen in den Alpen. *Sprachwissenschaft* 43. 221–250.
- Gaeta, Livio. 2020. Remotivating inflectional classes: an unexpected effect of grammaticalization. In Bridget Drinka (ed.), *Historical Linguistics 2017. Selected papers from the 23rd International Conference on Historical Linguistics*, 205–227. Amsterdam & Philadelphia: John Benjamins.
- Gaeta, Livio. 2023. Phonology and conventionalization: Naturalness and beyond in German(ic) sonorants. In Jennifer Brunner, Petra Hödl, Veronika Mattes & Thomas Schwaiger (eds.), *Für Bernhard Hurch*. Special issue of *Grazer Linguistische Studien* 94. 121–134.
- Gaeta, Livio, Marco Angster, Raffaele Cioffi & Marco Bellante. 2022. Corpus linguistics for low-density varieties. Minority languages and corpus-based investigations. In Maria Grossmann, Fabio Montermini, Diana Passino, Diego Pescarini & Florence Villoing (eds.), *Corpus et données en morphologie*. Special Issue of *Corpus* 23; available at <https://journals.openedition.org/corpus/7345>.
- Gaeta, Livio, Raffaele Cioffi & Caterina Saracco. In press. Korpuslinguistik am Beispiel der walserdeutschen Sprachinseln in Italien. Sprachkontakt, Spracherhaltung, Sprachwandel. In Angélika Prediger (ed.), *Deutsche Sprachminderheiten in der Welt: Empirische Studien zur Sprachvariation und Sprachideologie*. Berlin: de Gruyter.
- Gaeta, Livio, Caterina Saracco & Raffaele Cioffi. 2024. *Greschòneytsch. Nuovo Dizionario Titsch – Italiano / Tedesco*. Alessandria: Edizioni dell'Orso.
- GrW 1998 = Walser Kulturzentrum (ed.), *Greschòneytsch Wörterbuch: Vocabolario Titsch / Italiano*. Quart: Musumeci.

- Hotzenköcherle, Rudolf. 1956. Umlautphänomene am Südrand der Germania. In Elisabeth Karg-Gasterstädt & Johannes Erben (eds.), *Fragen und Forschungen im Bereich und Umkreis der germanischen Philologie. Festgabe für Theodor Frings*, 221–250. Berlin: Akademie-Verlag.
- Lass, Roger. 1990. How to do things with junk: Exaptation in linguistic change. *Journal of Linguistics* 26. 79–102.
- Moulton, William G. 1941. *Swiss German Dialect and Romance Patois*. Baltimore: Linguistic Society of America.
- Rizzi, Enrico. 1993. *Storia dei Walser*. Anzola d'Ossola (Novara): Fondazione Arch. Enrico Monti.
- Russ, Charles V. J. 1990. High Alemannic. In Charles V. J. Russ (ed.), *The Dialects of Modern German: A Linguistic Survey*, 364–393. London: Routledge.
- Vennemann, Theo. 1986. Ruckümläut [sic]. In Dieter Kastovsky & Aleksander Szwedek (eds.), *Linguistics across Historical and Geographical Boundaries*. Vol 1: *Linguistic Theory and Historical Linguistics. In Honour of Jacek Fisiak on the Occasion of His Fiftieth Birthday*, 701–723. Berlin: Mouton de Gruyter.
- Zürrer, Peter. 1982. *Wörterbuch der Mundart von Gressoney. Mit einer Einführung in die Sprachsituation und einem grammatischen Abriss*. Frauenfeld: Huber.
- Zürrer, Peter. 1999. *Sprachinseldialekte. Walserdeutsch im Aostatal (Italien)*. Aarau: Sauerländer.
- Zürrer, Peter. 2009. *Sprachkontakt in Walser Dialekten. Gressoney und Issime im Aostatal*. Stuttgart: Steiner.
- Zürrer, Peter. 2011. Conservazione e mutamento nei dialetti walser in Italia. In Walter Breu (ed.), *L'influsso dell'italiano sul sistema del verbo nelle lingue minoritarie*, 101–121. Bochum: Brockmeyer.



Craig Alan Volker

# The convergence of Nalik with Tok Pisin: Two languages becoming one linguistic repertoire

**Abstract:** The Nalik language of New Ireland, Papua New Guinea has been increasingly heavily influenced by universal bilingualism in Tok Pisin and, to a much lesser extent, the English-medium education system. This can be seen in the loss of the phoneme /ɣ/, the adaptation of Tok Pisin-like morphological and syntactic constructions, the shrinkage of the Nalik lexicon, changes in the semantic range of some words to be more like Tok Pisin, and a more direct pragmatic environment in public speaking. This merging of Nalik with Tok Pisin must be taken into account in any Nalik maintenance or revitalisation efforts.

**Keywords:** Tok Pisin, Nalik, convergence, Papua New Guinea, bilingualism

## 1 Introduction

Papua New Guinea, by far the largest South Pacific island nation, is well known for having the largest number of languages of any nation in the world, 840 for a population of approximately nine million according to the 2024 edition of *Ethnologue* (Eberhard et al. 2024). Of these, 25 indigenous languages, as well as Tok Pisin (an English-based pidgin-creole language and lingua franca), English, several Chinese languages, and numerous languages spoken by migrants from other Papua New Guinea provinces are spoken by the residents of New Ireland Province in the Bismarck Archipelago in the nation's northeast, who numbered 232,351 in 2021 (Igitoi 2021), very few of whom are monolingual. This study<sup>1</sup> will look at language contact

---

<sup>1</sup> Unless otherwise stated, the data on which this study is based come from observations made in Madina and nearby Nalik-speaking villages during my parttime residence there since 1989.

---

**Note:** I would like to thank Professor Thomas Stolz for his many probing questions about language use in Papua New Guinea during my time some years ago as a visiting professor at the University of Bremen, a catalyst that eventually led to my writing this chapter.

---

**Craig Alan Volker**, The Cairns Institute, James Cook University / Beijing Foreign Studies University, e-mail: craig.volker@jcu.edu.au

in the homeland of the speakers of one of these languages, Nalik, and examine what effect this language contact, in particular the universal acquisition of Tok Pisin and, through education, exposure to English, is having on the generation of Nalik children now growing up in the third decade of the twenty-first century. This will help us to anticipate what the rapidly changing linguistic ecology of the area might look like in the rest of this century and in turn, assist elders in the community decide what, if any, measures should be taken to deal with this new linguistic ecology.

Nalik is an Austronesian language and closely related to other languages in northern New Ireland, Djaul, and Lavongai (New Hanover) islands in New Ireland Province in what Ross (1988: 291) has called the Tungag/Nalik Network of New Ireland languages. The traditional Nalik homeland is in the Tikana Local Level Government Area of northern central New Ireland in a band across this long and narrow island. Nalik settlements today are located along the coasts, with five Nalik villages on the west coast and eleven on the east coast, all 80 to 120 kilometres southeast of the provincial capital, Kavieng. In 2015 an estimated 5000 people considered themselves Nalik with an estimated 4000 of them fluent or semi-fluent speakers of the language (Volker 2015). Many of them live outside the traditional Nalik homeland, and many of the people living in the Nalik area are immigrants from other parts of Papua New Guinea who sometimes, but not always, have a passive knowledge of Nalik, but generally do not have an active knowledge of the language.

The last reported monolingual speaker of Nalik was an elderly man in a west coast village who died in 1988 (Volker 1998). Today in addition to the English that they learn in school, all Nalik people speak Tok Pisin, which has become the dominant language for most younger people, especially those living on the east coast, where villages have easy access to transportation to Kavieng, and where many migrant workers have settled.

Clan affiliation in New Ireland Province is based on matrilineal genealogies, is far more important than any linguistic identity, and often crosses linguistic borders and therefore creates a tolerance of language shift and multilingualism. In addition to its use in its traditional homeland, Nalik is spoken, or at least passively understood, as a second language by many speakers of closely related Lakurumau to the north of the Nalik area and by many speakers of Kuot, the only indigenous non-Austronesian language in New Ireland Province, to the south. Both of these languages have far fewer speakers than Nalik. Speakers of all three languages belong to the Malagan cultural area of northern New Ireland Province, with shared customary beliefs, clan structures, rituals, and art. Oral history suggests that Kuot was previously spoken in a much larger area than it is today, and that language shift from Kuot to Nalik has been occurring since before the first imposition of colonial

control in 1884. Many Nalik speakers have knowledge of their Kuot roots and some trace their matrilineal clan home to areas where Kuot is still spoken.

Tok Pisin, now spoken by almost everyone in New Ireland, is an English-based pidgin language, with its genesis on the multiethnic trading vessels of the nineteenth century South Pacific and the “blackbirding” indentured labour system that brought Melanesians from many Austronesian language backgrounds to work for years on plantations in Queensland and German Sāmoa at the end of the nineteenth century. This early form of Melanesian Pidgin English was brought to the Bismarck Archipelago in what was then German New Guinea by returning labourers, where it was heavily influenced by the Kuanua language of the Gazelle Peninsula of New Britain and lexically by German.<sup>2</sup>

English, a legacy of Australian colonial rule from the end of World War I to 1975, is also part of contemporary life in New Ireland, as it is the language of instruction at all levels of public education in Papua New Guinea. New Ireland was one of the first areas of Papua New Guinea where Western formal education was developed in the colonial era. This has resulted in much higher levels of primary and secondary school attendance today than in the country as a whole. Even though the estimated national adult literacy rate in 2024 is only 65% (CIA 2024), it is virtually 100% among adult Naliks and, unlike much of Papua New Guinea, primary school education is universal in Nalik villages and among diaspora Naliks living in urban areas. For several decades, a majority of Nalik youth has gone on to at least some years of secondary education, usually at multiethnic boarding schools where no one language group is dominant in the student population. This means that virtually all Nalik youth and adults alive today have at least a passive understanding of basic English and many have spent their formative teenage years living in boarding schools where English was used in class and where Tok Pisin was the dominant language in student dormitories.

Because it is the medium of education, English is also the language of most formal written communication in the national and provincial governments and is the language of the only two national newspapers in the country. While most people retain a passive command of basic English after school, usually enough to read the daily national newspapers, individuals’ ability to produce enough English to use communicatively varies greatly. It is not normally used as a spoken language in New Ireland outside of schools except with visiting tourists and other foreigners.

---

<sup>2</sup> See Mosel (1980) for an analysis of the relationship between Kuanua and Tok Pisin and Volker (forthcoming) for an overview of the history of, and relationship between, Tok Pisin and English in Papua New Guinea.

Tok Pisin is the normal medium of communication with non-Nalik people, even with those who reside in Nalik villages or who are married to Nalik partners.

## 2 Today's rapidly changing Nalik linguistic ecology

Today all Nalik children learn Tok Pisin as infants. This childhood acquisition of Tok Pisin is a relatively new phenomenon. Although New Ireland was one of the first areas where the use of Tok Pisin became widespread, which is reflected in the large number of words with New Ireland origins in Tok Pisin, until recently Tok Pisin was a language learned in adolescence or adulthood and used mainly for interethnic communication. This changed with the rapid changes in mobility and technology Papua New Guinea experienced after attaining independence from Australia in 1975. A growing number of Naliks were able to take advantage of their relatively high levels of formal education and obtained employment outside their traditional homeland, living in urban areas for part or even all of their lives. As a result of this and the large number of Naliks meeting future partners while attending multiethnic secondary and tertiary institutions, the number of interethnic marriages has increased. Children who grow up in these families tend to speak Tok Pisin as their dominant home language, even if the family returns to the Nalik homeland. Today there would be only a handful of families who use mainly Nalik with their children, although there are still a number of families where there is frequent code-switching between Nalik and Tok Pisin. Even families where both parents and all four grandparents are Nalik often use Tok Pisin with their children today. One grandmother told me she does this because “Tok Pisin is easier for children. They can learn Nalik when they are older.” This rarely happens however.

When I first moved to the Nalik area in 1989, I lived next to a village primary school. During recess time, almost all the children playing outside my window used Tok Pisin, and it was rare to hear children playing with each other in Nalik. Nevertheless, almost all could speak in Nalik when addressed in Nalik by an adult, and I could use Nalik to interview all the children in the grade 5 and 6 classes about their daily life. This survey of grade 5 and 6 children from two Nalik villages who attended Madina Primary School showed that Tok Pisin was already the dominant home language in most families, as only about a quarter of children reported using “mainly Nalik” at home.

Today I could not conduct a survey with primary school children speaking only in Nalik. On several occasions in the past several years when I have had the opportunity to address primary school assemblies, I have found that I have had to use Tok

Pisin or English, as a sizeable minority of children no longer have even a passive knowledge of Nalik. Further evidence of this rapid loss of the ability to use Nalik is the fact that I am well known in the village where I live as a promoter of the use of Nalik, and to please me, children will sometimes try to initiate a conversation with me in Nalik. Many times, however, they are unable to proceed very far and end up giggling and having to switch to Tok Pisin. For some children, their grasp of Nalik is so weak that they will greet me with the wrong greeting when they come across me in the street (e.g., saying the Nalik equivalent of “Good morning” when it is the afternoon).

The relationship between the Nalik and Tok Pisin languages in contemporary Nalik society can be explained by what Fred Anderson (2015) has described as a continuum of linguistic power in multilingual societies, with “power” being related to the relative social strength each language has both in the community as a whole and in the linguistic repertoire of individual speakers. While all languages have a capacity for social interaction and creativity, they differ greatly in the degree to which members of a society use them in daily life and for the purposes for which they are the means of communication. There is always a tendency for languages that are used by more people and for more purposes to have the greater social strength and to crowd out languages with lesser social strength.

In the Nalik area, it is obvious that Tok Pisin is what Anderson (2015) calls the “language of greater power” and Nalik the “language of lesser power”. Tok Pisin is used for a majority of social interactions and the number of social domains where Nalik is used either alone or together with Tok Pisin is declining. While the concept of “linguistic imperialism” has generally been used in discussions of the dominance of English over national and regional languages (e.g., Phillipson 1992), it can also be used here, where the national lingua franca, Tok Pisin, is dominating the local language, Nalik, in an unequal power play. Here the sociolinguistics of power proposed by Blommaert (2010) to describe discourse between people speaking different languages is relevant. In Blommaert’s paradigm, people choose among the linguistic resources available to them in their linguistic repertoire and choose those most appropriate for the level (*scales* in his terminology) to which they aspire or in which they operate. Tok Pisin is a language of mobility and symbolises the world beyond the small area where Nalik is spoken, and so one cannot be surprised that it has greater sociolinguistic power than Nalik and is therefore chosen more often than Nalik.

This situation is not unique to Papua New Guinea, as there are many countries where languages of lesser power and languages of greater power are spoken in the same society. In these societies not only is there a tendency for languages of lesser power to be dominated by languages of greater power and to be used less often (see Volker and Anderson (2015) for examples of this in the Asia-Pacific region), there is

a tendency for bilingual and multilingual individuals to try to make the grammars and semantics of the languages they use converge.

This phenomenon has received particular attention by linguists in India. The convergence of the grammars and phonologies of the varieties of languages spoken in adjacent areas but belonging to separate languages families was described as early as 1971 for a community in southwestern India (Gumperz and Wilson 1971) and more recently for other areas in India, such as by Roy and Subbārāo (2023) for northeastern India. These researchers report that there is a tendency for bi- and multilingual speakers to use the same grammatical patterns or phonological systems in all the languages they speak. Where this is done over time and by enough speakers, this can result in language change in one or more of the languages involved so that their grammars, phonologies, or semantics can become more similar. Theoretically, this could eventually result in the different languages having nearly identical phonologies, grammars, and semantics and being distinguished only by the lexical items used to express these nearly identical systems.

Given the current sociolinguistic situation in New Ireland, we can expect the phonologies, grammars, and semantics of Tok Pisin and Nalik to converge in a similar way. Because of the dominance of Tok Pisin in relation to Nalik and the role of Tok Pisin as a medium to communication beyond the immediate local area, we can also expect that it is more likely that Nalik will adopt Tok Pisin features than the reverse. Moreover, because of most persons' low levels of English proficiency and limited opportunities outside of school to use English, we should not expect English to exert a strong influence on the phonologies, grammars, and semantics of the other two languages. As the analysis that follows will show, these expectations are indeed valid.

### 3 Phonology

The phonological systems of both Nalik and the New Ireland dialect of Tok Pisin phonology are still poorly described. Nevertheless, to even a casual observer, it is obvious that some young children are not acquiring /ɣ/, a prominent Nalik phoneme that does not exist in Tok Pisin, but they are acquiring /z/, another prominent Nalik phoneme that does not exist in Tok Pisin, but that is taught in primary school as an important English phoneme.

No study has been published of the phonology of the New Ireland variety of Tok Pisin. The most comprehensive description of the phonology of Tok Pisin at a national level is still Laycock (1985). A basic overview of Nalik phonology has been

described in Volker (1998). While recognising the lack of specifically New Ireland focus in the former study and the lack of detail in the latter, these can still be used as a basis for examining the speech of children today.

Laycock (1985) has described what he calls the “Tok Pisin core phonology” (i.e., the phonology that is shared by most geographic and social varieties of Tok Pisin) in his Table 2, which is reproduced in Table 1 in which I have added the relevant Standard Tok Pisin orthographic representations and, where contrastive, information about voicing in square brackets.

**Table 1:** Tok Pisin consonants (adapted from Laycock 1985: 297).

I	stops [unvoiced]	p [P, p]	t [T, t]	k [K, k]
II	stops [voiced]	b [B, b]	d [D, d]	g [G, g]
III	Clusters [unvoiced]	mp [mp]	nt [nt]	ŋk [nk]
IV	clusters [voiced]	mb [mb]	nd [nd]	ŋg [ng]
V	nasals	m [M, m]	n [N, n]	ŋ [ŋg]
VI	continents	w [W, w]	l [L, l]	y [Y, y]
VII	continents	-	s [S, s]	h [H, h]
VIII	flaps	-	r [R, r]	-
IX	fricatives [unvoiced]	f [F, f]	ʃ [S, s]	-
X	fricatives [voiced]	v [V, v]	ʒ [S, s]	-

Using the information in Volker (1994), we can draw a similar chart in Table 2 with the same terminology to represent the Nalik consonants. It should be pointed out that in Nalik there is a very small number of words that differentiate minimal pairs of words between the pairs of sounds written as ⟨f/v⟩, ⟨s/z⟩, ⟨p/w⟩, ⟨t/r⟩, and ⟨k/x⟩, but in most – but not all – words they are allophones, with the first of each pair of sounds tending to be used when adjacent to a consonant other than /l/ and the second sound tending to be used between vowels or between a vowel and /l/. The rules governing this are complex and not yet fully understood, as there are numerous individual lexical items for which this rule does not apply, and the rule can be reversed for emphasis. For all of these reasons, the fact that there are a small number of words for which they differentiate between meaning, and because native speakers were adamant that they should be represented as separate letters when the Nalik orthography was being settled in the 1980s, they have been analysed in Volker (1994) and subsequent work as separate phonemes.

**Table 2:** Nalik consonants (from Volker 1998).

I	stops [unvoiced]	p [P, p]	t [T, t]	k [K, k]
II	stops [voiced]	β [B, b]	d [D, d]	g [G, g]
III	Clusters [unvoiced]	-	-	-
IV	clusters [voiced]	mb [mb]	nd [nd]	ŋg [ŋg]
V	nasals	m [M, m]	n [N, n]	ŋ [ŋg]
VI	continuents	w [W, w]	l [L, l]	y [Y, y]
VII	continuents	-	s [S, s] voiceless z [Z, z] voiced	ɣ [X, x] voiced
VIII	flaps	-	r [R, r]	-
IX	fricatives [unvoiced]	f [F, f]	ʃ [S, s]	-
X	fricatives [voiced]	v [V, v]	ʒ [S, s]	-

Given the linguistic power differential between Tok Pisin and Nalik and universal Tok Pisin – Nalik bilingualism, we would expect that any convergence between the two languages to be for Tok Pisin to have an influence on the Nalik phonological system rather than the reverse. This is indeed the case among many young children today.

Comparing Tables 1 and 2, we see that Nalik has two consonants that are absent in Tok Pisin: both what Laycock would label voiced continuents: a voiced coronal sibilant /z/ and a voiced velar fricative ɣ (written ⟨x⟩). I have not heard any children having difficulty producing /z/ in Nalik names or words, possibly because this sound is present in English and taught in school from grade one onwards.

This is not the case, however, with /ɣ/, which is common in Nalik, but has no English equivalent. I have heard many pre-adolescent children being unable to produce /ɣ/ in Nalik, even in the common greetings which tend to end in *doxo* ‘good’ (e.g. the equivalents of *good morning*, *good midday*, *good afternoon*, and *good evening*). These children usually substitute the voiceless glottal fricative /h/ – which does appear in Tok Pisin and English, but in Nalik only in loanwords – for the voiced velar fricative /ɣ/, e.g., *paanaraan doho* for *paanaraan doxo* ‘good morning’. While a majority of children can still produce /ɣ/, I do notice that for an increasing number of children, this is no longer possible.

Interestingly, before the establishment of a fixed orthography for Nalik in the 1980s, many Nalik people used the letter *h* to represent /ɣ/ when writing Nalik. At that time almost all Nalik persons had a good command of Nalik as well as at least primary English-medium school education. Since the letter *h* was not otherwise needed to write Nalik, the choice of this letter representing a fricative sound made

in the back of the mouth was an obvious choice to represent /ɣ/. Some people still spell their Nalik names with *h* rather than *x* for this reason.

There does not yet seem to be a convergence of the vowel systems of the two languages. Both Tok Pisin and Nalik have a five-vowel system, but unlike Tok Pisin, vowel length in Nalik can be phonemic, most notably with /a/ and /a:/, written ⟨a⟩ and ⟨aa⟩, respectively. If the vowel systems converged, we would see speakers confusing common Nalik words that are differentiated only by vowel length, such as *laraf* ‘yesterday’ and *laraaf* ‘afternoon’, or converting them into homonyms. This is not yet the case.

## 4 Morphology

In contemporary Nalik there are a number of morphological constituents that are disappearing in the speech of younger speakers. These are all distinctions, morphological categories, or irregularities that do not have parallels in Tok Pisin.

In the noun phrase this can be seen in the disappearance of the irregular plural marker *fu*. Both Nalik and Tok Pisin mark plurality with prenominal particles, *mun* and *ol*, respectively. In Nalik a small number of female humans, however, use the irregular marker *fu* instead of *mun*. Many younger speakers, however, use *mun* with these nouns or else use both together.

It can also be seen in the disappearance of the Nalik dual marker *u*, probably originally a shortened form of the numeral *uru(a)*<sup>3</sup> ‘two’. Unlike numerals, which appear in postnominal positions, *u* takes the prenominal position of the article *a*. Tok Pisin does not have a dual marker distinct from the numeral *tupela* ‘two’, which, like Tok Pisin adjectives, has a prenominal position. While, as described below, most Nalik speakers use Tok Pisin numerals, for a minority of those who still use Nalik numerals for small numbers, *uru(a)* appears in a prenominal position, like its Tok Pisin equivalent.

The influence of Tok Pisin can be seen with the disappearance in the speech of many young Nalik speakers of the difference between alienable and inalienable possession, which Blust (2013: 484) has called a “nearly universal” characteristic of Oceanic languages. In Nalik the possession of terms relating to genetic kinship, the deity, physical and spiritual body parts, and traditionally owned land have traditionally been marked as inalienable by using inalienable suffixes rather than separate possessive pronouns (see Volker 1998). This distinction does not exist in Tok Pisin, nor in the speech of many younger speakers of Nalik.

---

3 In the northeastern dialect *uru*, elsewhere *urua*.

The loss of this distinction seems to have spread unevenly through the lexicon. In the 1990s I observed that in Madina Community School, children from Madina Village no longer used inalienable possessive suffixes with *iza* ‘name’, whereas children from neighbouring Luaupul Village at the same school did. At that time all children did use inalienable suffixes with words denoting members of their immediate family (the equivalents of ‘mother’, ‘father’, ‘grandparent’, etc.). Today even this use has almost completely disappeared in the speech of those children who still speak Nalik. Similarly, even adults rarely use inalienable possessive suffixes when referring to the possession of customary land or names of the deity, such as *Nakmai* ‘God’ or *Piran* ‘Lord’. With the commonly used word *das* ‘brother’, for many speakers the third person inalienable possessive suffix *-na* has been fused onto the root, with many speakers using the alienable possessive pronouns equivalent to ‘my’ and ‘your’ with *dasna*. For these speakers, as in Tok Pisin, there is no longer a marked category of inalienable possession in Nalik.

In the verb phrase both Tok Pisin and Nalik mark durative aspect by preverbal particles. In Tok Pisin this is with *save* (ultimately from Portuguese *sabe* ‘knows’), while Nalik has traditionally used *i* before intransitive verbs and *t* before transitive verbs. Today many Nalik speakers no longer use *t*, so that as in Tok Pisin, the distinction between transitive and intransitive verbs is no longer made. The Nalik durative marker *i* has the same form as the Tok Pisin predicate marker *i*, which is undoubtedly the motivation for the choice of *i* rather than *t* as the sole remaining Nalik durative marker.

While transitive durativity may no longer be marked, transitivity itself is still marked in Nalik, as it is in Tok Pisin. For many Nalik verbs, transitive verbs are formed by adding the transitive suffix *-ing* to the intransitive verb (e.g., intransitive *rexaas* ‘to know [about something]’ and transitive *rexaasing* ‘to know [something]’). This is similar to Tok Pisin, in which almost all transitive verbs are marked with the suffix *-im*, which is added to an intransitive verb to form its transitive equivalent (e.g., *rit* ‘to read [intransitive]’ and *ritim* ‘to read [something]’). When loanwords that are transitive are absorbed into Nalik, they are marked with the Tok Pisin *-im* suffix and not the indigenous *-ing* transitive suffix, even if they are relatively recently acquired words related to government bureaucracy or technology that have been taken directly from English and not Tok Pisin (e.g., *transferim* ‘to transfer [students or workers]’ and *downloadim* ‘to download [something]’). This is a strong indicator of the dominance of Tok Pisin grammar in the linguistic repertoire of bi- and trilinguals.

Another verbal category that has almost completely disappeared in Nalik is the past tense locative copula *vinai*. This has been used optionally for sentences in the past where the predicate is a locative (e.g., the equivalent of ‘she was at the market’). *Vinai* is the only copula form present in Nalik; there are no present tense

and no non-locative copulas. Tok Pisin does not have a copula of any sort, which has been a motivation for the loss of *vinai* in Nalik. Today the use of *vinai* is very rare and we can assume that it will disappear completely with the current generation. This is another example of a morphological category disappearing in Nalik if there is no equivalent in Tok Pisin.

In their personal pronoun systems, Nalik and Tok Pisin differ in that traditionally, at least for many older speakers with high status, in Nalik the third person singular pronoun, *naan*, could not refer to a nonhuman. This rule seems to be a stylistic rule that these speakers have used to separate their speech from that of ordinary speakers. For those speakers who make this distinction, it is made no matter what the grammatical relation of the pronoun is (subject, direct object, or prepositional object), so that these speakers lack a third person singular pronoun for nonhuman referents. Today only a very few older clan leaders make this morphological distinction.

We can assume that the motivation for the disappearance of this distinction is the convergence of the Nalik personal pronoun system to that of New Ireland Tok Pisin. In the New Ireland dialect of Tok Pisin there is no distinction between the use of the third person pronoun *em* for human or nonhuman referents. In some other dialects of Tok Pisin, a human / nonhuman distinction does exist for some speakers, but it is only when the third person singular pronoun is the object of a preposition. In that position, *en* is used for a non-human referent instead of the otherwise default Tok Pisin third person singular pronoun *em*. This use of the variant *en* is a feature of the dialect spoken in Madang Province on the New Guinea mainland. That dialect is the basis of the standard written form of Tok Pisin used in the Tok Pisin translations of both the Bible and Bahā'ī scriptures with their many pronominal references to God, a nonhuman (e.g., *long En* 'to Him/Her'), so it is familiar to many New Irelanders through religious services. It is also familiar to many New Irelanders who have lived on the New Guinea mainland. A few Naliks do use *en* rather than *em* after prepositions, but this is overcorrected to apply to all referents, whether human or nonhuman. The presence of this distinction does not seem to be prevalent enough in the Tok Pisin of New Irelanders to motivate the retention of the distinction in Nalik.

## 5 Syntax

In the changes we have examined so far, the influence of English has been minimal or non-existent. When we look at the changes in the syntax used by younger Nalik-speakers, however, we can see the influence of English as well as Tok Pisin.

This affects the construction of comparatives and the lack of passive voice. Tok Pisin has further been the motivation for the modern construction of existential clauses.

The influence of English is particularly evident in changes in the construction of comparatives. Nalik has traditionally had two ways of forming the comparative of adjectives. Both involve the noun or pronoun being described, a construction ending in the transitive suffix *-ing* and as the direct object of this transitive verb the noun or pronoun with which the subject is being described. The first option uses the relevant adjective as the main predicate, followed by a second verb, *paamu* ‘first’ with the transitive suffix *-ing*, as in (1). The second uses the comparative verb *vu* as the predicate, followed by the relevant adjectival verb with the transitive suffix *-ing*, which is followed by a grammatical object, which is the pronoun or noun with which the subject is being compared, as in (2).<sup>4</sup>

(1) Nalik (own competence)

*A naalik ka doxo paamu-ing a fnalik.*  
 ART son 3SG good first-TR ART daughter  
 ‘A son is better than a daughter.’

(2) Nalik (own competence)

*A fnalik ka vu doxo-ing a naalik.*  
 ART daughter 3SG compare good-TR ART son  
 ‘A daughter is better than a son.’

Today neither of these constructions is in common use, especially among younger speakers. It is more common either to use the Tok Pisin / English loan *moa* ‘more’ followed by the relevant adjective (with no transitive suffix), which is then followed by the oblique preposition *pan*, the object of which is the noun or pronoun with which the subject is being compared, as in (3), or to avoid a syntactic construction entirely and use a phrase stating one noun or pronoun has a particular quality and the other does not, as in (4).

(3) Nalik (own competence)

*A naalik ka moa doxo pan a fnalik.*  
 ART son 3SG more good OBL ART daughter  
 ‘A son is better than a daughter.’

---

<sup>4</sup> The different genders in the subject and object in examples (1) and (2) are not relevant here.

## (4) Nalik (own competence)

*A fnalik ka doxo ma a naalik kavit.*  
 ART daughter 3SG good and ART son not  
 ‘A daughter is good and a son is not.’

With the use of *moa*, which has the same pronunciation as *more* in Papua New Guinea English, and the preposition *pan*, which rhymes with English *than*, the motivation for the type of comparison shown by (3) appears to be English and not Tok Pisin. The motivation for the second and more common option shown by (4) uses the same circumlocution normally used in Tok Pisin to express comparison (‘a daughter is good and a son is not’).

Another characteristic of modern Nalik is the lack of passive voice. While Nalik did have a passive construction, only a very few older clan leaders have any knowledge of how it works and even they rarely produce sentences using it. It remains only in phrases in a few older songs and oral literature. The motivation for its loss is probably the lack of a passive construction in Tok Pisin. This is reinforced by the weak command of the Standard English passive, which few speakers of Papua New Guinea English have. This is because passive constructions are usually not taught until students are in high school. Many children do not attend high school and even for those who do, passive constructions are often not assimilated into their everyday English usage. When used, the verb *be* is often omitted, so that sentences such as *tax returns rendered at the end of the year* are common, in which there is no surface marking of passive voice.

The construction of Nalik existential clauses has been strongly influenced by Tok Pisin. Three lexemes have been recorded to form Nalik existential clauses: *bawai*, *roxon*, and *gaat*. With *bawai* there is no subject marker and the existential verb *bawai* appears after the subject, as in (5). This construction does not have a parallel in Tok Pisin and has disappeared almost completely in modern Nalik.

## (5) Nalik (own competence)

*A raafin bawai.*  
 ART sardine EXIST  
 ‘There are sardines.’

The equivalent sentence with either *roxon* ‘have’ or *gaat* ‘have’ requires the third person subject marker *ka*, with the subject following the existential verb, as in (6).

## (6) Nalik (own competence)

*Ka roxon/gaat a raafin.*  
 3SG have ART sardine  
 ‘There are sardines.’

*Roxon* is the original Nalik word for ‘have’. It had almost disappeared by the 1990s in favour of *gaat*, a loanword from Tok Pisin (*gat*, from English *got*), but in a short revival of traditional Nalik by a prominent clan leader, Michael a Xomerang (also spelt Michael Homerang), that coincided with the development of a Nalik orthography and the writing of the first books in Nalik, the word was revived. It is now again in common use, although still not used as much as the Tok Pisin loanword *gaat*.

The Nalik construction with words meaning ‘have’ is a parallel of the equivalent of the Tok Pisin construction with *gat*, which also has the additional meaning ‘have’, as in (7).

## (7) Tok Pisin (own competence)

*I gat talai.*  
 SM/PM have sardine  
 ‘There are sardines.’

Tok Pisin *i* is usually described as a predicate marker, but Keesing (1988) has shown that it can also be interpreted as a subject marker inherited from its initial substrate Austronesian languages in which, as in Nalik, a preverbal subject marker is usually obligatory. With this interpretation, except for the absence of an article (which does not exist in Tok Pisin), the structure of the Nalik sentence in (6) is identical to that of the Tok Pisin sentence in (7).

## 6 Nalik lexical shrinkage

Lexical shrinkage refers to the replacement of indigenous words by loanwords from other languages, the loss of certain registers and the complete loss of lexical items, and the diminished use of a language in certain domains.

Before the introduction of Tok Pisin during the colonial era, Nalik was used in all lexical domains. For both young men and young women, the development of specialised vocabulary to describe the knowledge being imparted was an important part of their training. In the absence of writing, precision in oral communication was vital for the continuation of cultural and social life, and there was a rich lexicon for the discussion of spiritual concepts that adults were expected to know.

Until formal Western education was introduced during the colonial period and became universal after World War II, adolescent males lived together apart from their families in separate houses known as *haus boi* ('boys' house') in Tok Pisin, which belonged to their maternal clan. Young men in the *haus boi* were under strict discipline by elder men in their clan, often their maternal uncles, who combined a strict regime of intense physical labour, practical training in life skills (such as gardening, fishing, and combat), and clan lore (such as history, rhetorical skills, songs, and dance). This training culminated in their being taken into the bush for training in secret knowledge belonging to the clan, after which they were eligible for marriage and living as adult men. When clan elders recognised leadership skills in particular young men, they were taken aside for special training as apprentice sorcerers, assassins, carvers, or for eventual nomination as *maimai* (clan leaders).

Adolescent females were given similar training by elder clan matriarchs, although they continued to live with their families until marriage. This training emphasised traditional medicine, childrearing, gardening, nutrition, and female health (including birth control and abortion techniques), and like young men, they were also instructed in secret clan knowledge. Since New Ireland societies are matrilineal and land ownership is held by women, particular attention was paid to information related to land and to the history of land transfers and obligations between clans.

As primary education became universal in the years following World War II, an increasing number of young people attended the two high schools located in the Nalik-speaking area, often as boarding students. These young people have no longer been under the continual control of clan elders and no longer had time to acquire the same amount of indigenous knowledge that earlier generations had. Today many clans no longer have *haus boi*, and even where they do, they are more like clan meeting houses than places where young men gather to be taught how to behave and speak. At the same time, young women spend much of their time with their peers rather than with their grandmothers and clan matriarchs. Many elders claim this lack of traditional training for adolescents is the cause of an increase in the number of premarital pregnancies and a decrease in the acquisition of customary knowledge in recent decades. While there is no objective proof for the first claim, the second is undoubtedly a cause of the shrinkage of the Nalik lexicon.

It is not a surprise that the introduction of new technology, flora, fauna, and religious beliefs have brought new words into Nalik, before World War I in the German colonial era from German and Latin, and for the past 100 years since the Australian takeover of New Ireland during World War I from English and Tok Pisin. What is surprising is that Nalik words for an increasing number of items that were

common even before the onset of colonialism, such as *varama* ‘knife’ and *win* ‘cup’, have been completely replaced by their English and/or Tok Pisin equivalents (*naip* ‘knife’ and *kap* ‘cup’, respectively).

In addition to the replacement of these words, Nalik has lost whole fields of terminology. The most pervasive has been the field of indigenous spirituality. With the growing influence of Christian missions, foreign missionaries strove to eliminate “pagan” practices such as traditional dancing, birth control, and beliefs that they connected to sorcery. This became easier with the breakdown of the *haus boi* system and young people having fewer opportunities to have regular access to instruction by custodians of Nalik knowledge. As the awareness of Nalik cosmology and spirituality has been lost, a great many of the words used to describe them have also been lost. In recent years this has widened to include the loss of indigenous words that were adapted to use in the introduced religions that have to a great extent supplemented the indigenous belief system.

The rapidity of this loss can be seen in worship services and scripture study meetings among Nalik Bahā’īs, who published a prayer book in Nalik with Christian, Muslim, and Bahā’ī prayers and a translation into Nalik of an important book of Bahā’ī scripture, *Kalimāt-i-Maknūnih* (*Hidden Words* in English, *A Doring Bangbaang* in Nalik), soon after the adoption of a Nalik orthography in the 1990s. In these translations a conscious effort was made to incorporate terminology from traditional Nalik cosmology. While being actively used at that time, which coincided with a short period of Nalik revitalisation led by a respected clan leader and Bahā’ī teacher, Michael a Xomerang, today four decades later they are used only by some elderly people and, except for a few hymns, have disappeared from active public use. Even if they can use Nalik for everyday communication, few younger and middle-aged people in the 2020s have enough understanding of higher registers of Nalik to use these materials comfortably and prefer using scriptures in English instead.

A field in which Nalik terminology has disappeared to an even greater extent is mathematics. Nalik originally had a well-developed mathematics system based on a hybrid base-five and base-ten system that could be used to count up to 999, with a separate completely base-ten system to count up to ten for ceremonial purposes, such as counting pigs donated for a memorial service.<sup>5</sup> With universal education in English and its less complicated counting system and the almost identical counting system used in the New Ireland dialect of Tok Pisin, this system fell out of active use after World War II. An attempt was made in the 1990s to revive it for use in the Nalik-medium lower primary schools being set up at that time,

---

<sup>5</sup> See Volker (1996) for a detailed description of Nalik mathematic terminology.

and even to expand it by coining Nalik terms for words needed to teach modern mathematics. With the collapse of that experiment and the reintroduction of an all-English school system in this century, however, there has been no incentive to hold on to it. Today even people who speak Nalik frequently use the English / Tok Pisin counting system and its numerals in Nalik for all but the first few numbers. When they do use original Nalik numerals, some speakers place them in a pre-nominal position as in English and Tok Pisin, rather than after the noun as has been the case in Nalik in the past. The ceremonial counting system has almost completely vanished except for a few clan leaders who have learned it by reading Volker (1996).

The examples of Nalik lexical shrinkage mentioned so far have included some kind of replacement: the replacement of an indigenous word by a loanword (e.g., ‘knife’), the replacement of indigenous terms for spiritual concepts with new introduced terms, and the wholesale replacement of the Nalik mathematics system by the English / Tok Pisin system. A more serious instance is terminology related to flora and fauna, as this is a disappearance of a whole register rather than just the displacement of an indigenous word with an introduced system. Primary school children who participated in a project looking at the role of local birds in Nalik culture showed a surprisingly low awareness of the names of local flora and fauna (da Silva 2017). In some cases, even with the names of their own clan totems they knew only the Tok Pisin and not the indigenous name of the bird. More concerning than that was the fact that for many animals and fish they could only describe the creature, not give its name in any language. This indicates that they may not recognise it when an elder talks about it, even when the Nalik name is used in a discourse that is otherwise all in Tok Pisin. This limits their ability to have access to indigenous knowledge of the environment in which they live.

It should be pointed out that there is public awareness of Nalik lexical shrinkage. This is due in part to the attempt by Michael a Xomerang mentioned above to revive indigenous “classical” vocabulary and to return to indigenous words that had been replaced by loanwords. While since a Xomerang’s death in 1999, that attempt at linguistic revival is no longer very active, several words are still used as linguistic emblems by people who wish to indicate their desire to use “classical” Nalik. These include *roxon* instead of *gaat* in constructions such as (6) and using indigenous *taamon* ‘if’ instead of the Tok Pisin loan *sapos* (from English *suppose*) ‘if’, as well as the use of the respectful formal greeting *maluvio* ‘respect and love’ that was revived as a daily greeting in primary school assemblies in the 1990s.

## 7 Semantics

Together with actual lexical shrinkage, we can notice that the semantics of a number of Nalik words and phrases has moved in the direction of Tok Pisin. To some extent this is because of changes in society in the century and a half that Tok Pisin has been part of the New Ireland linguistic ecology, but often there is a change in the direction of the semantic range of a Nalik word or phrase in the Tok Pisin equivalent that cannot be explained solely by non-linguistic changes in society.

Perhaps the most fascinating example of this is the word *lagaf*. To understand how the semantic range of this word has changed with bilingualism in Tok Pisin, we first need to understand how the English words *hot* and *hard* changed phonologically as they were adopted into Tok Pisin. In many words, English [ɒ] (as in *hot*) and [ɑ:] (as in the Australian pronunciation of *hard*) were merged in the five-vowel Tok Pisin phonemic system as /a/. At the same time, with the exception of only an extremely small number of words, Tok Pisin does not permit voiced final consonants, so a final [d] in English became a final [t] in Tok Pisin. In this way, both English *hot* and English *hard* became homonyms in Tok Pisin, *hat*.

Like English, Nalik originally had two words meaning ‘hot’ and ‘hard’ or ‘difficult’, *lagaf* and *vulvulazaai*, respectively. But with bilingualism in Tok Pisin, an increasing number of people used *lagaf* for both words to the extent that by the 1990s, *vulvulazaai* was remembered by some older Nalik speakers, but no longer in common use. Today, just as there are two homonyms for *hat* in Tok Pisin, there are now the two same meanings for the two homonyms of Nalik *lagaf*.

As indigenous words are replaced by Tok Pisin and/or English loanwords, semantic distinctions originally made in Nalik can be lost. This is different from the creation of homonyms, such as with the modern use of *lagaf* just described. With *lagaf*, people use the same phonemes, but are still just as aware of the difference in meaning between ‘hot’ and ‘difficult’ as when there were two separate lexical items. When asked, even linguistically unsophisticated people analyse the two meanings as homonyms, and school children have no problems learning and differentiating the two English words *hot* and *hard*.

An example of semantic distinctions being completely lost is the introduction and wide use in modern Nalik of the Tok Pisin/English loanword *lav* ‘love’ with all its Western meanings. This word has for the most part replaced the indigenous Nalik words *olavaai* and *olaving*. *Olavaai* is an intransitive verb meaning ‘to love with no expectation of that love being reciprocated’, for example, when one falls in love with a happily married person who shows no interest in an extramarital affair. *Olaving* is a transitive verb (with the transitive suffix *-ing* already mentioned above) meaning ‘to love with a reasonable expectation of that love being reciprocated’, such as the love of a child for its mother or the romantic love of someone

who obviously shows interest in being seduced. With both *olavaai* and *olaving* being replaced by English *love* / Tok Pisin *lav*, distinctions that once could be easily made in song or oral literature with the two indigenous words can no longer be made succinctly, and the immediate awareness of the status of a potential relationship has been lost.

Sometimes changes in society can cause the semantic range of a Nalik word to change so that the surface form of the indigenous word remains, but it ends up with the same semantic range as a different word in Tok Pisin and / or English. *Ziparaas* is an example of this. In precolonial society this word referred to a protective fence around a village to keep out invaders. With this idea of protection, the word was also used to describe the protection that legal decisions made by clan leaders had. This latter meaning was transferred to the legal decisions and protections of the colonial legal system. With the imposed peace that the Australian colonial administration provided, the protective fences were no longer needed and were forgotten by everyone except a few people who were interested in oral history, so the word lost its meaning of a fence. As this was occurring, introduced religious and political ideologies introduced Western ideas of the ethics of justice, which were attached to the word describing the legal system that enforced these ethics. Today the meaning of *ziparaas* has moved to being almost identical to that of *justice* in English and *jastes* in Tok Pisin, the ethics and practice of being fair, just, and reasonable. Any meaning related to its original definition as a protective fence has been completely lost.

The semantic range of *mus* has changed in a similar circular way. When cars and trucks were first introduced to New Ireland, they were given the name *mus*, which originally meant 'large canoe', but by being applied to land vehicles, the meaning became widened to mean 'transport'. But in recent years, the Tok Pisin loanword *ka* (from English *car*) has become more widely used, pushing out the Nalik word originally used for this introduced technology, so that *mus* is returning to its original meaning of 'water transport'.

There are a number of expressions in Nalik that may be calques from Tok Pisin. Because of the influential role that New Ireland languages played in the development of Tok Pisin from Melanesian Pidgin in the early colonial era, it is difficult at this stage to know if certain idioms were originally from Nalik and other New Ireland languages and became calques in early forms of Tok Pisin or if the idiom originated in Tok Pisin and was later calqued to make a similar Nalik phrase. The use of *roxon*, and later *gaat*, 'have' in existential phrases in a construction similar to Tok Pisin *gat* has been discussed above. Because there was already a different construction for existential phrases using *bawai*, it is quite likely that the use of the equivalent of 'have' is a calque, particularly since the loanword *gaat* edged out using the indigenous word *roxon*.

Similarly, with Nalik expressions such as *dor texaasing* ‘to announce’ or ‘to explain’, a word-for-word translation of the Tok Pisin idiom with the same meaning, *tok save* (literally *speak knowledge*), or *mara na vaal* (Tok Pisin *ai bilong haus*) ‘front door’ (in both languages literally *eye of the house*), it is difficult to tell if the Nalik phrase is a relatively late calque of the Tok Pisin phrase, if the Tok Pisin phrase developed much earlier from Nalik and / or the other Austronesian languages that were the substrate of the language, or, less likely, the two were independent developments. Unfortunately, this question will be difficult to answer, as we do not have archival evidence from Nalik from the early part of the twentieth century to see how far back these Tok Pisin-like idioms go in Nalik.

The examples so far have been words and phrases that Nalik has borrowed or imitated from Tok Pisin. As the language of greater power, we would expect that Tok Pisin would have taken fewer words or phrases from the language of lesser power. Indeed, the only loan in the opposite direction from Nalik into the New Ireland dialect of Tok Pisin that I have documented is the idiom *fangan faat* ‘embezzle’ or ‘embezzlement’. Literally ‘eat stones’, it uses *faat* ‘stone(s)’, a Nalik slang word for coins. It is used unchanged by speakers of many languages in New Ireland to speak humorously or ironically in Tok Pisin about embezzlement. I have not seen it used in Tok Pisin outside of New Ireland Province.

## 8 Pragmatics

As society has changed, public rhetoric styles have changed greatly. This is undoubtedly due to the introduction of Western ideas of equality and democracy and the resulting drastic changes in power dynamics caused by the breakdown in the power of clan leaders that came as the German and then Australian colonial administrations and missions exerted their power and later introduced the Westminster political system, with its politicians who did not necessarily have traditional titles. Many Nalik people identify these changes with the introduction and spread of Tok Pisin and English, which are thought of as much more “direct” than Nalik, so the adoption of more direct ways of speech in Nalik is identified as being more like Tok Pisin and English.

It is understandable why these languages are thought of as more utilitarian. Tok Pisin had its origins in plantation labour gangs comprised of mainly young men. As it became a more general interethnic language, it was not used for ceremonial events such as *malagans* (traditional northern New Ireland memorial ceremonies) until quite recently. It is used in musical lyrics today, but still rarely used for other creative oral or written literature. Similarly, although English has a rich

literature, it is introduced in primary and lower secondary schools in Papua New Guinea as a utilitarian tool needed for the acquisition of knowledge in other subjects. Only a small number of students get to upper secondary school or university where more literary uses of English are introduced.

Naliks are headed by clan leaders, *maimai* (translated into English today as ‘chiefs’), who are assisted by *wangpaang* (literally ‘messengers’, but translated into English today as ‘assistant talking chiefs’). In pre-colonial Nalik society, the clan was the highest social unit. The pre-colonial clan hierarchy of leaders has been described in an as yet unpublished manuscript by Michael a Xomerang (n.d.). In precolonial society, a *maimai* spoke in public in very high registers and avoided speaking directly to the public when giving orders. Often these orders and pronouncements would be opaque to the general public. He would have at least one *wangpaang*, who would act as an interpreter, ensuring that the message the *maimai* wanted to deliver was understood and obeyed. For example, if there were a case of illicit sexual conduct, such as between two people of the same clan (an act thought of an incestuous in New Ireland cultures), he might make a public announcement that “pigs have broken the fence that needs to be repaired”. The offending couple would be expected to understand that they were guilty and should go away and commit suicide, but if they were slow to understand the indirect condemnation, the *wangpaang* would go to the couple and explain in direct words what the meaning of the *maimai*’s pronouncement was.

Today a *maimai* will publicly denounce illicit behaviour and issue punishments with very direct and strong words, often backed up by village court decisions (although not to the extent of demanding suicide!). Whether speaking in Nalik or, today increasingly often, Tok Pisin, the *maimai* is likely to use the same style of direct speech that would have been used by a Tok Pisin-speaking colonial administrator in the past or an English-speaking national court judge or teacher today.

Whereas as late as the 1990s, Nalik was the language heard most at a weekly village meeting or a more special ceremonial event, such as a funeral or *malagan* memorial celebration, Tok Pisin has now become the default language. Often the Tok Pisin oratory will be framed by Nalik expressions to open and close a talk, but the most important explanations will be in Tok Pisin. This is explained as a recognition of the inability of young persons to understand Nalik, the presence of residents in the community from outside the Nalik area and even from outside the province, and a breakdown in multilingualism that older generations tended to have in several New Ireland languages. In this century even an increasing number of men being appointed clan *wangpaang* and *maimai* do not speak Nalik and will go to the extent of making the customary greetings and ceremonial utterances used at *malagan* memorial ceremonies in Tok Pisin. Where they do use Nalik, they often

do not have a command of the higher ceremonial registers of Nalik that their predecessors had, such as the special ceremonial counting system described above.

## 9 Conclusions and the implications for Nalik society

Today, all persons who speak Nalik also have Tok Pisin, and to a lesser extent English, as part of their linguistic repertoire. As we have seen, Tok Pisin is the dominant language, which is a motivation for these bilingual speakers to use Nalik a way that is more like Tok Pisin in its phonology, morphology, syntax, lexicon, and pragmatics. This phenomenon has implications for Nalik society and the role the Nalik language plays in the modern identity of individuals and Nalik society.

The reality of rapid language change in the direction of Tok Pisin means that persons interested in the maintenance and revitalisation of Nalik must make decisions about whether to accept this modern, more Tok Pisin-like Nalik or to try to revive older forms. At present, when asked how Nalik should be presented and used in any printed or oral public form, many people respond that preserving and promoting “pure” Nalik should be the goal of any Nalik revitalisation project. These older forms can be elicited from Nalik elders as well as from the few recordings that exist, the few older Nalik writings that exist by persons such as Michael a Xomerang, and the unpublished notes from the inter-war years left by the Nalik-speaking German Catholic missionary Gerhard Hoffman.<sup>6</sup> This is currently the goal of a small number of elders in Madina Village who meet once a week to exchange vocabulary and discuss the meaning of words with the aim of producing a Nalik – English bilingual dictionary.<sup>7</sup> With the rapid change in Nalik in recent years, the importance of this documentation cannot be underestimated.

For educators there are pedagogical implications of the drift in Nalik towards having a more Tok Pisin-like structure. At the moment there is little training in comparative linguistics in teachers’ pre-service education. With a greater awareness of comparative linguistics and the dynamics of language change in bi- and multi-lingual societies, teachers can be more aware of changes in Nalik. This in turn can

---

<sup>6</sup> Copies of Hoffman’s Bible and liturgical translations into Nalik have been deposited in the Lugagon Village Catholic Church archives and in the Pacific Collection of the University of Hawai’i Library.

<sup>7</sup> An early rough draft of this dictionary with questions that remain unanswered is available online for public comment at the James Cook University Cairns Institute Language Archives (<https://www.jcu.edu.au/language-and-culture-research-centre/resources/language-archives>).

help instructors, both teachers in the formal educational sector and clan leaders in the informal traditional preparation of new clan leaders to be able to teach students more easily to keep the languages they speak separate and to develop rhetorical skills in all of them. Efforts such as this may or may not preserve older forms of Nalik that are thought to be more “classical” and linked to higher registers, but they will at least help people make informed decisions about the linguistic changes occurring in their society.

If the consensus of the community is that Nalik, at least in its present form, will eventually disappear and that efforts to maintain its “pure” presence are futile, Nalik-speaking society will have to make important decisions about what defines its own identity. This may lead to an acceptance of some emblematic use of Nalik being maintained by persons who identify as Nalik, but who in their everyday lives are Tok Pisin speakers and who, if they do have a knowledge of Nalik, use it in a way that is almost a relexification of Tok Pisin. Or it may lead to an indigenous identity that is separated from the use of the Nalik language entirely. It is hoped that discussions around these topics will lead to informed discussions around identity and language, and the degree to which fluency in Nalik and the ability to use certain registers of Nalik will be necessary for adults to participate meaningfully in Nalik culture in the twenty-first century.

## Abbreviations

3	third person
ART	article
EXIST	existential
OBL	oblique
PM	predicate marker
SG	singular
SM	subject marker
TR	transitive

## References

- Anderson, Fred. E. 2015. Education, power and sociolinguistic mobility. In Craig Alan Volker & Fred Anderson (eds.), *Education in languages of lesser power: Asia-Pacific perspectives*, 13–29. (IMPACT: Studies in Language and Society 35). Amsterdam/Philadelphia: John Benjamins.
- Blommaert, Jan. 2010. *The sociolinguistics of globalization*. (Cambridge Approaches to Language Contact). Cambridge: Cambridge University Press.

- Blust, Robert. 2013. *The Austronesian languages*. (Asia-Pacific Linguistics A-PL 008). Canberra: Australian National University Press.
- CIA. 2024. *World Factbook*. <https://www.cia.gov/the-world-factbook/> (checked 29 November 2024).
- Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2024. *Ethnologue: Languages of the world*. 27th edn. Dallas: SIL International. Online: <http://www.ethnologue.com> (checked 1 November 2024).
- Gumperz, John Joseph & Robert Wilson. 1971. Convergence and creolization: A case from the Indo-Aryan/Dravidian border. In Dell Hymes (ed.), *Pidginization and creolization of languages*, 151–168. Cambridge: Cambridge University Press.
- Igitoi, John. 2021. *New Ireland Province*. Port Moresby: National Statistics Office (poster).
- Keesing, Roger. 1988. *Melanesian Pidgin and the Oceanic substrate*. Stanford: Stanford University Press.
- Laycock, Don. 1985. Phonology: substratum elements in Tok Pisin phonology. In Stephen Adolphe Wurm & Peter Mühlhäusler (eds.), *Handbook of Tok Pisin (New Guinea Pidgin)*, 297–310. (Pacific Linguistics C70). Canberra: Australian National University Press.
- Mosel, Ulrike. 1980. *Tolai and Tok Pisin. The influence of the substratum on the development of New Guinea Pidgin*. (Pacific Linguistics B-73). Canberra: Australian National University Press.
- Phillipson, Robert Henry Lawrence. 1992. *Linguistic imperialism*. Oxford: Oxford University Press.
- Ross, Malcolm David. 1988. *Proto Oceanic and the Austronesian languages of western Melanesia*. (Pacific Linguistics C-98). Canberra: Australian National University Press.
- Roy, Gargi & Kārumūri V. Subbārāo. 2023. Language contact and convergence in Northeastern India. In Jelle J.P. Wouters & Tanka Subba (eds.), *The Routledge Companion to Northeast India*, 209–306. Milton Park, Oxon/New York: Routledge.
- da Silva, Cláudio. 2017. *Aves na cultura Nalik: (re)Conhecendo a voz da comunidade por meio da investigação-ação (Papua Nova Guiné)*. (Masters in Social Education, Development and Dynamics Locals), Universidade de Coimbra, Coimbra. Retrieved from <http://hdl.handle.net/10316/84159>.
- Volker, Craig [クレイグ・フォルカー]. 1996. “ナリク語の辞書作成の経過報告 NO.1: ナリク語の数学 / A Lingmaam 1: A Ikakating wan a ling a Nalik / Nalik Dictionary Project 1: Nalik Mathematics,” 聖徳学園岐教育大学紀要 / *Annals of the Gifu University for Education and Languages* 32. 1–33.
- Volker, Craig. 1998. *The Nalik language of New Ireland, Papua New Guinea*. (Berkeley Models of Grammars Series 4). New York: Peter Lang Publishing.
- Volker, Craig Alan. 2015. Vernacular education in Papua New Guinea: Reform or deform? In Craig Alan Volker & Fred Anderson (eds.), *Education in languages of lesser power: Asia-Pacific perspectives*, 201–217. (IMPACT: Studies in Language and Society 35). Amsterdam/Philadelphia: John Benjamins, 2015.
- Volker, Craig Alan (forthcoming). Melanesian Pidgin, Tok Pisin and English in Papua New Guinea. In Kate Burridge & Raymond Hickey (eds.), *The new Cambridge history of the English language*, volume 6, chapter 38. Cambridge: Cambridge University Press.
- Volker, Craig Alan & Fred Anderson (eds.). 2015. *Education in languages of lesser power: Asia-Pacific perspectives*, 201–217. (IMPACT: Studies in Language and Society 35). Amsterdam/Philadelphia: John Benjamins, 2015.
- a Xomerang, Michael [Michael Homerang]. n.d. *A xuxzaai a vaaribuking sin a xastunaan a xulmu*. [Unpublished manuscript].

Tom Güldemann

# A Kwadi perspective on Khoe juncture-verb constructions

**Abstract:** Kwadi, a now extinct click language of southwestern Angola with poor documentation, was once a candidate for a linguistic isolate in Africa. The historical analysis of its pronoun system and available lexicon eventually led to establishing its genealogical relation to the southern African Khoe family. In this paper, I entertain further evidence for the higher-order lineage Khoe-Kwadi by looking at verbal morphology. I propose that the so-called “juncture” morpheme of Khoe is related to a grammatical element in Kwadi that attaches to verb roots and turns them into dependent forms. This hypothesis also informs the historical analysis of the Khoe juncture itself as well as the possible role of language contact in the development of multi-verb constructions in Khoe-Kwadi as a whole.

**Keywords:** dependent verb, Khoe-Kwadi, multi-verb construction, verb juncture, reconstruction

## 1 Introduction

Kwadi is an extinct click language once spoken by a group of small-scale pastoralists in southwestern Angola along the lower Curoca River (Guerreiro 1971). It already had very few speakers in the 1960s when E. Westphal recognized it to be an important linguistic research topic. In 2014, A.-M. Fehn encountered only two individuals who remembered a few words and expressions of the language (Fehn and Rocha 2023). Kwadi’s extinction over the course of the second half of the 20th century is mainly due to the shift of the small speech community to the southwestern Bantu language Kuvale.

Westphal (1964/5, n.d.a–c) himself collected the most extensive, albeit restricted data on Kwadi, including audio recordings. Regarding its genealogical relations, he first considered it to be an isolate (Westphal 1962: 8, 1963: 247) but later pondered

---

**Note:** The present topic was presented first at the 5th International Symposium on Khoisan Languages and Linguistics, Riezern, 14–16 July 2014 (see Güldemann and Fehn 2014). My co-author Anne-Maria Fehn did not pursue a publication of this joint research but I am grateful for the initial collaboration without which this article would lack several insights into Kalahari Khoe.

---

**Tom Güldemann**, Humboldt-Universität zu Berlin, e-mail: tom.gueldemann@hu-berlin.de

the possibility of its link to the Khoe family (1965: 137, 1971: 380). This idea was taken up by Köhler (1981: 469) and Ehret (1982). None of these proposals was based on the inspection and analysis of sufficient data within historical-comparative methodology. Voßen’s (1997) sound reconstruction of Proto-Khoe, including a large amount of morphology, ultimately paved the way for a dedicated assessment of the Khoe-Kwadi hypothesis. That is, my analysis of Westphal’s Kwadi material (Güldemann 2001) and its comparison with Voßen’s (1997) Khoe reconstructions in Güldemann (2004) and Güldemann and Elderkin (2010) revealed a cognate pronoun system and more than 60 potential lexical isoglosses – a sizeable number of them involving emerging regular sound correspondences.<sup>1</sup> This robust evidence for a genealogical relation between the two units has by now been consolidated and extended by Fehn and Rocha’s (2023) extensive historical-comparative reconstruction of Proto-Khoe-Kwadi lexicon. Using old as well as previously unavailable data, their research proposes more than 130 lexical items for the proto-language based on a detailed phonological analysis and a fuller establishment of regular sound correspondences.

Possible future revisions of Voßen’s (1997) internal Khoe classification aside, the structure of Khoe-Kwadi with Kwadi and Khoe as sister branches is as in Figure 1.

Khoe-Kwadi			
	Kwadi		single language†
	Khoe		
		Kalahari Khoe	
		East	Shua LC:           Cara, Deti, IXaise, Danisi, etc. Tshwa LC:         Kua, Cua, Tsua, etc.
		West	Khwe LC:         !IXom, !IXo, Buga, !IAni, etc. Gllana LC:        Gllana, Glui, etc. Naro LC:         Naro, Ts’ao, etc.
		Khoekhoe	(Cape)† LC (!Ora-Xiri) LC (Eini)† LC Nama-Damara LC Haillom ‡Aakhoe

Notes: † = extinct, LC = language complex, (. . .) = only older and/or limited data

**Figure 1:** The current sub-classification of Khoe-Kwadi (Güldemann 2014: 27).

<sup>1</sup> Blench (2017: 180–181) still features Kwadi as an African isolate language whose similarities to Khoe are supposedly due to contact without, however, engaging with the evidence proposed by then in favor of the genealogical relationship.

Apart from the overall typological similarity between Khoe and Kwadi (cf. Güldemann 2013b) and the historical relation between their pronominal systems, little evidence exists so far in terms of cognate grammar. In this regard, Güldemann (2010) raised the hypothesis about an additional isogloss in proposing that the so-called “juncture” morpheme of Khoe languages has a possible cognate in Kwadi. This paper is to substantiate this further piece of morphological evidence for the Khoe-Kwadi hypothesis. After briefly introducing the juncture concept (Section 2.1), I give a cross-family survey of the Khoe juncture regarding its function (Section 2.2), morpho-syntax (Section 2.3), and phonology (Section 2.4), and discuss hypotheses about its origin (Section 2.5). In Section 3, I present data on two complex predicate structures of Kwadi that are pertinent to the historical link I advance here, namely a construction expressing volition (Section 3.1) and a pattern involving verb root reduplication (Section 3.2), and then summarize the results (Section 3.3). I finally present my historical conclusions (Section 4).

## 2 A survey of juncture-verb constructions in Khoe

### 2.1 The verb juncture in Kalahari Khoe – an introduction

Köhler (1981) developed the concept and coined the original French term “joncture” in order to describe suffixes in West Caprivi Khwe without an apparent meaning of their own that join verb roots to other grammatical suffixes. Due to distinct sets of allomorphs, he identified two distinct morphemes, namely “joncture I” before non-past suffixes and “joncture II” before suffixes for past tense and verb derivation, as illustrated in (1) and (2).

(1) West Caprivi Khwe (Köhler 1981: 498–499)

- a. *kyámà-à-tè*  
follow.spoor-JUN.I-PRS  
‘... follow a spoor’
- b. *kyámà-nà-há*  
follow.spoor-JUN.II-PST  
‘... followed a spoor’

## (2) West Caprivi Khwe (Köhler 1981: 497–499)

- a. *kwê-é-tè*  
not.want-JUN.I-PRS  
'... reject'
- b. *kwê-ré-hǎ*  
not.want-JUN.II-PST  
'... rejected'

Later historical-comparative work, notably Voßen (1997, 2010), established the relevance of the juncture for the entire Kalahari branch of Khoe and proposed a base form *a* with a range of different allomorphs. Since then, the juncture has been described in detail in a number of individual languages and has also been the subject of historical analyses. In the following, I present a brief cross-family survey of so-called “juncture-verb constructions”, leading to a partly new approach to their synchronic description and diachronic assessment.

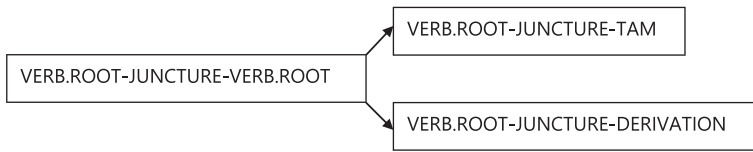
## 2.2 Functional characteristics

A first important step is the functional evaluation of the juncture. Depending on the nature of the following element, there exist three types of juncture-verb construction throughout Kalahari Khoe. That is, an initial verb root is linked by the juncture to: a) another verb root in a multi-verb construction (henceforth just MVC), b) a derivational suffix, and c) a TAM suffix. This is exemplified in Ts'ixa, respectively, with the collocation of ‘fly’ and ‘pass’ in (3a), the benefactive suffix *-ma* in (3b), and the past suffixes *-ta* and *-ha* in (3b/c) (note in (3c) that the juncture does not always have a segmentally overt form for which see Section 4).

## (3) Ts'ixa (Fehn 2014: 206, 223, 84)

- a. *g'íni=sì tè xalási=m l|xè //abuù-à ngèè*  
fly=F.SG NEAR.PST glass=M.SG LOC fly-JUN pass  
'The fly flew past the glass.'
- b. *kuí=m mǎ=m kà tí kà kúì.k'èè=sà tí*  
dress=M.SG DEM=M.SG MPO 1SG POSS sister=F.SG 1SG  
*kyúù-à-mà-nà-tà*  
buy-JUN-BEN-JUN-PST1  
'I bought that dress for my sister.'
- c. *xóo-hà ìi=sà*  
dry:JUN-PST3 tree=F.SG  
'the dry tree'

According to general knowledge about language change and concrete evidence from Khoe itself, these three types would appear to be related to each other historically, insofar as the final TAM and derivational suffixes requiring the juncture are plausibly derived via grammaticalization from final verbs in earlier MVCs. This applies in fact to two elements in the above Ts'ixa examples, namely benefactive *-ma* originating in Proto-Khoe \**ma* 'give' (cf. Voßen 1997: 440–441) and past3 *-ha* deriving from the lexeme *háǎ* 'exist' (cf. Fehn 2014: 45). This assumed historical network is schematized in Figure 2 (see Section 4 for further discussion).



**Figure 2:** Semantic map of the three juncture-verb constructions in Khoe.

However, not all verb suffixes occur with the juncture. For example, it is not used with the impersonal/passive suffix \**-e~\*i* (Voßen 1997: 191). This motivated Kilian-Hatz (2008) to consider the juncture in Khwe to be an active marker, an untenable interpretation for the rest of Kalahari Khoe in line with Vossen (2010: 48). Likewise, the juncture does not co-occur with verbal object indexation in the languages that have it (Voßen 1997: 191). In most languages, some predicate markers are not suffixes but pre- or postverbal particles, which also do not trigger the verb to take the juncture.

In the following, I outline the three types of juncture-verb constructions in more detail. I start with the MVC type in which the juncture joins two (or sometimes more than two) verbs into a complex predicate, called variably verb compounding (cf. Nakagawa 2006, Visser 2010, Rapold 2014) or verb serialization (cf. Kilian-Hatz 2006, 2010; Haacke 2014). Across the group, juncture-based MVCs display a wide variety of functions, as illustrated in (4a–d) from Ts'ixa. The routinized use of one of the combined verb roots leads to asymmetrical combinations, as appears to be the case in the patterns of (4d). These are the potential locus of grammaticalization, which largely affects the final verb and turns it into a grammatical suffix to be discussed further below.

## (4) Ts'ixa (Güldemann and Fehn 2017: 510–511)

## a. Sequential cause-effect

*nóxá=m ín=mà tí kò muùn-à 'ààn*  
 snake=M.SG DEM.REF=M.SG 1SG IPFV see-JUN know  
 'I recognize this snake.'

## b. Accompanying manner

*tí kò pere gllài*  
 1SG IPFV flee:JUN run  
 'I run like a fugitive.'

## c. Accompanying posture ('sit', 'stand', 'lie' etc. + verb root 2)

*tí kò nyúun-a ll'àm katsí-sà 'à*  
 1SG IPFV sit-JUN beat cat-F.SG OBJ  
 'I beat the cat sitting.'

## d. Action-path (verb root 1 + 'exit', 'enter', 'pass' etc.)

*nguú-m 'à tí kò gllai-a ky'òà*  
 house-M.SG LOC 1SG IPFV run-JUN exit  
 'I run out of the house.'

MVCs as in (4) are hardly dealt with by Voßen (1997) but have received more detailed attention in descriptions of individual Khoe languages such as Nakagawa (2006) for G!ui, Kilian-Hatz (2006, 2010) for West Caprivi Khwe, Visser (2010) for Naro, Fehn (2014) for Ts'ixa, and Fehn and Phiri (2022) for northeastern Kalahari Khoe as a whole. Such constructions also exist in southeastern Kalahari Khoe, for example, Kua, as shown in (5).

## (5) Kua (Chebanne and Collins 2017: 99)

*à-bè kúnī ʔòà nllàì-ā #ãã*  
 3-M.SG cart LOC jump-JUN enter  
 'He just jumped into the cart.'

A second context of the juncture is before derivational suffixes. A relevant cross-Khoe survey has been provided by Voßen (1997, 2010). His work discusses close to twenty different derivations but only six of them regularly appear with the juncture in one or more of the languages surveyed. Table 1 lists them and presents language-specific examples. The only forms that require the juncture in all Kalahari Khoe languages, although not in all dialects, are the two deriving transparently from the final verb of an earlier asymmetrical MVC, namely dative \*-ma and terminative-itive \*-xu. Other suffixes display a higher degree of dialectal variation and only sporadically appear with the juncture, some possibly only in analogy to suffixes with a true verbal origin.

**Table 1:** Derivation suffixes with the juncture in Kalahari Khoe (Vossen 2010: 53–54).

Function	Reconstruction (source)	Example
Reflexive	PK *-sani	<i>kúǀm-á-hi</i> ‘hear oneself’ (Kua)
Directive-locative	PKalK *-!’o	<i>péé-á-ò</i> ‘jump ahead’ (Cara)
Dative or benefactive	PK *-ma (<PK *ma ‘give)	<i>góm-á-mà</i> ‘smoke for’ (Tsua)
Terminative-itive	PKalK *-xu (<PK *xu ‘leave’)	<i>giám-a-xú</i> ‘get rid of’ (Buga)
Causative III	PK *-si	<i>hĩĩ-hĩĩ-á-sí</i> ‘use, seduce’ (IIAni)
Repetitive or iterative	(? < *X+Causative III)	<i>kúú-a-kási</i> ‘keep going’ (Danisi)

The third type of juncture-verb construction involves suffixes in the TAM domain. The most elaborate paradigm of such suffixes exists in West Caprivi Khwe (cf. Köhler 1989: 122–123, Kilian-Hatz 2008: 98–105). While five past suffixes appear with juncture II, which is the form also found in other Kalahari Khoe languages, four non-past suffixes take juncture I, which is assumed to constitute a Khwe innovation. Köhler confidently derives all but one suffix from an earlier final verb of a MVC. The entire suffix system is given in Table 2.

**Table 2:** TAM suffixes with the juncture in Khwe (after Köhler 1989: 123).

Function	Form	Lexical source	Juncture
Present, imperfective	-tè	‘stand’	I
Progressive	-n#ùè	‘sit’	
Habitual	-llòè	‘lie, sleep’	
Future	-gòè	? ‘approach, go to meet’	
Past hodiernal	-tà	‘rise, stand up’	II
Past hesternal	-ll’òm	‘sit/sleep on tree (of bird)’	
Past proximal	-tĩ	‘be there, stay’	
Past remote	-hĩ	‘do make’	
Perfect	-hā	‘exist, be there’	

Cognates of inflectional suffixes requiring the juncture also exist in other Kalahari Khoe languages, notably in the past domain. One suffix is attested throughout the group, namely \*-hā/ha marking perfect, current relevance, stative, and past (Voßen 1997: 365, cf. (3c) above from Ts’ixa). Voßen (1997: 231) also records the past form \*-hĩ in Tshwa varieties – a likely cognate of Khwe -hĩ. In Ts’ixa, Fehn (2014: 147–149) identifies three past suffixes requiring the juncture, -hā/ha for remote past, -ta for hodiernal past, and -’o for recent past, whereby the first two are assumed to be related to corresponding Khwe forms.

## 2.3 Morpho-syntactic scope

The very term “juncture” (or “linker” as per Vossen 2010) conveys the notion of a plain conjunction between an initial verb root and a following lexical root or grammatical suffix – a **symmetrical** structure that can be schematized as in (6). Likewise, the MVC subtype in particular appears to be viewed as a syntactically balanced structure by those authors who characterize it as plain verb serialization (Kilian-Hatz 2006, 2010; Haacke 2014).

(6) [[VERB.ROOT]-JUNCTURE-[VERB.ROOT or SUFFIX]]

There have been few alternative analyses, notably in Heine’s (1986) historical account to be discussed in Section 2.5. However, there is ample synchronic evidence for an **asymmetrical** character of juncture-verb constructions. Such is provided especially by Nakagawa’s data on Glui. With reference to the phonological effects of the juncture, Nakagawa (2006) aptly observes that it has the function of marking the altered verb root 1 of a “compound verb” (called here neutrally MVC), which can be transferred to other Khoe languages. That is, the segmental and prosodic interaction of the juncture concerns the verb root preceding it, while the following root or grammatical element remains unaffected. For the specific case of MVCs, this has been schematized by Nakagawa as in Figure 3 (VR = verb root).

INPUT			OUTPUT		
VR1	+	VR2	>	VR1'	VR2
↑		↑			
Alternation		No alternation			

**Figure 3:** Juncture-based MVC formation in Glui (after Nakagawa 2006: 66).

Other data from Glui point in the same direction. The variable reflexes of the juncture together with its preceding verb root not only represent a morpho-phonological unit but also form a morpho-syntactic sub-constituent. This is reflected by its possible separation from verb root 2, so that the two verbs that are normally “joined” by the juncture are no longer adjacent.

- (7) Glui (Nakagawa 2006: 76)
- a. *η!abo-sera ca ts'āũ-a mǎǎ*  
sandal-F.DU 1SG make-JUN give
  - b. *η!abo-sera ts'āũ-a ca mǎǎ*  
sandal-F.DU make-JUN 1SG give
  - c. *ts'āũ-a η!abo-sera ca mǎǎ*  
make-JUN sandal-F.DU 1SG give  
'Make a pair of sandals for me.'

The possible disjunction of the juncture-marked VR1 and VR2 is shown in (7). The expected MVC pattern with *ts'āũ-a mǎǎ* 'make for' establishing the benefactive derivation (cf. Table 1, Section 2.2) can be seen in (7a). In (7b) and (7c), however, the two verbs are interrupted by one or both arguments of the complex predicate, whereby the juncture is retained on VR1 (cf. also Fehn and Phiri (2022: 153) for such verb separation in Ts'ixa).

I thus conclude that juncture-verb constructions have a morpho-syntactic configuration that is more structured than commonly assumed in that initial verb root and juncture form a constituent opposed to the final verb root or suffix. The last element can be viewed as the head of the complex constituent in line with the overall head-final syntax of Khoe languages. I thus propose an analysis in terms of morpho-syntactic asymmetry in (8).

- (8) [[VERB.ROOT-JUNCTURE]-VERB.ROOT or SUFFIX]

Treating MVCs with the juncture as verb serialization should now be reassessed against the canonical typological approach. Thus, Aikhenvald (2006: 1) and many others require that the verbs in a serial construction “act together as a single predicate, **without any overt marker of coordination, subordination, or syntactic dependency of any other sort**” [bold TG]. Juncture-verb constructions do not meet such a definition. Instead, they are better analyzed as non-symmetrical MVCs whose crucial element is a marker of dependency with scope over the initial verb root, turning it into a non-finite construct form.

## 2.4 Phonological variation

Since Köhler's (1981, 1989) earliest description of West Caprivi Khwe it is known that the juncture involves extensive segmental allomorphy as well as tonal perturbations on the initial verb. Voßen's (1997, 2010) surveys have focused on the segmental variation, which concerns the assimilatory interaction of the juncture

with the final mora of the preceding verb root, involving such different shapes as CV, V, and N, and, if relevant, variable vowel qualities. Voßen only deals with a single juncture, because the distinction between two such elements in the Khwe group is unique and assumed to be a later innovation. Rather than reconstructing a proper proto-form, the author simply posits a base form *-a*, which can be preceded by such consonants as *r* or *n*; these are explained in line with Heine (1986) by the earlier presence of such consonants in the second C slot of CVCV verb roots.

Language-specific descriptions focused on the co-occurring segmental and prosodic juncture variation. A crucial contribution is again Nakagawa (2006) on G!ui, introducing among other things the term “flip-flop” for the tonal changes, which is inspired by Haacke’s (1999) description of this phenomenon in Namibian Khoekhoe (see below). Nakagawa identifies five juncture allomorphs in G!ui, which are represented in Table 3. Similar patterns are attested elsewhere, for example, in Khwe, Ts’ixa, Shua, and Tjwao (Kilian-Hatz 2008: 108–121, Fehn 2014: §4.2, Elderkin 2016, Fehn and Phiri 2022), whereby tonal flip-flop turns out to be universal and is thus a general effect of the juncture morpheme.

**Table 3:** Juncture allomorphy in G!ui (Nakagawa 2006: 66–67).

	Allomorph	Verb base		Verb+Juncture before <i>mǎǎ</i> ‘give’	
1	Flip-flop only	<i>kláá</i>	‘skin’	<i>kláǎ mǎǎ</i>	‘skin for someone’
2	Flip-flop and <i>a</i> -suffixation	<i>ʃbú</i>	‘wrap’	<i>ʃbǔ-a mǎǎ</i>	‘wrap for someone’
3	<i>a</i> -suffixation only	<i>qʰáǎ</i>	‘talk’	<i>qʰáǎ-a mǎǎ</i>	‘talk for someone’
4	<i>ná</i> -suffixation only	<i>kʰǒbē</i>	‘rub’	<i>kʰǒbē-ná mǎǎ</i>	‘rub for someone’
5	<i>r</i> -insertion only	<i>kʰáē</i>	‘stab’	<i>kʰá-r-ē mǎǎ</i>	‘stab for someone’

A segmental juncture is not attested in Khoekhoe varieties, so that they were excluded from Voßen’s (1997, 2010) comparative discussion. However, the tone change called flip-flop associated with the Kalahari Khoe juncture also occurs in the formation of complex predicates in Khoekhoe. As mentioned, the very term “flip-flop” was in fact coined by Haacke (1999) in his detailed tonal analysis of this language. Based on the research findings of this author, Rapold (2014) advanced a hypothesis that is crucial for the topic at issue here. He surveyed the types of complex predicates in Namibian Khoekhoe where weak flip-flop occurs and concluded that there is considerable overlap between them and semantically and formally related contexts of the Kalahari Khoe juncture (cf. Table 1 and 2 above).

Table 4 shows that all but one derivational element and the perfect/past marker from *hǎǎ* ‘exist’ of Kalahari Khoe are also attested with weak flip-flop on relevant Khoekhoe verbs. Rapold thus concludes that the prosodic phenomenon

**Table 4:** Kalahari Khoe juncture vs. Khoekhoe weak flip-flop (after Rapold 2014).

Grammatical context	Kalahari Khoe (juncture)	Khoekhoe (weak flip-flop)	Khoekhoe marker
Reflexive	(✓)	(✓)	-sèn
Directive-locative	(✓)	✓	? /'óá = 'meet'
Dative-benefactive	✓	✓	-pǎ
Terminative-itive	✓	✓	-xúú < 'leave'
Causative III	(✓)	(✓)	-sǐ
Perfect-past	✓	✓	hǎǎ < 'exist'

Note: (✓) attestation restricted

in Khoekhoe is a late developmental stage of the verb juncture in Kalahari Khoe. That is, the segmental loss of the juncture is found in Kalahari Khoe only with some verbs, while its reduction to a mere supra-segmental morpheme has been completed in Khoekhoe. This implies that the juncture should be reconstructed back to Proto-Khoe rather than just one subbranch.

## 2.5 Historical origin

Against the background of Voßen's (1997) assumption of a basic form *a*, the juncture has been subject to various attempts to assess its origin or at least its relation to other grammatical elements. Before discussing two more concrete historical hypotheses, I briefly mention various other suffixes in Khoe languages that may appear akin to Voßen's juncture base form *a*. Some discussions of the juncture mention them because they partly have a linking function, or at least a morphosyntactic behaviour somewhat similar to the juncture.

- (9) Naro (Visser 2010: 181)

*thuu=r*      ***bóǎ-a***  
 PST=1SG.SBJ see-PFV  
 'I have seen'

- (10) Naro (Visser 2010: 181)

***#úú-a***      *te*  
 head-LINKER 1SG  
 'my head'

- (11) G!ui (Nakagawa 2013: 400)  
 ʔà-bì !áì jà #áó  
 3-M.SG good LINKER heart  
 ‘He is good in the heart.’ (= He is happy)
- (12) Namibian Khoekhoe (Haacke 2014: 139)  
 #gà̀n-à-dò̀m < #gà̀n-xà-dò̀m  
 be.close-ADJR-throat  
 to become hoarse

First, Naro possesses a perfective verb suffix *-a*, shown in (9). Second, there is an *-a* attaching to the bare noun in one type of possessive construction in Khwe (called “genitive” by Kilian-Hatz 2008) and Naro (called “linker” by Visser 2010), as illustrated for Naro in (10). Third, G!ui has a morpheme *a*, possibly a suffix, which links predicative adjectives to body part nouns conveying experiencer constructions of the pattern “adjective-noun-*ed*”, as in (11). Finally, a suffix *-a* in Namibian Khoekhoe, as in (12), links a verb and a noun for deriving a complex verb stem; it is assumed to be a shortened version of the adjectivizer *-xa*.

It should be clear from the brief characterizations and examples that none of these elements is a good match of the verb juncture at issue, because they attach to different hosts in distinct grammatical contexts and do not display the allomorphy or cause the prosodic changes typical for the juncture. The situation is arguably different for two other elements, which Heine (1986) and Elderkin (1986) have entertained to be related to the juncture; both authors deal, like Voßen (1997), with a single morpheme *-a*.

Heine (1986) traces the juncture back to an identificational marker *’a*, called by him “copula”, which is attested in some Khoe languages. This *’a* marked different types of nominal constituents, as schematized in his complex grammaticalization network in Figure 4.

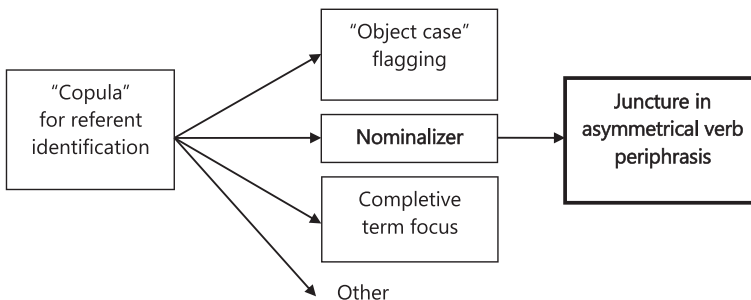


Figure 4: Semantic map of the identification particle *’a* (after Heine 1986).

(13) !Ora (Meinhof 1930: 53, 61)

*ham /xa-b-a tje ta !ũ 'a-b ka?*  
 which side-M.SG-? 1PL FUT go “COP”-M.SG want

‘Nach welcher Seite wollen wir gehen?’ [Which side do we want to go to?]

Heine’s hypothesis regarding the juncture as derived from some nominalizing element relies heavily on a single sentence from Meinhof’s (1930) description of the extinct Khoekhoe language !Ora, reproduced in (13). The crucial part is the final string *!ũ 'a-b ka* with the relevant *'a* followed by *-b*, which with all probability is the person-gender-number marker (henceforth just PGN) for (third-person) masculine singular. Heine (1986: 13–14) analyzes the one-off pattern [[VERB *'a*-PGN] AUXILIARY] as a verbal periphrasis in which the “copula” *'a* is said to nominalize the preceding content verb. Meinhof (1930) himself explicitly admits not to understand this string, and it is indeed opaque both in general and within the grammar of !Ora, so that Heine’s syntactic analysis must be regarded as ad-hoc.

(14) [[VERB-*'a*-PGN]-AUXILIARY] > [VERB-*'a*-Ø-GRAM]

Heine further proposes that the grammaticalization ended in !Ora at the above stage but progressed further in Kalahari Khoe. As schematized in (14), the PGN would have been lost and the “copula” *'a* developed into the juncture, the following auxiliaries becoming TAM markers. While Heine (1986) did not treat juncture-based derivational suffixes and MVCs, Kilian-Hatz (2004) and Vossen (2010) extended this hypothesis at least to the first domain.

Elderkin (1986) entertains another hypothesis. He derives the juncture from a conjunction *'à* that came to form a phonological word with the preceding verb root. He assumed that a conjunction of this shape no longer exists in modern Khoe (see also Vossen 2010: 47) but later research revealed the existence of such a predicate coordinator in such languages as Naro, Ts'ixa, Shua, and Tjwao (cf. Fehn and Phiri 2022). Haacke (2014) even equated this *'à* with the juncture in Naro on the basis of examples like (15).

(15) Naro (Haacke 2014: 131)

- a. *ga-sa ko //õa 'a kx'aa*  
 [PRO-F.SG IPFV descend] ?JUN drink
- b. *//õa=s ko 'a kx'aa*  
 [descend=3F.SG.SBJ IPFV] ?JUN drink
- Both: ‘She is coming down to drink.’

However, the examples in (15) reveal considerable differences of the coordinating 'a vis-à-vis the juncture in that the former combines a complex clausal entity with a following verb (phrase) and need not be adjacent to the first verb, as is the case in (15b). Evidence to the same effect is presented by examples of the conjunction 'a (written orthographically as *a*) in Visser (2010), who explicitly distinguishes this element from the juncture suffix.

(16) Naro (Visser 2010: 180)

*tàà-è=r ko a lōò tama*  
 [defeat-PASS=1SG.SBJ CONT] CONJ [go not]  
 'I am defeated and don't go.' > 'I can't go.'

Thus, in (16) negation is not shared between the coordinated predicates, the first verb is marked for passive voice, which precludes the real juncture, and 'a is detached from the first verb. As was shown above in examples with the juncture, however, sharing of clause operators and voice as well as adjacency between the initial verb and the gram itself are defining characteristics of such constructions in Kalahari Khoe. The distinct nature of the two elements is confirmed by a comparison of their constructional meaning, which can be seen in the minimal pair in (17): the juncture in (17a) expresses a single event, while the conjunction in (17b) renders a sequence of two separate events.

(17) Kalahari Khoe (Visser 2010: 179–180)

- a. *!xóó-(a) gùì*  
 hold-JUN lift  
 'pick/lift up'
- b. *!xóó a gùì*  
 hold CONJ lift  
 'hold/touch and (then) lift'

Overall, the two historical scenarios proposed by Heine (1986) and Elderkin (1986) are problematic for several reasons. First, the identification particle as well as the conjunction of the form 'a still exist in modern Khoe languages and are clearly distinct from the juncture. Moreover, the authors do not account for the absence vs. presence of the glottal stop between \*-a and 'a, which is not trivial in view of the fact that [ʔ] in Khoe is phonemic. Likewise, the prosodic effects associated with the juncture (see Section 2.4) are not found on the elements preceding the conjunction or the identification particle. Finally, while morpho-syntactic changes certainly occur in grammaticalization, the authors

fail to present detailed scenarios as to how the purported source element became a suffix between a particular type of host and a particular set of following elements.

Last but not least, all previous hypotheses start out from a reconstructed juncture of the form \*-a. The following section brings in data from Khoe's extinct relative Kwadi, which provides a new perspective on this and other issues revolving around the history of the Khoe juncture. That is, in the quest for further grammatical comparisons for assessing Kwadi's genealogical status, I entertain another potential morphological cognate by observing that the Kwadi suffix *-la* (and its allomorphs) seems to be a marker of non-finiteness – this element and other verb morphemes potentially inform the historical analysis of the juncture in Khoe" (Güldemann 2010: 14–15) – an idea I flesh out in Section 3.

### 3 Complex verb constructions in Kwadi

#### 3.1 The volition construction with *-(a)la*

The first relevant Kwadi structure is straightforward, despite the scarcity of data. The less than a dozen tokens present an overall consistent picture according to which a verb root is joined with a final auxiliary 'want' by means of a suffix *-(a)la*, as exemplified in (18).<sup>2</sup>

(18) Kwadi (Westphal 1964/5; Güldemann 2001: 53, 56)

- a. *ta k''ɔle jwala-xε* cf. *ʔũ(ũ)* 'eat' (Fehn and Rocha 2023, Suppl.: 64)  
 1SG meat eat:?-want  
 'I want to eat meat.'
- b. *ta ʔl'áme 'õala-xε* cf. *ʔũã* 'buy' (Fehn and Rocha 2023, Suppl.: 62)  
 1SG knife buy:?-want  
 'I want to buy a knife.'

<sup>2</sup> The Kwadi data below are from Westphahl's (1964/5) fieldnotes. These consist of loose sheets of paper, which I copied and then numbered in the order they were in when first consulting them. Others may have used the fieldnotes afterwards and reordered the loose sheets, so that the page numbers in the following examples from my unpublished morphosyntactic analysis (Güldemann 2001) are unlikely to be recoverable from the original pages in the archive. While the linguistic annotation of the examples is mine, I did not change the recurrently inconsistent transcription.

This pattern represents a periphrastic construction conveying volition in a morpho-syntactic configuration that is compatible with Kwadi's overall head-final syntax. In the structure [[VERB-(a)la]-xe], the final auxiliary 'want' governs a verbal complement suffixed by *-(a)la*, which itself appears to encode the dependent non-finite status of its host. Note that the initial *a* of *ala* only appears rarely and could merely be the effect of regressive assimilation from the suffix vowel onto the verb root, instead of being a genuine part of the suffix.

### 3.2 The intransitive reduplication stem with *-la*

A recurrent but far more intricate phenomenon in Kwadi, referred to already by Westphal (1963: 247) and Güldemann (2013a: 262), are complex verb stems. In terms of token frequency, they happen to establish the most prominent verb pattern in Westphal's data. This does not necessarily reflect its central role in the language but could just as well be an artifact of the content and nature of his data elicitation. Whatever its status, it reflects a morphological pattern I argue to inform the historical comparison with the Khoe family.

The entire phenomenon is formally diverse and thus difficult to describe precisely – also owing to the lack of sufficient and/or coherent data, which to a large extent arises from the preliminary character of Westphal's transcriptions. This problem needs to be taken into account with respect to the following discussion. The basic pattern can be characterized as a stem formation that involves the reduplication of (part of) the verb root and the insertion of a suffix *-la*, which in certain tokens is transcribed as stronger *-da*, weaker *-ya*, and nasal *-na*.

**Table 5:** Four patterns of reduplication stems in Kwadi.

Pattern	REDUPLICAND	<i>-la</i>	=ROOT	Number of verb lexemes
I	CV(V)	<i>-la</i>	=CV(V)	25/26
II	CV	<i>-la</i>	=CVN	3/4
III	CV	<i>-la</i>	=CVCV(CV)	6/7
IV	CVCV	∅	=CVCV	6/7

In Table 5, I identify four different sub-patterns, which appear to depend on the shape of the input root and appear to target with very few exceptions two syllable templates as output, namely  $\sigma\sigma$  in the case of the patterns I and II, and  $\sigma\sigma\sigma$  in the case of III and IV, whereby the second syllable is almost universally /la/. The

four patterns display a diverse frequency in terms of verb lexemes, as shown in the rightmost table column.<sup>3</sup>

(19) Kwadi (Westphal 1964/5; Güldemann 2001: 62)

*ta kx'v-la-kx'v*

1SG DUP-?-fear

'I am afraid'

The most frequent reduplication type I is illustrated in (19). It also subsumes a few verbs ('bite', 'buy', 'cough', and possibly 'dance') where the base apparently has a diphthong and it remains unclear whether the reduplication process copies both or only the first vowel.

(20) Kwadi (Westphal 1964/5; Güldemann 2001: 80, 81)

a. *kú-lá-kũŋ* 'hear'

b. */hũ-dà-/hũŋ* 'smell'

c. *sé-lá-sèŋ* 'sleep'

d. *kxá-lá-kàm.mà* 'clap'

Potential cases of pattern II, given exhaustively in (20), are few and provide an inconsistent picture. The verb 'hear' has an alternative in pattern III; 'hear', 'smell', and 'sleep' involve a final velar nasal in the base, which could well be just a nasal vowel changing the structure to  $C\tilde{V}-la=C\tilde{V}$  of pattern I; and the base of 'clap' differs significantly from its reduplicand.

(21) Kwadi (Westphal 1964/5; Güldemann 2001: 68)

*ta tu-la-tumu*

1SG DUP-?-swallow

'I swallow'

Type III, as in (21), can be explained by a targeted syllable template  $\sigma\sigma\sigma$  that simplifies an expected pattern  $CVCV(CV)-la=CVCV(CV)$  of a polysyllabic base root by reduplicating only the first syllable to  $CV-la=CVCV(CV)$ . The data only contain two

<sup>3</sup> Two verbs appear to occur in more than one pattern, explaining the alternative numbers. They are 'hear' in the patterns II and III (*ku-la-kũŋ* from *kũŋ* vs. *ku-laa-kumu* from *kumu*) and 'sing, (dance)' in the patterns I and IV (*'e-la-'e* from *'e* vs. *'ela-'ela* from *'ela*). These data may be due to erroneous transcription and/or morphological analysis rather than represent genuine lexical alternation.

cases with a trisyllabic base ('like; lick, taste' and 'write'), the second of which is a Kuvale borrowing.

(22) Kwadi (Westphal 1964/5; Güldemann 2001: 52)

- a. *tà wólá-wòlà*  
1SG DUP-work
- b. *tà wó-lá-wòlà*  
1SG DUP?-work  
'I work'

Finally, pattern IV looks at first glance like full reduplication, as in (22a). However, almost all verb bases in this pattern have a second syllable *la* (or in one case *da*). The single exception is *tanga-tanga* from *tanga* 'read' – a formation based on a Kuvale loan. In Table 5, I give the account in terms of full reduplication but without the last case, the reduplicand could also be viewed to be just the initial CV of the lexical root followed by grammatical *-la*, as in (22b), so that pattern IV could be subsumed under III.

The hypothesis about output constraints related to weight and template of the root base also receives support from the unique case of *la-labɔla* from *labɔla* 'speak' (cf. (23) below). While the output would suggest yet another pattern *la=CVCVCV*, which is close to but not identical to pattern III, I assume that this verb is another exception due to its initial syllable being itself *la*. Pattern III is expected to yield *la.la.labɔla*; so the shorter form encountered may derive from dropping one *la* for reasons of weight and/or euphony.

Overall, I consider pattern I to be the original and all others to be derived from it in one way or another in connection with different verb base shapes. Thus, the phenomenon as a whole is assumed to have started out from a structure [[REDUPLICATION-*la*]-VERB.ROOT], in which all components were monosyllabic.

**Table 6:** Conjugational forms in Kwadi attested with the reduplication stem.

Form	(Possible) meaning	Example
stem	Present ~ progressive	(23a)
stem- <i>na</i>	Present ~ progressive	(23b/d)
<i>ka</i> stem <sup>4</sup>	Future	(23c)

<sup>4</sup> Cf. the future form of Kuvale, the target of language shift, which has a verb prefix *-ka-*. Thus *mi-ka-popya* 'I will speak' (Westphal 1964/5, Güldemann 2001: 58).

Reduplication stems are found in the data in three different TAM contexts of the attested conjugation system, as listed in Table 6 and exemplified with one verb in (23a–c).

(23) Kwadi

- a. (Westphal 1964/5; Güldemann 2001: 71)  
*tá lá-lábòlà*  
 1SG DUP-speak  
 ‘I speak.’
- b. (Westphal 1964/5; Güldemann 2001: 58, 71)  
*tá lá-lábòlà-nà*  
 1SG DUP-speak-TA  
 ‘I speak.’
- c. (Westphal 1964/5; Güldemann 2001: 58)  
*ta ka la-labòla*  
 1SG FUT DUP-speak  
 ‘I will speak.’
- d. (Westphal 1964/5; Güldemann 2001: 71)  
*ta kwade tʃi lap<sup>ba</sup>la-na*  
 1SG Kwadi ? speak-TA  
 ‘I speak Kwadi.’

Given the profile of the three TAM forms, one is tempted to associate the stem pattern with imperfective meaning to help identify some more general common denominator. However, a quite different but significant observation about the occurrence of the stem pattern emerges from the contrast between (23b) and (23d) and concerns syntax. That is, the base form of the verb is used in (23d) where it is preceded by the object argument *kwade tʃi* ‘Kwadi language’. Examples (24)–(26) present further minimal sentence pairs with one and the same verb but different syntactic configurations.

(24) Kwadi (Westphal 1964/5; Güldemann 2001: 36, 71)

- a. *ala ʔnyũ-la-ʔnyu*  
 1PL DUP-?-eat  
 ‘We (are) eat(ing).’
- b. *ta kɔ-la ʔnu*  
 1SG meat-? eat  
 ‘I eat meat.’

- (25) Kwadi (Westphal 1964/5; Güldemann 2001: 51)
- a. *ta pɛ-la-pɛ*  
1SG DUP-?-put  
'I put.'
  - b. *tshǒ\_khàʃ<sup>b</sup>à pɛ*  
in.front put  
'Put in front.'
- (26) Kwadi (Westphal 1964/5; Güldemann 2001: 53, 63)
- a. *ta kǒ-lá-kǒ*  
1SG DUP-?-go  
'I go.'
  - b. *ta thú-la: kǒ*  
1SG night-? go  
'I (will) come [presumably: go] at night.'

These data suggest a more general complementary distribution: reduplication stems lack preceding objects or adjuncts, while sentences with such constituents lack the marked stem pattern (except for a single and doubtful case). This apparent regularity does not just indicate that the reduplication stem is associated with **intransitive** predicates. The fact that the string [REDUPLICAND-*la*] appears in linear terms where non-verbal participants, including verb complements, occur also suggests that both constituent types can be equated. The reduplication stem would then be (derived from) a hierarchical syntactic structure in which the final verb root is the head that controls a dependent verb form, just like a plain verb controls its argument. It should also be observed that some preverbal non-subject constituents are marked themselves by an element *la*, as is the case in (24b) and (26b). The exact nature of this element remains unclear but if related to the verb affix *-la-* of the reduplication stem, the present hypothesis would, if anything, be supported.

The phenomenon whereby a complex predicate comprises a verb root and a cognate root copy (in potentially reduced form) may look strange at first glance but is in fact quite common elsewhere in Africa and beyond (see Bond and Anderson (2014) for an extensive cross-African survey). There are different variants that can be placed on a scale between more syntactic and more morphologized forms but their ultimate origin appears to lie in constructions of a marked information structure with mostly predicate-centered focus. For the more syntactic end of the scale, reference can be made to such cases as Ewe (cf. Essegbey 1999) featuring so-called “inherent complement verbs” in cognate-object constructions triggered in intransitive clauses. A morphologized variant may be found in certain Chadic languages

which possess conjugation paradigms with root reduplication (cf., e.g., Wolff (2007) for Lamang and close relatives).

The apparent restriction of the reduplication stem in Kwadi to intransitive clauses may indicate its persistent syntactic nature, despite its strongly morphological appearance, but the insufficient data do not allow one to assess this issue conclusively. In any case, I conclude that the reduplication stem arguably also originates in a hierarchical structure where *-la* marks the initial verbal constituent that is dependent on a following cognate verb root.

### 3.3 Summary

Summarizing the two preceding sections, Kwadi possesses two complex verbal predicates with similar morphemes. It is not fully clear but probable that *-(a)la* and *-la* are cognate, as their constructions can be analyzed in a parallel fashion, as shown in (27).

- (27) Kwadi
- |    |   |  |
|----|---|--|
| a. | [[VERB.ROOT <i>-(a)la</i> ] <i>-‘want’</i> ]    | Volition construction<br>(Section 3.1)           |
| b. | [[VERB.ROOT <i>-la</i> ] <i>-VERB.COGNATE</i> ] | Intransitive reduplication stem<br>(Section 3.2) |

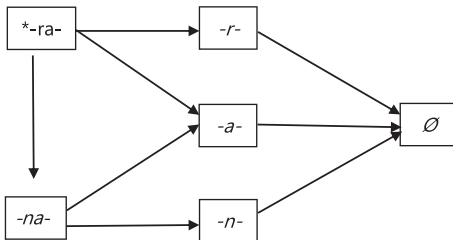
The initial strings [VERB-*(a)la*] are implied to originate in a non-finite verb form depending syntactically on the final verb. Westphal himself, although never coming to provide a fuller and explicit description and analysis, had the same intuition when speaking about the “formation of verbo-nominals (infinitives) with an infix *-la-*” (1963: 247).

## 4 Khoe-Kwadi and the history of the Khoe juncture

In Section 2 I have provided an updated survey of juncture-verb constructions in the Khoe family with several crucial findings. Following Rapold (2014), such a morpheme should be reconstructed for the entire Khoe family, not just its Kalahari branch. This originates in an asymmetrical MVC of the form \*[[VERB.ROOT-JUNCTURE]-VERB.ROOT]. The final verb of this hierarchical syntagm is the head (developing in certain environments to a grammatical suffix), while the initial verb is a dependent form controlled by the former. The juncture itself causes tonal change

on the preceding verb and marks it as in construction with a following lexeme or gram, thus functioning as a dependency marker. The major open problem is to account for the juncture's segmental allomorphy. The available hypothesis by Heine (1986) and Voßen (1997, 2010) of an original *\*-a* whose modern allomorphs in *-r(a)-* and *-n(a)-* are due to etymological consonants in the preceding verb root does not provide a fully plausible explanation of the modern picture (cf. Voßen 1997: 358–359).

This is where Kwadi comes into play: as argued in Section 3, Kwadi possesses two asymmetrical MVCs of the original structure [[VERB.ROOT-(*a*)*la*]-VERB.ROOT]. In view of its virtual identity with the Proto-Khoe juncture pattern and the similarity between Kwadi *-(a)la* and the set of juncture allomorphs *r(a)*, *n(a)*, and *a*, a new hypothesis emerges: similar to *-(a)la* in Kwadi, the proto-Khoe form could have been *\*-ra* rather than simple *\*-a*. Its allomorph *-na* would reflect an areally recurrent phonetic variant of *\*-ra* and its remaining shorter reflexes would be part of a historical cline of lenition and segmental erosion, as schematized in Figure 5. This scenario can be seen as a variation of a more general areal theme in that the segmental erosion of the juncture suffix restores the default bimoraic lexical template (cf. Nakagawa et al. 2023). Since the tonal changes caused by the juncture affect the preceding verb root, it comes as no surprise that this is its historically most resilient trace across the entire family.



**Figure 5:** Allomorphy and segmental erosion of the Khoe juncture.

- (28) Kwadi (Westphal 1964/5; Güldemann 2001: 56, 53)  
*ta kw-ãna-xε* cf. *kũ(ũ)* ‘go, walk’ (Fehn and Rocha 2023, Suppl.: 22)  
 1SG go-?-want  
 ‘I want to go away.’
- (29) Kwadi (Westphal 1964/5; Güldemann 2001: 56)  
*tàà ũ-ná-ʒnũ* cf. *ʔũ(ũ)* ‘eat’ (Fehn and Rocha 2023, Suppl.: 64)  
 ?1SG DUP-?-eat  
 ‘He [possibly: I] eat[s].’

It is also significant that some changes assumed for the Khoe juncture turn up in Westphal's (1964/5) variable transcriptions of the Kwadi suffix *-(a)la*, suggesting that its lenition was already a latent process there. This is shown in the volition construction in (28) and the intransitive reduplication stem in (29) where Westphal perceived the suffix as *-na*, apparently due to assimilation to the nasal vowel of the preceding roots.

Table 7 shows the comparison between the three subtypes of juncture-verb construction in Khoe and the two verbal complex predicates in Kwadi. All five patterns have a similar basic structure, suggesting the Proto-Khoe-Kwadi reconstruction in (30).

**Table 7:** Constructional comparison between the juncture in Khoe and *-(a)la* in Kwadi.

Domain		Khoe	Kwadi
Form		*-ra	-(a)la
Syntax		*[[ROOT-ra] <sub>DEPENDENT</sub> -ROOT <sub>HEAD</sub> ]	[[ROOT-(a)la] <sub>DEPENDENT</sub> -ROOT <sub>HEAD</sub> ]
Function	1	*ROOT-ra -TAM	ROOT-(a)la -TAM
	2	*ROOT-ra -DERIVATION	-
	3	*ROOT <sub>x</sub> -ra -ROOT <sub>y</sub>	-
	4	-	ROOT <sub>x</sub> -la -ROOT <sub>x</sub>

(30) \*[[ROOT-Ra]<sub>DEPENDENT</sub>=ROOT<sub>HEAD</sub>]

What differs between Khoe and Kwadi is the functional profile of the constructions derived from the proto-structure in (30). The only shared pattern is that in Function line 1 of Table 7, where the final main verb is an auxiliary that encodes TAM features. The three remaining patterns, two in Khoe and one in Kwadi, seem to be unique to each unit. How can this be reconciled with the hypothesis of common descent?

A first observation in this respect is that only the shared type for TAM marking in Function line 1 of Table 7 and the intransitive reduplication stem of Kwadi in line 4 fully conform with the head-final structure assumed for Proto-Khoe-Kwadi in that the syntactic head is straightforwardly also the semantic head. This is in line with my assumption that the asymmetric MVC of (30) is a typologically recurrent periphrastic structure in which a final auxiliary as the syntactic main verb controls a preceding non-finite verb, marked as such by means of the suffix \*-Ra. This is harder to transfer to the two Khoe patterns in Function lines 2 and 3 of Table 7. Especially in the diverse MVC types, the initial dependent verb must often be construed as the main verb from a semantic and syntactic viewpoint. For example, in a sequential cause-effect pattern like 'see-know' = 'recognize' in (4a), or in a

motion-path pattern like ‘run-exit’ = ‘run out’ in (4d), it is the first verb that encodes the temporally prior and thus principal state of affairs.

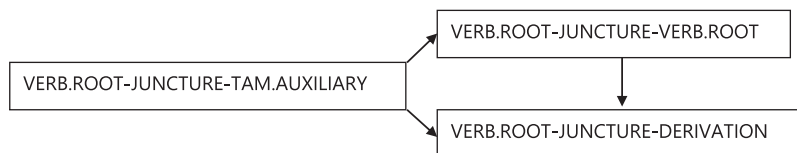
However, this apparent tension of (earlier) syntactic constituency vs. current semantics should be scrutinized against the background of the areal context of the Khoe family. As opposed to Kwadi, Khoe languages are deeply entrenched in the Kalahari Basin area which also hosts head-initial languages of two other families, namely Kx’a and Tuu. These display head-initial serial verb constructions precisely in the functional domains of the two relevant types of juncture-verb construction in lines 2 and 3 of Table 7 (cf. Güldemann and Fehn 2017: 510–511, Fehn and Phiri 2022). It is thus not unlikely that these two patterns developed within the areal context of the Kalahari Basin and are thus innovative vis-à-vis Proto-Khoe-Kwadi and Kwadi.

There is also a formal parallel between juncture-verb constructions in Khoe and serial verb constructions in Kx’a and Tuu languages, where the verbs generally do not display a linker morpheme. I have proposed above the gradual decline of the segmental substance of a proto-juncture \*-ra in Khoe, namely from CV over V or C to Ø, leaving in the end only suprasegmental traces (cf. Figure 5). Such changes have made the Khoe structure more similar to verb serialization and compounding in the contact languages of Kx’a and Tuu.

That juncture erosion is at least partly due to language contact rather than plain language-internal drift toward reduction of phonetic material in grams is suggested by its geographical distribution profile. In particular, Khoekhoe, the group that encroached most on the Kalahari Basin, is the one that completed the overall trend toward reducing the juncture segmentally. Güldemann’s (2006: 116–119) study on the Cape linguistic area is the first work that proposed contact influence from local forager languages on those of colonizing Khoe groups in the domain of predicate formation and in particular MVCs. The hypothesis there must be amended in light of the above discussion. That is, Khoekhoe like its predecessor possessed the juncture-verb construction and thus certain types of complex predicates before the contact with Tuu, as convincingly argued by Rapold (2014). Assuming Tuu substrate influence in Khoekhoe MVCs remains viable, notably for the segmental erosion of the juncture, the increase in the variation of relevant MVC types, including the occurrence of structures with more than two verbs, and possibly the increase in the MVC token frequency. It does not account, however, for the salient presence of MVCs as such.

At the same time, the emergence of a good portion of MVC types in Khoe may still have been mediated by areal convergence, namely during the very formation of this family, which involved contact with Kx’a languages (see Güldemann 2008).

This resembles the situation for the marking of person, gender, and number for which Güldemann (2019) also argues for intensive early contact between Kx'a and Pre-Khoe.



**Figure 6:** Historical chronology of the three juncture-verb constructions in Khoe.

The picture thus emerging is that Proto-Khoe possessed the inherited structure in (30) for auxiliary periphrasis but later language states came to re-use it increasingly for new functions expressed in Kx'a and Tuu contact languages by verb serialization. This also calls for a change of the theoretically expected semantic map in Figure 2 of Section 2.2 toward the different historical chronology in Figure 6, whose lack of semantic plausibility can be motivated by contact interference that is not steered primarily by language-internal factors.

Language contact may also be relevant for the emergence of the unique Kwadi structure in line 4 of Table 7 – here, however, with different contact partners. It is beyond doubt that Bantu languages had a strong impact on Kwadi and recent research (Güldemann, Smith and Bajić forthcoming) even entertains the idea that Kwadi may have a Bantu substrate. Against this background, it is relevant that constructions involving the repetition of the verb root are not just pervasive in Africa but in fact particularly prominent in the Bantu family, including languages in the northwest of the Kalahari Basin close to Kwadi (Güldemann and Fiedler 2022). Hence, it is not farfetched to assume that the intransitive reduplication stem is partly a Bantu legacy not shared by Khoe languages further south.

I close by way of a more general point on historical linguistics. Proposals of new non-obvious genealogical relationships, if viable, can crucially inform the evaluation of the genealogical units, which are implied in the hypothesis and have already been partially reconstructed. This is indeed the case for Khoe-Kwadi with Güldemann's (2004) establishment of a shared system of person-gender-number marking and the related Proto-Khoe reconstructions. It is also what I hope to have achieved with this contribution in that my proposal has considerably altered the historical reconstruction of juncture-verb constructions in Khoe by associating them with Kwadi predicate structures.

## Abbreviations

1/3	first/third person
ADJR	adjectivizer
BEN	benefactive
C	consonant
CONJ	conjunction
CONT	continuative
COP	copula
DEM	demonstrative
DU	dual
DUP	reduplication
F	feminine
FUT	future
IPFV	imperfective
JUN	junction
LC	language complex
LOC	locative
M	masculine
MPO	multi-purpose oblique
MVC	multi-verb construction
N	noun
NEAR.PST	near past
OBJ	object
PASS	passive
PFV	perfective
PGN	person-gender-number
PL	plural
PK	Proto-Khoe
PKaIK	Proto-Kalahari-Khoe
POSS	possession
PRO	pronoun
PRS	present
PST	past
R	resonant
REF	referential
SBJ	subject
SG	singular
TA(M)	tense-aspect-(modality)
V(R)	verb (root)

## References

- Aikhenvald, Alexandra Y. 2006. Serial verb constructions in typological perspective. In Alexandra Y. Aikhenvald & Robert M. W. Dixon (eds.), *Serial verb constructions: a cross-linguistic typology*, 1–68. Oxford: Oxford University Press.
- Blench, Roger M. 2017. African language isolates. In Lyle Campbell (ed.), *Language isolates*, 162–192. London: Routledge.
- Bond, Oliver & Gregory D. S. Anderson. 2014. Aspectual and focal functions of cognate head-dependent constructions: evidence from Africa. *Linguistic Typology* 18(2). 215–250.
- Chebanne, Anders M. & Chris Collins. 2017. Tense and aspect in Kua: a preliminary assessment. In Anne-Marie Fehn (ed.), *Khoisan languages and linguistics: proceedings of the 4th International Symposium July 11–13, 2011, Riezlern/Kleinwalsertal*, 91–108. (Quellen zur Khoisan-Forschung 36). Cologne: Rüdiger Köppe.
- Ehret, Christopher. 1982. The first spread of food production to southern Africa. In Ehret, Christopher & Merrick Posnansky (eds.), *The archeological and linguistic reconstruction of African history*, 158–181. Berkeley: University of California Press.
- Elderkin, Edward D. 1986. Kxoe tone and Kxoe ‘jonctures’. In Rainer Vossen & Klaus Keuthmann (eds.), *Contemporary studies on Khoisan*. 2 vols, vol. 1: 225–235. (Quellen zur Khoisan-Forschung 5). Hamburg: Helmut Buske.
- Elderkin, Edward D. 2016. Tonal patterns in Khwe verb conjugation. In Rainer Voßen & Wilfrid H. G. Haacke (eds.), *Lone Tree scholarship in the service of the Koon: essays in memory of Anthony T. Traill*, 119–141. Cologne: Rüdiger Köppe.
- Essegbey, James. 1999. *Inherent complement verbs revisited: towards an understanding of argument structure in Ewe*. (MPI Series in Psycholinguistics 10). Wageningen: Ponsen and Looijen.
- Fehn, Anne-Maria. 2014. *A grammar of Ts’ixa (Kalahari Khoe)*. Köln: Universität zu Köln Ph.D. thesis. <http://kups.ub.uni-koeln.de/7062/>
- Fehn, Anne-Maria & Admire Phiri. 2022. Juncture-verb constructions in Northeastern Kalahari Khoe: a comparative perspective. *Stellenbosch Papers in Linguistics Plus* 65. 147–190.
- Fehn, Anne-Maria & Jorge Rocha. 2023. Lost in translation: a historical-comparative reconstruction of Proto-Khoe-Kwadi based on archival data. *Diachronica* 40(5). 609–665.
- Guerreiro, Manuel Viegas. 1971. Vida humana no deserto de Namibe: Onguaia. *Finisterra* 6(11). 84–124.
- Güldemann, Tom. 2001. *Linguistic analysis of Westphal’s unpublished field notes on Kwadi*. Unpublished manuscript.
- Güldemann, Tom. 2004. Reconstruction through ‘de-construction’: The marking of person, gender, and number in the Khoe family and Kwadi. *Diachronica* 21(2). 251–306.
- Güldemann, Tom. 2006. Structural isoglosses between Khoekhoe and Tuu: The Cape as a linguistic area. In Yaron Matras, April McMahon & Nigel Vincent (eds.), *Linguistic areas: convergence in historical and typological perspective*, 99–134. Hampshire: Palgrave Macmillan.
- Güldemann, Tom. 2008. A linguist’s view: Khoe-Kwadi speakers as the earliest food-producers of southern Africa. *Southern African Humanities* 20. 93–132.
- Güldemann, Tom. 2010. Kwadi: from family-level to family-internal isolate. Paper presented to the International Workshop ‘Language Isolates in Africa’, Lyon, 3–4 December, 2010. <https://www.iaaw.hu-berlin.de/de/region/afrika/afrika/linguistik/mitarbeiter/prof-dr-phil-tom-gueldemann/dokumente-unpublished/13-gueldemann-tom-2010-kwadi-from-family-level-to-family-internal-isolate.pdf>

- Güldemann, Tom. 2013a. Morphology: Kwadi. In Rainer Vossen (ed.), *The Khoesan languages*, 261–263. London: Routledge.
- Güldemann, Tom. 2013b. Typology. In Rainer Vossen (ed.), *The Khoesan languages*, 25–37. London: Routledge.
- Güldemann, Tom. 2014. “Khoisan” linguistic classification today. In Tom Güldemann & Anne-Maria Fehn (eds.), *Beyond ‘Khoisan’: historical relations in the Kalahari Basin*, 1–41. Amsterdam / Philadelphia: John Benjamins.
- Güldemann, Tom. 2019. Person-gender-number marking from Proto-Khoe-Kwadi to its descendants: a rejoinder with particular reference to language contact. In Rainer Voßen & Christa König (eds.), *Patterns of linguistic convergence in Africa*, 53–73. (Frankfurter Afrikanistische Blätter 27/8). Cologne: Rüdiger Köppe.
- Güldemann, Tom & Edward D. Elderkin. 2010. On external genealogical relationships of the Khoe family. In Matthias Brenzinger & Christa König (eds.), *Khoisan Languages and Linguistics: Proceedings of the 1st international symposium January 4–8, 2003, Riezlern/Kleinwalsertal*, 15–52. (Quellen zur Khoisan-Forschung 24). Cologne: Rüdiger Köppe.
- Güldemann, Tom & Anne-Maria Fehn. 2014. A Kwadi perspective on Khoe juncture-verb constructions. Paper presented at the 5th International Symposium “Khoisan Languages and Linguistics”, Riezlern/Kleinwalsertal, 14–16 July, 2014. <https://www.iaaw.hu-berlin.de/de/region/afrika/afrika/linguistik/mitarbeiter/prof-dr-phil-tom-gueldemann/dokumente-unpublished/19-gueldemann-tom-and-anne-maria-fehn-2014-a-kwadi-perspective-on-khoe-juncture-verb-constructions.pdf>
- Güldemann, Tom & Anne-Maria Fehn. 2017. The Kalahari Basin area as a “Sprachbund” before the Bantu expansion. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 500–526. Cambridge: Cambridge University Press.
- Güldemann, Tom & Ines Fiedler. 2022. Predicate partition for predicate-centred focus and Meeussen’s Proto-Bantu “advance verb construction”. In Koen A. G. Bostoen, Gilles-Maurice de Schryver, Rozenn Guérois & Sara Pacchiarotti (eds.), *On reconstructing Proto-Bantu grammar*, 537–580. (Niger-Congo Comparative Studies 4). Berlin: Language Science Press.
- Güldemann, Tom, Andrew B. Smith & Vladimir Bajić. forthcoming. The archeolinguistics of Kalahari Basin area languages. In Martine Robbeets & Mark Hudson (eds.), *Oxford Handbook of Archaeology and Language*. Oxford: Oxford University Press.
- Haacke, Wilfrid H. G. 1999. *The tonology of Khoekhoe (Nama/Damara)*. (Quellen zur Khoisan-Forschung 16). Cologne: Rüdiger Köppe.
- Haacke, Wilfrid H. G. 2014. Verb serialization in northern dialects of Khoekhoegowab: convergence or divergence? In Tom Güldemann & Anne-Maria Fehn (eds.) *Beyond ‘Khoisan’: historical relations in the Kalahari Basin*, 123–149. Amsterdam/Philadelphia: John Benjamins.
- Heine, Bernd. 1986. Bemerkungen zur Entwicklung der Verbaljunkturen im Kxoe und anderen Zentralkhoisan-Sprachen. In Rainer Vossen & Klaus Keuthmann (eds.), *Contemporary studies on Khoisan*. 2 vols., vol. 2: 9–21. (Quellen zur Khoisan-Forschung 5). Hamburg: Helmut Buske.
- Kilian-Hatz, Christa. 2004. Verb derivation in Khwe (Central-Khoisan). *Annual Publication in African Linguistics* 2. 109–135.
- Kilian-Hatz, Christa. 2006. Serial verb constructions in Khwe (Central-Khoisan). In Alexandra Y. Aikhenvald & Robert M. W. Dixon (eds.), *Serial verb constructions: a cross-linguistic typology*, 108–123. Oxford: Oxford University Press.

- Kilian-Hatz, Christa. 2008. *A grammar of modern Khwe (Central Khoisan)*. (Quellen zur Khoisan-Forschung 23). Cologne: Rüdiger Köppe.
- Kilian-Hatz, Christa. 2010. Serial verb constructions vs. converbs in Khwe. Matthias Brenzinger & Christa König (eds.), *Khoisan Languages and Linguistics: Proceedings of the 1st international symposium January 4–8, 2003, Riezlern/Kleinwalsertal*, 115–143. (Quellen zur Khoisan-Forschung 24). Cologne: Rüdiger Köppe.
- Köhler, Oswin. 1981. Les langues Khoisan, section 1: Présentation d'ensemble. In Jean Perrot (ed.), *Les langues dans le monde ancien et moderne, première partie: Les langues de l'Afrique subsaharienne*, 455–482. Paris: Centre National de la Recherche Scientifique.
- Köhler, Oswin. 1989. *Die Welt der Kxoe-Buschleute im südlichen Afrika: Eine Selbstdarstellung in ihrer eigenen Sprache, Band I: Die Kxoe-Buschleute und ihre ethnische Umgebung*. Berlin: Dietrich Reimer.
- Meinhof, Carl. 1930. Der Koranadialekt des Hottentottischen. *Zeitschrift für Eingeborenen-Sprachen*, Beiheft 12. Berlin: Dietrich Reimer.
- Nakagawa, Hiroshi. 2006. *Aspects of the phonetic and phonological structure of the Glui language*. Johannesburg: University of the Witwatersrand Ph.D. dissertation.
- Nakagawa, Hiroshi. 2013. Syntax: IIGana subgroup: IGui. In Rainer Vossen (ed.), *The Khoisan languages*, 394–401. London: Routledge.
- Nakagawa, Hiroshi, Alena Witzlack-Makarevich, Daniel Auer, Anne-Maria Fehn, Linda Ammann Gerlach, Tom Güldemann, Sylvanus Job, Florian Lionnet, Christfried Naumann, Hitomi Ono & Lee J. Pratchett. 2023. Towards a phonological typology of the Kalahari Basin Area languages. *Linguistic Typology* 27(2). 509–535.
- Rapold, Christian. 2014. Areal and inherited aspects of compound verbs in Khoekhoe. In Tom Güldemann & Anne-Maria Fehn (eds.), *Beyond 'Khoisan': historical relations in the Kalahari Basin*, 151–175. Amsterdam / Philadelphia: John Benjamins.
- Visser, Hessel. 2010. Verbal compounds in Naro. In Matthias Brenzinger & Christa König (eds.), *Khoisan Languages and Linguistics: Proceedings of the 1st international symposium January 4–8, 2003, Riezlern/Kleinwalsertal*, 176–200. (Quellen zur Khoisan-Forschung 24). Cologne: Rüdiger Köppe.
- Voßen, Rainer. 1997. *Die Khoe-Sprachen: Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. (Quellen zur Khoisan-Forschung 12). Cologne: Rüdiger Köppe.
- Vossen, Rainer. 2010. The verbal 'linker' in Central Khoisan (Khoe) in the context of deverbal derivation. *Journal of Asian and African Studies* 80. 47–60.
- Westphal, Ernst O. J. 1962. A re-classification of Southern African non-Bantu languages. *Journal of African Languages* 1. 1–8.
- Westphal, Ernst O. J. 1963. The linguistic prehistory of southern Africa: Bush, Kwadi, Hottentot, and Bantu linguistic relationships. *Africa* 33. 237–265.
- Westphal, Ernst O. J. 1964/5. Kwadi field notes. Cape Town: Manuscripts and Archives Department, University of Cape Town, unpublished manuscript.
- Westphal, Ernst O. J. 1965. Linguistic research in S.W.A. and Angola. In *Die ethnischen Gruppen in Südwestafrika*, 125–144. (Wissenschaftliche Forschung in Südwestafrika 3). Windhoek: S.W.A. Wissenschaftliche Gesellschaft.
- Westphal, Ernst O. J. 1971. The click languages of southern and eastern Africa. In Jack Berry & Joseph H. Greenberg (eds.), *Linguistics in Sub-Saharan Africa*, 367–420. (Current Trends in Linguistics 7). The Hague/ Paris: Mouton.
- Westphal, Ernst O. J. n.d.a. *Kwepe vocabulary*. Cape Town: Manuscripts and Archives Department, University of Cape Town, unpublished manuscript.

- Westphal, Ernst O. J. n.d.b. *The Kwadi language (for the use of Prof. António de Almeida)*. Cape Town: Manuscripts and Archives Department, University of Cape Town, unpublished manuscript.
- Westphal, Ernst O. J. n.d.c. *Word list of Angolan click languages*. Cape Town: Manuscripts and Archives Department, University of Cape Town, unpublished manuscript.
- Wolff, H. Ekkehard. 2007. Reduplication, aspect, and predication focus in Central Chadic: what Lamang and Hdi tell about Malgwa verb morphology. In Tourneux, Henry (ed.), *Topics in Chadic linguistics IV: comparative and descriptive studies; Papers from the 3rd Biennial International Colloquium on the Chadic Languages, Villejuif, November 24–25, 2005*, 129–155. (Chadic Linguistics 5). Cologne: Rüdiger Köppe.

Julia Nintemann, Maike Vorholt, Deborah Arbes & Kevin Behrens

# OUT OF our minds and ONTO the paper: On the morphological complexity of configurational spatial relations

**Abstract:** This study investigates the morphological complexity of spatial relations across a sample of 30 areally and genealogically diverse languages from a functional-typological perspective in both qualitative and quantitative terms. The focus lies on configurational constructions encoding different Directionalities (Place, Goal, Source) and Configurations (inside, on, behind, in front of, under). A comparison is drawn to constructions encoding general location wherever possible. Our findings indicate a hierarchy of complexity from Place via Goal to Source, supporting prior research.

**Keywords:** spatial relations, configuration, general location, directionality, morphological complexity

## 1 Introduction

In this study, we address a topic that has been one of the focal points in recent years in the linguistic research of Thomas Stolz and his research team in Bremen, viz. spatial relations (cf., e.g. Stolz et al. 2014, 2017a, 2017b; Stolz 2018; Stolz and Nintemann 2024; Robbers and Hober 2018; Nintemann and Robbers 2019; Nintemann et al. 2020; Nintemann and Hober 2023)<sup>1</sup>. The focus of our present study lies on constructions in which Directionality, i.e. location at, movement to, or movement from, and Configuration, i.e. the relative position of an entity to a reference-object, are combined (see Section 2). This is exemplified with the superior Configuration ON in Kolyma Yukaghir in (1).

---

<sup>1</sup> However, Stolz's interest in this topic began even before his time in Bremen, as his 1992 monograph on local case systems (*Lokalkasussysteme*) shows.

---

**Julia Nintemann**, University of Bremen, e-mail: jnintemann@uni-bremen.de

**Maike Vorholt**, University of Bremen, e-mail: vorholt@uni-bremen.de

**Deborah Arbes**, University of Bremen, e-mail: d.arbes@uni-bremen.de

**Kevin Behrens**, University of Bremen, e-mail: kevin.behrens@uni-bremen.de

(1) Kolyma Yukaghir [Yukaghir] (Maslova 2003: 270)<sup>2</sup>

## a. AT + ON

... *irk-in pajpā-die pie budie modo-t*  
 ... one-ATTR woman-DIM mountain **SUPER** sit-SS:IPFV  
*pon'-ō-l'el*  
 leave-RES-INFR(3SG)

'... one woman remained sitting **on top of** the mountain.'

## b. TO + ON

*pulun-die-gele tude budie-n im-delle tāt kōu-de-j-m*  
 old.man-DIM-ACC he **SUPER-DIR**<sup>3</sup> put-SS:PFV CA go-CAUS-PFV-TR:3SG  
 'She put the old man **on** her back and brought him away that way.'

## c. FROM + ON

*tāt unu-d+ørd'e-gen qon-de-ge pie budie-t*  
 CA river-ATTR+middle-PROL go-3SG-DS mountain **SUPER-ABL**  
*jaqte-lek medū-nu-l*  
 sing-PRED be.heard-IPFV-SF

'When he was going along the middle of the river, he could hear a song coming **from the top of** the mountain.'

Configuration in Kolyma Yukaghir is expressed with a postposition, in this case *budie* 'on (the top/surface of)'. Directionality on the other hand is usually expressed with case suffixes. As shown in (1a), however, Place is not overtly coded in combination with a configurational postposition. Goal in (1b) and Source in (1c), however, are expressed with directional and ablative case suffixes, respectively. The example of Kolyma Yukaghir spatial constructions shows that different types of morphemes (i. e. free vs. bound) may be involved in expressing Directionality and Configuration. Furthermore, the examples demonstrate that there is asymmetry between the constructions in that Place is not overtly coded, while Goal and Source are both overtly expressed.

This study aims at investigating how Directionality and Configuration are expressed in a sample of 30 areally and genealogically diverse languages, while also taking account of asymmetries in the morphological complexity of the constructions. In Section 2, we provide our theoretical framework, explain the relevant

<sup>2</sup> The original glossing was retained as close as possible for the provided examples. To avoid doublets, we unified grammatical labels according to the Leipzig Glossing Rules (cf. Comrie et al. 2015). If the original source does not provide glosses and/or an (English) translation, this will be indicated in a footnote. Relevant spatial expressions are marked in bold.

<sup>3</sup> In the original example in Maslova (2003: 270), *-n* is falsely glossed as ABL. We corrected this according to an analogous example (Maslova 2003: 320).

concepts and terminology, and outline our hypotheses. Section 3 serves to outline our methodology, while Section 4 presents a qualitative analysis of different types of spatial constructions. A quantitative analysis is offered in Section 5. Section 6 concludes this paper.

## 2 Theoretical background, terminology, and hypotheses

According to Talmy (1985: 60–61), “[t]he basic motion event consists of one object (the ‘Figure’) moving or located with respect to another object (the reference-object or ‘Ground’)”. The Ground can assume different roles in a spatial situation. The three basic spatial relations (SR) which encode Directionality considered in this study are

- Place, i.e. the location of the Figure (= AT),
- Goal, i.e. the endpoint of the movement of the Figure (= TO), and
- Source, i.e. the starting point of the movement of the Figure (= FROM).

Creissels (2006: 19) assumes that “[a]ll languages must encode in some way or another the distinction between localisation, the source of motion, and the destination of motion, but they differ in the way spatial adpositions or case affixes participate in the encoding of this distinction.” Similarly, we expect that all languages have the means to express different types of spatial Configuration, i.e. the relative position of the Figure to the Ground, e.g. INSIDE, ON, UNDER, etc.

Based on these assumptions, we examine how languages express Directionality and Configuration in combination from a functional-typological perspective and use canonical typology as outlined in Corbett (2005) as a yardstick to compare the occurring forms:

In a canonical approach, we take definitions to their logical end point and build theoretical spaces of possibilities. Only then do we ask how this space is populated. [. . .] It follows that canonical instances (the best examples, those most closely matching the canon) may well not be the most frequent. They may indeed be extremely rare, or even non-existent. However, they fix a point from which occurring phenomena can be calibrated, and it is then significant and interesting to investigate frequency distributions. (Corbett 2005: 26)

The starting point for a comparison within the canonical framework is an idealised paradigm, which contains all logical combinations of the relevant features. This is beneficial to this study, as it enables a data-driven approach to document all possi-

ble forms used in different languages. Table 1 presents the canonical constructions for the expression of Directionality and Configuration including all relevant components.

**Table 1:** Canonical constructions for the expression of spatial relations (Stolz [to appear], following Lestrade 2010: 73).

SR	Figure	Verb	Directionality	Configuration	Ground
<b>Place</b>	Figure	±dynamic	X	Configuration	Ground
<b>Goal</b>	Figure	+dynamic	Y	Configuration	Ground
<b>Source</b>	Figure	+dynamic	Z	Configuration	Ground

Table 1 presents the structure of the sentences that are the object of this study as well as the terminology used. In this study, we focus on the cells highlighted with grey shading. We neglect the shape of the Figure entirely and mention the verb briefly when necessary, but exclude this category from the analysis. In the following, we apply the canonical framework only to the cells denoting Directionality and Configuration in combination with a very limited scope of possible Grounds (namely common nouns, see Section 3.1). The centre of attention therefore lies on the nominal domain and we also include examples of phrases without a verb. Additionally, it is important to keep in mind that the order in which the individual elements occur in Table 1 is an arbitrary example and varies from one language to another. Word order does thus not have an effect on the canonicity of a construction. Table 2 illustrates a canonical example in English.

**Table 2:** Source and BEHIND in English.

SR	Figure	Verb	Directionality	Configuration	Ground
<b>Source</b>	<i>She</i>	<i>came</i>	<i>from</i>	<i>behind</i>	<i>the house</i>

Both the Directionality and the Configuration cell are filled with one and only one form, i.e. every relevant element is expressed overtly. This is not always the case in English, as visible in Table 3:

**Table 3:** Place/Goal and BEHIND in English.

SR	Figure	Verb	Directionality	Configuration	Ground
<b>Place</b>	<i>She</i>	<i>was</i>		<i>behind</i>	<i>the house</i>
<b>Goal</b>	<i>She</i>	<i>went</i>		<i>behind</i>	<i>the house</i>

The examples for Place and Goal in combination with *behind* are less canonical in that the cell for Directionality is empty. In analogy to Source, one could expect prepositions such as *at* and *to*. However, stating them overtly (*\*she was at behind the house, \*she went to behind the house*) is not considered standard language. In these cases, Directionality seems to be implied without requiring a preposition.

The canon makes no statement about what kind of morpheme is canonical, i.e. different types of markers for Directionality and Configuration are possible, such as case markers, adpositions, or relational nouns. Hagège (2010: 291) notes that

in languages with both case affixes and [adpositions], spatial *directions*, such as “into”, “out of”, “across”, etc. will often be expressed by bound morphemes, while spatial *dimensions*, such as “inside”, “above”, “in front of”, “beside”, will tend to be expressed by independent morphemes like [adpositions]. [original italics]

For our qualitative analysis in Section 4, we consider what kind of morphemes are employed to express Directionality and Configuration, however, for the quantitative part (Section 5), their form does not play a role. Morphemes are counted no matter whether they are bound or free.

In many languages, it is possible to mark the Ground in a spatial construction without specifying the exact position in which the Figure is located relative to the Ground. In these cases, *general location* (GL) is expressed, see (2a). In contrast, a more exact localisation of the Figure in relation to the Ground can be expressed by including a Configuration, see (2b).

(2) Japanese [Japonic]

a. GL / Place (Hasegawa 2015: 95)

*Midori ga uchi ni iru.*  
Midori NOM home **at** exist  
‘Midori is **at** home.’

b. ON / Place<sup>4</sup> (Tanimori 1994: 289)

*Jisho wa tsukue no ue ni aru.*  
dictionary TOP desk **GEN on at** exist  
‘The dictionary is **on** the desk.’

The Japanese example in (2a) shows that GL is expressed with the particle *ni*, which, depending on the context, can mean ‘at, in, on, to, towards’ as well as ‘for, from, per, in order to’ in non-spatial constructions (Tanimori 1994: 137). As Ursini (2020: 496)

---

<sup>4</sup> The glosses are our own.

puts it for a similar example<sup>5</sup>: “This relation does not individuate a single location that the figure occupies [. . .]. Rather, it implies that the figure may be in any of a heterogeneous set of possible locations.” In contrast, the position of the Figure in relation to the Ground is specified in (2b) as ON. Although we do not expect all languages to have the possibility to express GL, we include it in our analysis for those languages for which GL is attested in our respective sources as a basis for comparison. Concerning the canonical constructions for GL, we assume similar spatial constructions as depicted in Table 1 above, however, without a morpheme that expresses a Configuration, i.e. a combination of Figure, Verb, Directionality, and Ground.

According to the canon, the number of morphemes should be the same for Place, Goal, and Source, both for GL and configurational constructions, respectively. However, past studies have already shown that there are asymmetries in the coding of spatial relations, especially between the two dynamic relations Goal and Source. Kopecka and Vuillermet (2021: 9–10), for example, explain that

[i]t has been observed that linguistic markers, such as adpositions and/or case markers, used by languages for expressing the Goal are morphologically less complex than those used for expressing the Source, and, furthermore, that the expression of the Goal tends to be less constrained and, in general, more ‘straightforward’ than the expression of the Source (cf. Ikegami 1987, Bourdin 1997).

Other studies have shown that there is a general tendency of increasing complexity from Place via Goal to Source (cf. Stolz et al. 2017b; Nintemann et al. 2020).

Different languages have different means to overtly or covertly code these spatial relations. One of the most prominent topics in the linguistic research of spatial relations is that of syncretism, i.e. “the formal identity of the expressions employed for two or more categories” (Stolz et al. 2017b: 11). If we consider all possible combinations of syncretism in the expression of Place, Goal, and Source, the patterns in Table 4 arise, where X, Y, and Z represent formally distinct constructions.

**Table 4:** Logically possible patterns of formal distinctions (adopted from Stolz et al. 2017b: 11).

Option	Place	Goal	Source	Pattern	Word-Forms
I	X	Y	Z	Place ≠ Goal ≠ Source	3
II	X	X	Z	(Place = Goal) ≠ Source	2

<sup>5</sup> The example given by Ursini (2020: 466) is *Mario is waiting at the pub* in contrast to the “more narrowly defined” location expressed in *Mario is waiting in front of the pub*.

Table 4 (continued)

Option	Place	Goal	Source	Pattern	Word-Forms
III	X	Y	Y	Place ≠ (Goal = Source)	2
IV	X	Y	X	Goal ≠ (Source = Place)	2
V	X	X	X	(Place = Goal = Source)	1

The patterns presented in Table 4 occur to various degrees in the languages of the world. Although different studies come to different conclusions as to the exact distribution of these patterns, the general picture is that Options I, II, and V are the most prevalent patterns, while Options III and IV occur less frequently (see, e.g. Creissels 2006; Pantcheva 2010; Lestrade 2010; Stolz et al. 2017b; Nintemann et al. 2020). Only with Option I is it possible to match the canon as depicted in Table 1 – if we exclude the verbal domain and focus solely on the nominal domain – as syncretism is one of the mismatches as postulated by Corbett (2005). Optional elements or different syntactic rules may result in one language attesting to more than one syncretism pattern. Theoretically, it is possible that languages attest to different options for different types of spatial constructions, e.g. there may be one pattern for UNDER in Place, Goal, and Source Constructions, but a different one for BEHIND. We assume, however, that languages tend to stick to the same option(s) independent of the type of Configuration (or GL) expressed.<sup>6</sup>

Based on the results of previous studies as discussed above, we formulate the hypotheses for our study as follows:

(3) Hypothesis 1

On average, morphological complexity increases from Place via Goal to Source both for GL and configurational constructions.

(4) Hypothesis 2

On average, configurational constructions are morphologically more complex than GL.

<sup>6</sup> However, other factors, e.g. the Ground's word class, may have a larger impact on the syncretism pattern. English, for example, attests to Option I (P≠G≠S) when the Ground is a common noun, but to Option II (P=G≠S) with spatial adverbs, see (i).

(i) English [Indo-European, Germanic]

a. Place: *He is **at** the market.* / *He is **∅** there.*

b. Goal: *He goes **to** the market.* / *He goes **∅** there.*

c. Source: *He comes **from** the market.* / *He comes **from** there.*

## (5) Hypothesis 3

Languages employ the same syncretism patterns throughout the paradigms of GL and configurational constructions.

In the following sections, we analyse the constructions under scrutiny in both qualitative and quantitative terms. The qualitative discussion of selected examples in Section 4 offers a look at actual manifestations of both GL and configurational constructions. It will be shown that languages have different means for the expression of spatial relations and that there is great variation between the languages considered in our study. With the quantitative analysis of our data in Section 5, we test our hypotheses as formulated in (3)–(5).

## 3 Methodology and data

### 3.1 Methods

The constructions under scrutiny express a combination of Directionality and Configuration. These configurational constructions can be read as ‘onto a Ground’, ‘from inside a Ground’, ‘behind a Ground’, etc. Given the limited extent of this paper, we restrict our focus to a subset of these. We investigate three spatial directions: location AT (= Place), movement TO (= Goal), and movement FROM (= Source) a given Ground. As for the Configuration, we centre our attention on five basic relations, viz. interior (INSIDE), superior (ON), posterior (BEHIND), anterior (IN FRONT OF), and inferior (UNDER). Some languages distinguish between [+contact] (ON) and [-contact] (ABOVE) in the superior configuration. In these cases, we consider only the option denoting [+contact]. We decided to restrict ourselves to spatial relations involving an inanimate common noun acting as the Ground mainly for two reasons. First of all, toponyms often follow their own rules in spatial constructions in that they are often zero-marked or need shorter markers in Ground position (cf., e.g. Stolz et al. 2014; Stolz and Nintemann 2024).<sup>7</sup> Anthroponyms and animate common nouns on the other hand often need special, usually longer marking (see Haspelmath 2019:

---

<sup>7</sup> Additionally, it is not uncommon that a language has a set of nouns that is closer to toponyms in their grammatical properties than to other common nouns. Haspelmath (2019: 322) establishes the term *topo-noun* based on the observation that “languages sometimes give special treatment to a diverse set of nouns that denote concepts which are commonly used as spatial landmarks, such as ‘(one’s) house’, ‘village’, ‘school’, ‘church’, ‘beach’”. However, for this study, we included all kinds of inanimate common nouns and thus did not exclude *topo-nouns* as Grounds from our analysis. A special case is the notion of ‘home, (one’s) house’, which is often expressed by non-nominal means

322; Nintemann 2024). Pronouns are also non-prototypical candidates for a Ground as they often behave like anthroponyms and animate common nouns in that they need special marking, and often appear in the form of indexed adpositions (see Nintemann 2024). The second reason is that we assume that Configurations like ON, INSIDE, or IN FRONT OF are most commonly used with inanimate common nouns, while they occur less frequently with toponyms, anthroponyms, animate common nouns, or pronouns due to their referents' semantic properties.

We proceeded exploratively in that we started out solely from the function of the investigated spatial relations and gathered the means of their expression in the examined languages. Our focus rests, however, on the marking at the nominal level. If a language marks Directionality only on the verbal level, it is considered zero-marked. The same goes for Configuration, which can be expressed by motion verbs. This can be seen in Yuwaalaraay in example (6).

(6) Yuwaalaraay [Pama-Nyungan] (Giacon 2014: 58)

INSIDE / Goal

*giirr ganunga / dhaymaa-yi wuu-waa-nha, ngandabaa*  
 true 3PL / ground-ABL **go\_in-MOV-PRS** snake

'They are all **going into** the ground, the snakes.'<sup>8</sup>

These verb-framed languages contrast with satellite-framed languages, where Configuration is expressed by an overt element that includes the nominal Ground (Talmy 1991: 486) like an adposition in Welsh, see example (7).

(7) Welsh [Indo-European] (King 2003: 166)<sup>9</sup>

GL / Place

*Mae dyn yn y stafell aros*  
 be.3SG.PRS man **in** DEF room wait.VN

'There is a man **in** the waiting room.'

Our investigation is based on grammars or grammatical descriptions of spatial relations, complemented by input from language experts and our own language competence. Since obtaining diachronic information for lesser-documented languages is difficult, we keep a solely synchronic perspective on all languages. This still poses a problem as not every grammatical description provides information

---

(Nintemann (in preparation)). For this study, we only include examples in which 'home, (one's) house' is expressed by nominal means.

<sup>8</sup> Some verbs counterintuitively use the ablative for a Goal expression (Giacon 2014: 60–61).

<sup>9</sup> All of the glosses for Welsh are our own.

on the analysability of a given segment. This raises questions such as whether the English preposition *into* consists of one or two morphemes. The same goes for the case system in Hungarian (see Section 4.1), which has fossilised case suffixes that are no longer productive on their own. Synchronically, forms containing these fossilised case suffixes are thus no longer analysable as consisting of two morphemes. In case a source does not provide evidence for the synchronic separability of a form, we count it as monomorphemic.

### 3.2 Data

For our investigation, we compiled a sample of 30 languages (see Appendix 1: Map 1). As a basis for the sampling, we made use of the Genus-Macroarea (GM) method as introduced by Miestamo et al. (2016). The aim was to cover as much of the diversity of the world's languages as possible. The method takes into account two varietal factors: genetic and areal diversity. The areal distribution is categorised into macroareas based on Dryer (1992): Africa, Eurasia, Southeast Asia and Oceania, Australia and New Guinea, North America, and South America. Several factors such as state of documentation, distribution and number of genera per macroarea are taken into account to calculate a selection among the 521 genera mentioned in Dryer (1992) that accounts for the best diversity per macroarea. Miestamo et al. (2016) generate several compilations of language lists that if one would use the top-down method and predetermine the sample size, they would find which combination of genera would be best to choose from.

As our undertaking is a crosslinguistic comparison, there is no necessity to determine which language represents each genus best. In a large-scale sampling such an effect becomes less important. However, the decision which language was chosen per genus cannot be considered without bias. Languages differ greatly in how well they are documented, so our choice always fell on those that have available sources for our research topic.<sup>10</sup> Following Miestamo et al.'s (2016: 250) approach, we selected the languages for each genus as randomly as possible. However, only languages for which our sources provide sufficient data for our research questions could be included. While spatial relations are covered in most grammatical descriptions, a thorough analysis of configurational spatial relations is hard to come by.

---

<sup>10</sup> At this point, we deem it necessary to acknowledge that the majority of our data is based on doculects, i.e. “a linguistic variety as it is documented in a given resource” (Cysouw and Good 2013: 342).

Of the 100 languages we investigated, we were only able to gather data for 30 languages, shown in its macroarea distribution in Table 5.

**Table 5:** Number of languages per Macroarea.

Macroareal	Africa	Eurasia	Southeast Asia and Oceania	Australia and New Guinea	North America	South America	Total
No. of languages	5	7	2	7	5	4	30

Additionally, some genera in a macroarea are (over)represented even though they were not included according to Miestamo et al.'s (2016) calculations, namely one language from each of the Japonic, Yukaghir, Eskimo-Aleut genus, one isolate and two Afro-Asiatic languages instead of one. We also added one sign language to further diversify our sample. Our genealogical classification is based on Glottolog 5.1 (Hammarström et al. 2024).<sup>11</sup>

## 4 The qualitative side of spatial Directionality and Configuration

In this section, we offer a qualitative analysis of our data on the basis of selected examples. We show how case, adpositions, relational nouns, or a combination of these strategies are used to express Directionality and Configuration (4.1–4.6). It is also shown that static and dynamic verbs play a major role in some languages, particularly when it comes to encoding Directionality (4.6–4.8). Furthermore, it will become clear that the presence or absence of overt markers can be influenced by language-specific factors, such as syntactic rules (4.8). Last but not least, we offer an analysis of how sign languages fit into the picture (4.9).

In the following subsections, we discuss the different types of constructions occurring in different languages. For each language, we offer the paradigm in the form of a table.<sup>12</sup> A few points have to be explained to ensure a proper understanding of these tables. X acts as a place-holder for the Ground noun. The position of X corresponds to the position of the Ground noun in a construction. If there is more

<sup>11</sup> An exception was made for German Sign Language as sign languages often lack enough research to be classified sufficiently into language families.

<sup>12</sup> The paradigms of the languages not discussed in this section are offered in Appendix 2: Tables 17–37.

than one possible construction, each construction is represented in a separate line. Optional elements are presented in brackets. Some languages do not employ GL, which is marked with  $\#$  (= non-existent). We use NA for columns where at least one Directionality could not be extracted from our sources. The last row presents the syncretism patterns (Options I–V as depicted in Table 4 above) attested for GL and each configurational construction in the respective language.

## 4.1 Case: Hungarian

In the Uralic language Hungarian, Place, Goal, and Source can be expressed with case affixes as in (8) for the Configurations INSIDE and ON, respectively, while for BEHIND, IN FRONT OF, and UNDER, postpositions are used as in (9). The postpositions expressing BEHIND, IN FRONT OF, and UNDER have a fossilised case suffix that is no longer productive in the nominal system of the language. A final long vowel indicated the lative, *-Vtt* was the ending for the locative, and *-Vl* was used for the ablative in Old Hungarian (cf. Stolz 1990: 345). Thus, as our study is purely synchronic, we do not count e.g. *alatt* ‘under’ as consisting of two morphemes (i.e. *al-att* [under-LOC]) but as only one.

(8) Hungarian [Uralic] (Kenesei et al. 1998: 76)

a. INSIDE / Place

*Anna a ház-ban lakik.*

Anna the house-INE lives

‘Anna lives **in** the house.’

b. INSIDE / Goal

*Anna a ház-ba lépett.*

Anna the house-ILL entered

‘Anna entered **into** the house.’

c. INSIDE / Source

*Anna a ház-ból érkezett.*

Anna the house-ELA arrived

‘Anna came **from** the house.’

(9) Hungarian [Uralic]

a. UNDER / Place (Dékány and Hegedűs 2021: 49)

*a híd alatt*

the bridge **under**

‘**under** the bridge’

- b. UNDER / Goal (Hegedűs and Dékány 2021: 238)  
*egyenesen az ágy alá*  
 straightly the bed **under\_to**  
 ‘straight **under** the bed’
- c. UNDER / Source (Hegedűs and Dékány 2021: 212)  
*a felszín alól érkező hangok*  
 the surface **under\_from** coming sound.PL  
 ‘the sounds coming **from under** the surface’

The  $\emptyset$  in the GL column in Table 6 indicates that this column has to remain empty for Hungarian as “the basic meanings of the three sets of local cases are interior, exterior, and surface, so there is no way of describing general location in Hungarian” (Kenesei et al. 1998: 237). The columns for INSIDE and ON are filled with several allomorphic forms due to vowel harmony.

**Table 6:** Paradigm of Hungarian [Uralic] (Kenesei et al. 1998; Dékány and Hegedűs 2021).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	$\emptyset$	<i>X-ban</i> <i>X-ben</i>	<i>X-n</i> <i>X-on</i> <i>X-en</i> <i>X-ön</i>	<i>X mögött</i>	<i>X előtt</i>	<i>X alatt</i>
Goal	$\emptyset$	<i>X-ba</i> <i>X-be</i>	<i>X-ról</i> <i>X-ről</i>	<i>X mögé</i>	<i>X elé</i>	<i>X alá</i>
Source	$\emptyset$	<i>X-ból</i> <i>X-ből</i>	<i>X-ra</i> <i>X-re</i>	<i>X mögül</i>	<i>X elől</i>	<i>X alól</i>
Option		I	I	I	I	I

As each cell in Table 6 is filled by a different configurational form, Hungarian follows Option I for all Configurations. However, it deviates from the canon in that the forms are synchronically unanalysable and thus encode both Directionality and Configuration (i.e. fused exponence).

## 4.2 Adpositions: Welsh

Welsh, like many Indo-European languages, marks spatial relations using adpositions. Place, Goal, and Source are expressed by three different prepositions, viz. *yn* ‘at’, *i* ‘to’, and *o* ‘from’, see example (10).

## (10) Welsh [Indo-European, Celtic] (King 2003: 32, 288, 291)

## a. GL / Place

*yn yr ysgol*  
 in DEF school  
 ‘at school’

## b. GL / Goal

*Dych chi 'n mynd i 'r dre heddiw?*  
 be.2PL 2PL PRED go.VN to DEF <sup>1</sup>town today  
 ‘Are you going to town today’

## c. GL / Source

*Diolch o galon i ti*  
 thanks from <sup>1</sup>heart to 2SG  
 ‘Thank you from [the bottom of] my heart’

For configurational constructions, Welsh employs a mix of prepositions and spatial nouns, which are also used as adjectives. These constructions are presented in Table 7.

**Table 7:** Paradigm of Welsh [Indo-European, Celtic] (King 2003; Iwan Rees, p.c.).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>yn X</i>	<i>tu fewn i X</i>	<i>ar X</i>	<i>tu ôl i X</i>	<i>o flaen X</i>	<i>(o) dan X</i>
Goal	<i>i X</i>	<i>i mewn i X</i> <i>i fewn i X</i>	<i>ar X</i>	<i>tu ôl i X</i>	<i>i flaen X</i> <i>i ffrynt X</i>	<i>(o) dan X</i>
Source	<i>o X</i>	<i>allan o X</i>	<i>oddiar X</i>	<i>o'r tu ôl i X</i>	<i>o flaen X</i> <i>o ffrynt X</i>	<i>oddidan X</i>
Option	I	I	II	II	I / IV	II

Although in Welsh it is possible to differentiate Place and Goal, this is not done for all configurational constructions, e.g. ON, see example (11). This syncretism causes them to be less canonical than GL, where all three cells are occupied by a different form.

## (11) Welsh [Indo-European, Celtic] (King 2003: 279, 292)

## a. ON / Place

*Mae 'r llyfr ar y bwrdd*  
 be.3SG.PRS DEF book on DEF table  
 ‘The book is on the table’

## b. ON / Goal

*Rho*            *'r*    *llyfr*    *ar*            *y*    *bwrdd*  
 put.2SG.IMP   DEF   book   **on(to)**   DEF   table  
 'Put the book **on(to)** the table'

## c. ON / Source

*Cymer*            *dy*            *bethau*    ***oddiar***    *y*    *bwrdd*,  
 take.2SG.IMP   2SG.POSS   <sup>L</sup>thing.PL   **from.on**   DEF   table  
*nei*                *di?*  
<sup>L</sup>do.2SG.FUT   2SG  
 'Take your things **off** the table, will you?'

The Source relation usually differs from both Place<sup>13</sup> and Goal in that it employs some form of the preposition *o* 'from', see example (11c). The spatial nouns used for the Configurations in question are *tu* 'side', *ôl* 'back', and *blaen* 'front'. *Tu* 'side' may be combined with the preposition *mewn* 'in' or *ôl* 'back' to mean 'inside' or 'backside', respectively.

Overall, Welsh exhibits a mixed pattern in terms of syncretism, as the forms for Place and Goal are identical in some cases and different in others.

### 4.3 Adpositions + case: Udihe

In the Tungusic language Udihe, Place, Goal, and Source are pervasively coded by locative (*-la* ~ *-dulA/-dîlA*)<sup>14</sup>, lative (*-tigi*), and ablative (*-digi*) suffixes, respectively. Udihe thus attests to Option I. These case markers can be attached directly to the Ground noun to express GL or a postposition that encodes Configuration. In the latter case, "[t]he changeable postpositions must be inflected with possessive affixes that refer to the argument of the postposition" (Nikolaeva and Tolskaya 2001: 403), i.e. the Ground noun. In line with Nikolaeva and Tolskaya (2001: 402), the forms presented in Table 8 are given in the 3rd person singular form and thus carry the possessive suffix *-ni*.

<sup>13</sup> One exception is the Configuration IN FRONT OF, where one option for Source (*o flaen X*) coincides with Place.

<sup>14</sup> The capital letter <A> is used to indicate that it is subject to vowel harmony. Which of the two forms is used depends on the noun class, i.e. *-la* is used for class I nouns, while *-dulA/-dîlA* is used for class II nouns (cf. Nikolaeva and Tolskaya 2001: 124).

**Table 8:** Paradigm of Udihe [Tungusic] (Nikolaeva and Tolskaya 2001).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X-la</i> <i>X-dulA/diA</i>	<i>X do-lo-ni</i>	<i>X xo:-lo-ni</i>	<i>X ca:-la-ni</i>	<i>X ñondu-le-ni</i>	<i>X xegie-le-ni</i>
			<i>X xo:n-dule-ni</i>	<i>X aka-la-ni</i>		
			<i>X xo:n-dile-ni</i>	<i>X tau-ze-le-ni</i>		
			<i>X ge:-le-ni</i>			
Goal	<i>X-tigi</i>	<i>X do-tigi-ni</i>	<i>X xo:n-tigi-ni</i>	<i>X ca:-tigi-ni</i>	<i>X ñondu-tigi-ni</i>	<i>X xegie-tigi-ni</i>
			<i>X ge:-tigi-ni</i>	<i>X aka-tigi-ni</i>		
				<i>X tau-ze-tigi-ni</i>		
Source	<i>X-digi</i>	<i>X do-digi-ni</i>	<i>X xo:n-digi-ni</i>	<i>X ca:-digi-ni</i>	<i>X ñondu-digi-ni</i>	<i>X xegie-digi-ni</i>
			<i>X ge:-digi-ni</i>	<i>X aka-digi-ni</i>		
				<i>X tau-ze-digi-ni</i>		
Option	I	I	I	I	I	I

GL in Udihe is completely in line with the canon in that there is exactly one morpheme, in this case a suffix, that expresses Place, Goal, or Source, respectively, and no other morpheme is involved.<sup>15</sup> The use of these suffixes is exemplified in (12).

## (12) Udihe [Tungusic] (Nikolaeva and Tolskaya 2001: 336, 121, 715)

## a. GL / Place

*Min-zurje omo škola-la ñansule-fi.*  
 me-both one school-LOC study-1PL.INCL  
 ‘We study with you **in** the same school.’

<sup>15</sup> There are some instances of a possessive suffix attached to the locative, lative, or ablative suffix, respectively. We assume, however, that these cases are either instances of actual possessive constructions or that the possessive is used to express definiteness, as in Udihe, “[a]rticles are missing, and definiteness may be expressed by means of 3<sup>rd</sup> person possessive affixes” (Nikolaeva and Tolskaya 2001: 479). This is illustrated in (ii).

## (ii) Udihe [Tungusic] (Nikolaeva and Tolskaya 2001: 521)

*Wakca-i ni mo:-digi-ni eu-gi-e-ni.*  
 hunt-PRP man tree-ABL-3SG climb.down-REP-PST-3SG  
 ‘The hunter climbed down [from] **the** tree.’

## b. GL / Goal

*We:-tigi ηene-ze-fi seutigi diga-na-za-fi.*  
 forest-LAT go-SBJV-1PL.INCL nut eat-DEST-SBJV-1PL.INCL  
 ‘Let us go **to** the forest and eat some nuts.’

## c. GL / Source

*Nua-ni ηele-ini teηku-digi tiηme-le-zeηe-zi.*  
 he-3SG be.afraid-3SG chair-ABL fall-SING-FP-INST.SS  
 ‘He is afraid to fall **from** the chair.’

The configurational forms, however, follow different rules. They follow the canon in that there is exactly one morpheme that expresses Configuration and exactly one morpheme that specifies Directionality. In contrast to the canonical construction as depicted in Table 1 in Section 2 above, however, an additional element is needed, viz. the possessive suffix that refers back to the Ground noun, see (13).

## (13) Udihe [Tungusic] (Nikolaeva and Tolskaya 2001: 885, 517, 521)

## a. ON / Place

*Jeu bi:-ni uti montom-ziga-i uti t'a xo:n-dile-ni?*  
 what be-3SG that circle-PL-FOC that tree **top-Loc-3SG**  
 ‘Why are circles there **on** the fallen tree?’

## b. ON / Goal

*Čeradaka xo:n-tigi-ni tukti:-ni.*  
 attic **top-LAT-3SG** climb-SG  
 ‘He is climbing **up to** the attic.’

## c. ON / Source

*Čeradaka xo:n-digi-ni eu-gi:-ni.*  
 attic **top-ABL-3SG** go.down-REP-3SG  
 ‘He is going **down from** the attic.’

The use of postpositions in Place (13a), Goal (13b), and Source (13c) constructions is exemplified with *xo:n-* ‘over, above, on’. As Nikolaeva and Tolskaya (2001: 402) state, “[t]he declension of other postpositions is completely analogous”. Overall, configurational constructions in Udihe are more complex than GL. The three spatial relations, however, show the same complexity – at least when it comes to the number of morphemes.

#### 4.4 Adpositions + relational nouns: Rapanui

In Rapanui, Directionality is marked by the prepositions *i*/<sup>16</sup>*i* ‘in, at, on’ for Place, *ki* ‘to’ for Goal, and *mai* ‘from’ for Source throughout the whole paradigm provided in Table 9. Example (14) illustrates the use of these prepositions to convey GL. The configurational forms on the other hand are more complex and consist of an initial preposition followed by a relational noun (referred to as locational by Kievit 2017: 121) and a second preposition. This second preposition “does not have any semantic contribution; it serves just to provide a syntactic link between the locational and its complement [ . . . ] the second preposition may be either *i* [ . . . ], *o* [ . . . ], or a copy of the first preposition [ . . . ]” (Kievit 2017: 124). In Table 9, only one form is listed, in this case a copy of the first preposition.

**Table 9:** Paradigm of Rapanui [Austronesian, Oceanic] (Kievit 2017; Du Feu 1996<sup>17</sup>).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>i</i> X	<i>i roto i</i> X	<i>i/a ruŋa i</i> X	<i>i tu'a o</i> X	<i>i mu'a i</i> X	<i>i raro i</i> X
Goal	<i>ki</i> X	<i>ki roto ki</i> X	<i>ki ruŋa ki</i> X	<i>ki tu'a ki</i> X	<i>ki mu'a ki</i> X	<i>ki raro ki</i> X
Source	<i>mai</i> X	<i>mai roto mai</i> X	<i>mai ruŋa mai</i> X	<i>mai tu'a mai</i> X	<i>mai mu'a mai</i> X	<i>mai raro mai</i> X
Option	I	I	I	I	I	I

(14) Rapanui [Austronesian, Oceanic] (Kievit 2017: 210, 213, 95)

a. GL / Place

*He noho i te hare o re huŋavai.*  
 NTR stay **at** ART house of ART parent\_in\_law  
 ‘She stayed **in** the house of her in-laws.’

b. GL / Goal

*E tahi mahana he turu a Tiare ki te hare hāpī.*  
 NUM one day NTR go\_down PROP Tiare **to** ART  
 house learn  
 ‘One day Tiare went down **to** school.’

<sup>16</sup> “In the accepted Rapa Nui orthography [ . . . ], this preposition is written either *i* or *i*, depending on its function [ . . . ] the preposition in a locative sense is written *i*, while the preposition occurring after locationals is *i*” (Kievit 2017: 210).

<sup>17</sup> In contrast to Kievit (2017), Du Feu (1996) writes the first preposition and the locational as one word, e.g. *kiroto ki te hare into* the house’. We adhere to the conventions used in Kievit (2017) here and used data provided in Du Feu (1996) only to complete the paradigm in Table 9.

## c. GL / Source

*Mai tai nei, mai te hopu ija mātou ko*  
**from** sea PROX **from** ART bathe NMLZ 1PL.EXCL PROM  
*kuā Tonere.*

COLL Tonere

‘We are coming **from** the shore, **from** swimming with Tonere.’

The examples in (15) illustrate this pattern for INSIDE. In (15a) and (15c), a copy of the first preposition, *i* or *mai*, respectively, is used, while in (15b), *i* is used as the second preposition.

## (15) Rapanui [Austronesian, Oceanic] (Kievit 2017: 121, 124)

## a. INSIDE / Place

*A mua ‘i roto i te hare.*  
 PROP Mum **at inside at** ART house

‘Mother is **in** the house.’

## b. INSIDE / Goal

*He uru te kurī ki roto i te hare.*  
 NTR enter ART cat **to inside at** ART house

‘The cat entered **into** (lit. **to inside**) the house.’

## c. INSIDE / Source

*He e’a mai roto mai te koro.*  
 NTR go\_out **from inside from** ART feast\_house

‘They went **out of** the feast house.’

Like Hungarian and Udihe, Rapanui follows Option I for all Configurations. However, Directionality is expressed by two morphemes, e.g. *ki . . . ki* ‘to . . . to’ and thus does not entirely follow the canon.

## 4.5 Case + relational nouns: Kalaallisut

West Greenlandic, or Kalaallisut, attests to the second-highest average number of morphemes in our sample, shortly after German Sign Language (see Section 4.9). This is clearly because of its general language type. Like all Eskimo-Aleut languages, Kalaallisut is highly polysynthetic and agglutinative so that nouns can carry numerous affixes<sup>18</sup> (Holst 2005: 55). In combination with the use of relational nouns for spatial relations, this leads to the employment of several morphemes. The basic

<sup>18</sup> In fact, Kalaallisut has almost exclusively suffixes, there is only one real prefix.

pattern is to put the Ground in the ergative case, employ the relational noun, refer back to the Ground by a possessive suffix for the third person, and then mark Directionality by case (Lybach 2022: 251). These cases by themselves also mark GL in Place, Source, and Goal constructions.

(16) Kalaallisut<sup>19</sup> [Eskimo-Aleut] (Nielsen 2019: 247–249)

## a. GL / Place

*Illu-mi ino-qa-nngila-q.*  
house-LOC human-there\_is-NEG-3SG.IND  
'There is no one **in** the house.'

## b. GL / Goal

*Pisiniarfim-mut iser-poq.*  
shop-ALL go\_in-3SG.IND  
'He went **into** the shop.'

## c. GL / Source

*Illu-miit ani-voq.*  
house-ABL come\_out-3SG.IND  
'He came **out** of the house.'

Kalaallisut possesses several relational nouns that denote not only various configurational but also coordinate relations, along with expressions of distant or close surroundings. For all investigated relations, we found corresponding relational nouns, as presented in Table 10 below.

**Table 10:** Paradigm of Kalaallisut [Eskimo-Aleut] (Lybach 2022).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	X-{ <i>m</i> }	X-{ <i>p</i> } { <i>ilu</i> } { <i>a</i> }{ <i>ni</i> }	X-{ <i>p</i> } { <i>qaa</i> } { <i>a</i> }{ <i>ni</i> }	X-{ <i>p</i> } { <i>tunu</i> } { <i>a</i> }{ <i>ni</i> }	X-{ <i>p</i> } { <i>saaq</i> }{ <i>a</i> } { <i>ni</i> }	X-{ <i>p</i> } { <i>atə</i> }{ <i>a</i> }{ <i>ni</i> }
Goal	X-{ <i>mut</i> }	X-{ <i>p</i> } { <i>ilu</i> } { <i>a</i> }{ <i>nut</i> }	X-{ <i>p</i> } { <i>qaa</i> } { <i>a</i> }{ <i>nut</i> }	X-{ <i>p</i> } { <i>tunu</i> } { <i>a</i> }{ <i>nut</i> }	X-{ <i>p</i> } { <i>saaq</i> }{ <i>a</i> } { <i>nut</i> }	X-{ <i>p</i> } { <i>atə</i> }{ <i>a</i> }{ <i>nut</i> }
Source	X-{ <i>mət</i> }	X-{ <i>p</i> } { <i>ilu</i> } { <i>a</i> }{ <i>nət</i> }	X-{ <i>p</i> } { <i>qaa</i> } { <i>a</i> }{ <i>nət</i> }	X-{ <i>p</i> } { <i>tunu</i> } { <i>a</i> }{ <i>nət</i> }	X-{ <i>p</i> } { <i>saaq</i> }{ <i>a</i> } { <i>nət</i> }	X-{ <i>p</i> } { <i>atə</i> }{ <i>a</i> }{ <i>nət</i> }
Option	I	I	I	I	I	I

<sup>19</sup> The morphemic dissemination is kept rudimentary to easily illustrate the GL marking here.

The examples in (17) and (18), are a comparison of the different uses for BEHIND and IN FRONT OF.

(17) Kalaallisut<sup>20</sup> [Eskimo-Aleut] (Stian Lybach, p.c.)

a. BEHIND / Place

*Qitsuk illup tunuaniippoq.*  
 {qitsuk} {illu}{p} {tunu}{-a}{ni}{ət}{vu}{q}  
 cat.NOM house.ERG **backside-POSS.3SG/SG-LOC-be\_located-IND-3SG**  
 ‘The cat was **behind** the house.’

b. BEHIND / Goal

*Qitsuk illup tunuanukarpoq.*  
 {qitsuk} {illu}{p} {tunu}{-a}{nut}{-kaq}{vu}{q}  
 cat.NOM house.ERG **backside-POSS.3SG/SG-ALL-move-IND-3SG**  
 ‘The cat went **behind** the house.’

c. BEHIND / Source

*Qitsuk illup tunuani(it) takkuppoq.*  
 {qitsuk} {illu}{p} {tunu}{-a}{nət} {takkut(ə)}{vu}{q}  
 cat.NOM house.ERG **backside-POSS.3SG/SG-ABL** appear-IND-3SG  
 ‘The cat appeared **from behind** the house.’

As illustrated in (17), the Ground is marked with the ergative case suffix {-p}, the relational noun {tunu} is then introduced, and can even be accompanied by another verbal suffix as in (17a) and (17b). In example (18), the same structure is seen with the relational noun {saaq} for IN FRONT OF.

(18) Kalaallisut [Eskimo-Aleut] (Stian Lybach, p.c.)

a. IN FRONT OF / Place

*Qitsuk illup saavaniippoq.*  
 {qitsuk} {illu}{p} {saaq}{-a}{ni}{ət}{vu}{q}  
 cat.NOM house.ERG **front-POSS.3SG/SG-be\_located-IND-3SG**  
 ‘The cat is **in front of** the house.’

<sup>20</sup> Since Greenlandic possesses numerous sound laws that truncate, add, assimilate, or change sounds and the orthography only shows the end result of these processes, we added the underlying form of each morpheme in curved brackets to better illustrate the structure of each word, a hyphen in front of a morpheme marks a morpheme that truncates the preceding consonant. For more information on sound rules in Greenlandic see Lybach (2022: Chapter 2).

## b. IN FRONT OF / Goal

*Qitsuk illup saavanukarpoq.*  
 {qitsuk} {illu}{p} {saaq}{-a}{nut}{-kaq}{vu}{q}  
 cat.NOM house.ERG front-POSS.3SG/SG-ALL-move-IND-3SG  
 ‘The cat went **in front of** the house.’

## c. IN FRONT OF / Source

*Qitsuk illup saavani(it) illup*  
 {qitsuk} {illu}{p} {saaq}{-a}{nət} {illu}{p}  
 cat.NOM house.ERG front-POSS.3SG/SG-ABL house.ERG  
*tunuanut ingerlavoq.*  
 {tunu}{-a}{nut} {ingerla}{vu}{q}  
 backside-POSS.3SG-ALL go-IND-3SG  
 ‘The cat went **from in front of** the house **to behind** the house.’

Of all the languages in our sample that employ relational nouns, Kalaallisut seems to have the most complex embedding, i.e. the highest number of morphemes within the expression, with four on average. As can be seen from Table 10, Kalaallisut shows Option I for all spatial relations, general or configurational. GL is solely marked by case, while all configurational expressions combine relational nouns with the respective directional case suffix. There is no syncretism in the expression of spatial relations in Kalaallisut.

## 4.6 Adpositions + verbs: Maltese

Maltese uses adpositions, more specifically prepositions, to express configurational constructions. For Place and Goal constructions expressing GL, zero-marking is used for “a place which is fairly predictable from the context” (Borg and Azzopardi-Alexander 1997: 155).<sup>21</sup> However, Source is always expressed overtly with the preposition *minn* ‘from’ as in (19c).

<sup>21</sup> The zero-marked expressions in (19a–b), ‘church’ and ‘school’, are prototypical candidates for topo-nouns, which are generally more prone to zero-marking. For a more thorough discussion of zero-marking of spatial relations in Maltese we refer the reader to Borg (1988), Stolz et al. (2014: Ch. 4.2) and Stolz and Vorholt (in preparation, Ch. 4.3.3.2 and 5).

(19) Maltese [Afro-Asiatic, Semitic] (Borg and Azzopardi-Alexander 1997: 49, 156)<sup>22</sup>

a. GL / Place

*Ommi* (qiegħda) Ø *l-knisja*  
 mother:3SG located:3SG.F **at** DEF-church  
 ‘His mother is **at** church.’

b. GL / Goal

*Ġrew* Ø *l-iskola*  
 run.3PL.PFV **to** DEF-school  
 ‘They ran **to** school.’

c. GL / Source

*Tlaqna* *mil*<sup>23</sup>-*ajruport* *fil-ħin*  
 leave:1PL.PFV **from**:DEF-airport in:DEF-time  
 ‘We left [**from**] the airport on time.’

Goal constructions are either only marked by the verb, or by use of a second preposition as indicated by the parenthesis in Table 11 below. The two options are shown in (20b) and (20c).

(20) Maltese [Afro-Asiatic, Semitic]

a. UNDER / Place (Borg and Azzopardi-Alexander 1997: 161)

*It-tfal* *iħobbu* *jistaħbew* **taħt** *il-mejda*  
 DEF-children 3.IPFV:love:PL 3.IPFV:hide:PL **under** DEF-table  
 ‘Children love to hide **under** the table.’

b. UNDER / Goal (Borg and Azzopardi-Alexander 1997: 162)

*Il-qarnita* *nizlet* *għal taħt* *il-blata* *biex*  
 DEF-octopus go\_down:3SG.F.PFV **for** **under** DEF-rock to  
*tistaħba*  
 3SG.F.IPFV:hide  
 ‘The octopus went **under** the rock in order to hide.’

c. UNDER / Goal [Korpus Malti 3.0, news79496]<sup>24</sup>

*Mort* **taħt** *it-tieqa* *tiegħu*  
 go:1SG.PFV **under** DEF-window of:3SG.M  
 ‘I went **under** his window [. . .].’

<sup>22</sup> We added the Ø here for better comprehensibility of the a. and b. example.

<sup>23</sup> The preposition *minn* ‘from’ fuses with the definite article (*il-*) to *mill-* ‘from the’.

<sup>24</sup> The glosses and translation are our own.

- d. UNDER / Source [Korpus Malti 3.0, literature13]<sup>25</sup>  
*Frighthom minn taħt l-iskipp*  
 take\_out:1SG.PFV:3PL.DO **from under** DEF-skip  
 ‘I took them out **from under** the skip.’

**Table 11:** Paradigm of Maltese [Afro-Asiatic, Semitic] (Borg & Azzopardi-Alexander 1997).<sup>26</sup>

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	∅ X	<i>fi</i> X <i>ġo</i> X <i>ġewwa</i> X	<i>fuq</i> X	<i>wara</i> X	<i>quddiem</i> X	<i>taħt</i> X
Goal	∅ X	<i>fi</i> X <i>ġo</i> X <i>ġewwa</i> X	( <i>għal</i> ) <i>fuq</i> X	( <i>għal</i> ) <i>wara</i> X	<i>quddiem</i> X <sup>27</sup>	( <i>għal</i> ) <i>taħt</i> X
Source	<i>minn</i> X	<i>minn</i> <i>ġo</i> X <i>minn</i> <i>ġewwa</i> X	<i>minn</i> <i>fuq</i> X	<i>minn</i> <i>wara</i> X	<i>minn</i> <i>quddiem</i> X	<i>minn</i> <i>taħt</i> X
Option	II	II	I / II	I / II	II	I / II

Source constructions, and in some cases Goal constructions as well, use a periphrastic form while Place is always indicated by a monolectic form in Maltese. Consequently, the Maltese paradigm is not in line with the canon.

## 4.7 Directionality via verbs: Wari’

Wari’ expresses Directionality exclusively with verbs. This strategy is an option in many languages (e.g. in English: *I entered the house* (Goal), *I left the house* (Source)), however, in Wari’ it is the only possibility to express Directionality. This leads to

<sup>25</sup> The glosses and translation are our own.

<sup>26</sup> It could be argued that *fi* ‘in’ could also be added to the GL column of the paradigm in Table 11. However, we chose to only include the forms provided in Borg and Azzopardi-Alexander (1997: 155–162) in this study. There is still ongoing research with regard to zero vs overt marking of Place and Goal, and the competition between zero, *fi*, *ġo*, and *ġewwa*, especially with regard to toponyms (cf. Stolz et al. 2014, 2017c, 2023), which we chose to exclude from this study (cf. Section 3). An in-depth analysis of configurational spatial relations in Maltese that is currently undertaken by Stolz and Vorholt (in preparation) reveals a more intricate paradigm.

<sup>27</sup> The possibility of *għal* ‘for’ and *quddiem* ‘in front of’ is not mentioned by Borg and Azzopardi-Alexander (1997). Stolz and Vorholt (in preparation, Ch. 5.1.2.2.10.1.2) discuss this case specifically and find only three genuine instances in the consulted text corpus.

the zero-marking of Directionality in nominal phrases, ultimately resulting in the syncretism of Place, Goal, and Source (= Option V), as displayed in Table 12.

**Table 12:** Paradigm of Wari' [Chapacuran] (Everett and Kern 1997).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	∅ X	<i>tequipa-in X</i>	NA	<i>wara-in X</i>	<i>quima-in X</i>	<i>xone-in X</i>
Goal	∅ X	<i>tequipa-in X</i>		<i>wara-in X</i>	<i>quima-in X</i>	<i>xone-in X</i>
Source	∅ X	<i>tequipa-in X</i>		<i>wara-in X</i>	<i>quima-in X</i>	<i>xone-in X</i>
Option	V	V		V	V	V

Wari' differentiates four types of movement or states: 'at rest', 'motion to', 'motion from', and 'motion past' (Everett and Kern 1997: 246). The verb *pe/to* 'to be at' is employed for 'at rest' location (see (21a) and (22a)), while *mao/mama* 'to go' expresses 'motion to' (see (21b) and (22b)). The Source relation requires a post-verbal modifier, e.g. *qui* 'coming this way' (see (21c)) or *ca* 'this neuter' (see (22c)).

(21) Wari' [Chapacuran] (Everett and Kern 1997: 247–248)

a. GL / Place

*Pe na-in ∅ xirim.*  
 be:at:SG 3SG:R.PST/PRS-3N ∅ house  
 'He is **at**, or **in**, the house.'

b. GL / Goal

*Mao na-in ∅ xirim*  
 go:SG 3SG:R.PST/PRS-3N ∅ house  
 'He went **to** the house'

c. GL / Source

*Tan' qui' nana-in ∅ xiri-nain pic*  
 arrive:PL coming:this:way 3PL:R.PST/PRS-3N ∅ house-3N rubber  
 They arrived **from** the rubber camp (lit. '... rubber's house')

As displayed in Table 12, Configuration is expressed through spatial nouns combined with a possessive pronoun referring to the Ground (see (22)). The nouns signify body parts, e.g. *tequipa* 'thorax', *wara* 'back', *quima* 'chest', and *xone* 'curve of back'. This is exemplified in (22) with INSIDE.

- (22) Wari' [Chapacuran] (Everett and Kern 1997: 253–254)
- a. INSIDE / Place  
*(Corom) to nana-in tequipa-in xirim.*  
 enter be:at:SG 3PL:R.PST/PRS-3N **thorax-3N** house  
 They are **inside** the house.' (lit. '...**in** the house's **thorax**.')
- b. INSIDE / Goal  
*Corom mama' nana-in tequipa-in xirim 'oro wari'.*  
 enter go:PL 3PL:R.PST/PRS-3N **thorax-3N** house COLL person  
 'The people entered going **into** the house.' (lit. '...**into** the house's **thorax**.')
- c. INSIDE / Source  
*Hwet ca' na-in tequipa-in xirim.*  
 appear:SG this:N 3SG:R.PST/PRS-3N **thorax-3N** house  
 'He came **out of** the house.' (lit. '...**out of** the house's **thorax**.')

Examples (22a) and (22b) illustrate how the verbs *to* 'be at (sg)' and *mama'* 'go (PL)' may be combined with an additional verb (*corom* 'enter') to convey Directionality. The verbs described above are just a small sample of the many verbs and combinations that Wari' employs to express Location and Direction.

## 4.8 Syntactic influence: Mandarin Chinese

Several language-specific factors may influence a spatial construction. In Mandarin Chinese, for example, both Place and Goal constructions may be either zero-marked or take a preposition. This is true for both GL and the configurational constructions as Table 13 demonstrates.

**Table 13:** Paradigm of Mandarin Chinese [Sino-Tibetan, Sinitic]<sup>28</sup> (Ross and Ma 2006; Jingting Ye, p.c.).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	(zài) X (nàr)	(zài) X (de) lītou (zài) X lǐ	(zài) X (de) shàngtōu (zài) X shàng	(zài) X (de) xiàtōu	(zài) X (de) qiántōu	(zài) X (de) xiàtōu (zài) X xià

<sup>28</sup> Ross and Ma (2006: 291) explain that "Mandarin location words consist of a base form and a location suffix". They list the location words with three different location suffixes, viz. *-tōu*, *-miàn*, and *biān*. For reasons of clarity, we decided to present only one of the forms in Table 13. Furthermore, not all of the location words can follow the Ground noun in their base form without a location suffix. According to Ross and Ma (2006: 293), this is only possible with *lǐ* 'inside', *shàng* 'above', and *xià* 'below' – as well as *wài* 'outside' which we do not consider in this study. In these cases, *de* 'of' does not occur, while it is optional in combination with the forms carrying a location suffix (Ross and Ma 2006: 292).

Table 13 (continued)

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Goal	(dào) X (nàr)	(dào) X (de) lǐtōu (dào) X (de) lǐ	(dào) X (de) shàngtōu (dào) X shàng	(dào) X (de) xiàtōu	(dào) X (de) qiántōu	(dào) X (de) xiàtōu (dào) X xià
Source	cóng X (nàr)	cóng X (de) lǐtōu cóng X lǐ	cóng X (de) shàngtōu cóng X shàng	cóng X (de) xiàtōu	cóng X (de) qiántōu	cóng X (de) xiàtōu cóng X xià
Option	I / II	I / II	I / II	I / II	I / II	I / II

As the brackets in Table 13 indicate, the prepositions *zài* ‘at, in, on’ and *dào* ‘to’ are not obligatory in every case. However, different factors determine the absence or presence of the preposition in these two relations. *Zài* can act as a preposition ‘at’ or a verb with the meaning ‘exist, be located at’ (cf. Ross and Ma 2006: 87–88).<sup>29</sup> If followed by an object noun phrase and a verb phrase, it is analysed as a preposition (23a). Without another verb phrase, however, it functions as a verb (23b).

(23) Mandarin Chinese [Sino-Tibetan, Sinitic] (Ross and Ma 2006: 88)<sup>30</sup>

- a. GL / Place with *zài* as a verb  
*Tā zài jiā chī fàn.*  
 he **at** home eat meal  
 ‘He is eating **at** home.’
- b. GL / Place with *zài* as a preposition  
*Tā zài jiā.*  
 he **exist** home  
 ‘He **is at** home.’

If we follow the consulted grammar by Ross and Ma (2006) and consider *zài* a verb in (23b), the Place construction is zero-marked. In (23a), however, *zài* is used as a preposition that marks the Ground *jiā* for the relation of Place. It is important to note that not all Grounds can be marked with only *zài*. *Jiā* ‘home’ is one of the few common nouns that may directly be marked with *zài* (and other locational prepositions). In Mandarin Chinese, there is a group of place nouns that differ from other nouns in that they may follow these prepositions without any additional

<sup>29</sup> The form *zài* is commonly analysed as a preposition that also functions as a verb (e.g. Ross and Ma 2006: 87–88). Alternatively, some authors suggest that this element should be analysed as a coverb (e.g. Li and Thompson 1989: 366; Po-Ching and Rimmington 2006: 152).

<sup>30</sup> The glosses are our own.

element (cf. Chao 2011: 532–544). This group consists of toponyms, position words like the configurational forms listed in Table 13 above – which may also function as a Ground by themselves – but also words like *zhèr* ‘here’ or *nàr* ‘there’, and a few common nouns like *jiā* ‘home’, *xuéxiào* ‘school’, or *fēijī chǎng* ‘airport’<sup>31</sup>. Other nouns have to be used with position words, e.g. *fángzi lǐ* ‘house in’ or *fángzi nàr* ‘house there’. Chao (2011: 543) defines these constructions as compounds which then form place nouns.

For Goal constructions, the preposition *dào* ‘to’<sup>32</sup> is used under certain conditions, i.e. if a verb phrase follows the Ground (24a). However, the verb may also precede the Ground, in which case the preposition is not used (24b).

(24) Mandarin Chinese [Sino-Tibetan, Sinitic] (Ross and Ma 2006: 303)<sup>33</sup>

- a. GL / Goal with preposition *dào*  
*Wǒ xiǎng dào túshūguǎn qù.*  
 I want to library go  
 ‘I want to **go to** the library.’
- b. GL / Goal without preposition *dào*  
*Wǒ xiǎng qù túshūguǎn.*  
 I want go library  
 ‘I want to **go to** the library.’

Depending on the syntactic properties, viz. the motion verb preceding or following the Ground, the Goal construction is either zero-marked or overtly marked with the preposition *dào* ‘to’. Similar rules apply when configurational forms are used, so that a variety of constructions is possible in Mandarin Chinese.

## 4.9 Spatial relations in sign languages: German Sign Language

Our sample also contains one sign language, viz. German Sign Language (DGS, ‘Deutsche Gebärdensprache’). The features it shows regarding our research question are comparable to many other sign languages. The principle of universalism dictates that all human languages, whether they use the oral-auditive or visual-manual modality, are typologically comparable. As Behrens (2024) has exemplarily shown, sign languages can for instance partake in the same language

<sup>31</sup> These place nouns thus correspond to Haspelmath’s (2019) notion of topo-noun (see fn. 7).

<sup>32</sup> Similar to *zài* ‘at’ or ‘exist, be located at’, *dào* ‘to’ can also be used as a verb with the meaning ‘to arrive’.

<sup>33</sup> The glosses are our own.

contact dynamics (e.g. *sprachbunds*) as spoken languages. The most prominent and very central feature of sign languages for this research question is the grammaticalisation of the sign space, i.e. the imaginary space in front of the signer's body that is reachable by the hands and arms. Virtually all points within this space can be served in an utterance, while economy results in certain areas on a level plane being more grammaticalised than others. The expressions under scrutiny contain several morphemes that can be referred to as Indices. Their free and bound forms place a Figure, Ground, or other parts of speech at a certain place in the signing space. Their indexical values have to agree anaphorically (Sandler and Lillo-Martin 2006: 25–26).

Another pivotal (but not exclusive) feature of sign languages are classifiers. The two main forms are substitutors, where a classifier hand form substitutes an entity in an expression, and manipulators, where the hand form represents how an entity is semantically handled (Eichmann et al. 2012: 94–96). Not only do they play an important role in verbal contexts but also in spatial expressions. The use of classifiers for spatial expressions instead of adpositions seems to be common among sign languages as a typological study by Eberle (2013) for five sign languages<sup>34</sup> suggests. While all sign languages show a great variety in the expression of spatial relations, classifiers are the most dominant method. The use of adpositions, however, is rare and often associated with influence from an oral language with late-deafened informants (Eberle 2013: 54–55).

In what follows, a general, possible template for the expression of a configurational spatial relation is given. The Figure is normally signed first. After this, the Ground is introduced, optionally by an indexical morpheme, i.e. signed on the side to embed it in the signing space. Hereafter follows the substitute classifier, marked with the same Index as the Ground. Lastly, a sign is expressed that consists of one morpheme for each the anaphorical Index, Directionality, and Configuration.<sup>35</sup>

Since sign languages have the facility to express morphemes simultaneously, indicated by square brackets, not all of them are in a sequential order.

---

<sup>34</sup> She investigates Catalan Sign Language, Estonian Sign Language, Nigerian Sign Language, Thai Sign Language, and Austrian Sign Language.

<sup>35</sup> To not exceed the scope of this paper, we have chosen a context-free pattern for this language. Since sign languages can make very productive use of iconicity, the exact morphological structure can differ vastly between different semantic contexts. It is not possible to take account of them all and calculate an average morpheme length for it.

- (25) German Sign Language (DGS)<sup>36</sup> [German Sign Language family]<sup>37</sup>  
 INSIDE / Place  
 BEAR POSS.3b CAVE-left [**SUBS:opening-left.a; INDEX-left.b**] LIVE INDEX-left.b  
 ‘The bear lives **inside** his cave.’

The glossing is partly simplified in order to represent the pattern as clearly as possible. What evokes the meaning of INSIDE in example (25) is the simultaneous expression of the substitutor classifier and the Index sign and their relation to each other. The orientation of the classifier sign (here described as positional -left.a, closer to the body) and the placement of the ending position of the Index (here as -left.b, further away from the body) gives the meaning of INSIDE as the indexical morpheme of the Index is within the iconic opening demonstrated by the substitutor. The optional doubling at the end of the sentence of indexical signs is used as reinforcement. It may be used here to underline the meaning of INSIDE but it is not generally obligatory (Quer et al. 2019: 400–402).

- (26) German Sign Language (DGS) [German Sign Language family] (own competence)
- a. INSIDE / Goal  
 WATER [**SUBS:surface-left; left.high-DIR.ALL-left.low**]  
 ‘**into** the water’
  - b. INSIDE / Source  
 WATER [**SUBS:surface-left; left.low-DIR.ABL-left.high**]  
 ‘**from out of** the water’

As to Directionality, DGS uses two locational morphemes on the Index sign when it encodes a direction, the starting and ending place of the sign path. When it has a stationary meaning, as in example (25), there is only one morpheme, the ending position of the Index sign. When two locational markers are present, the Index sign produces a path, this is why the Index is glossed here as DIR.ALL for Goal and DIR.ABL for Source.<sup>38</sup> The opposite meaning of (26a) is glossed in (26b) and would be conveyed by reversing the ending and starting locational morphemes on the Index and turning the pointing direction of the Index around, switching from DIR.ALL to DIR.ABL, to align with the resulting path.

<sup>36</sup> All glosses for DGS are our own.

<sup>37</sup> <https://media.spreadthesign.com/video/mp4/9/164048.mp4> (accessed 06.11.2024)

<sup>38</sup> See Sandler and Lillo-Martin (2009: Chapter 9) for more information on sign language phonology.

- (27) German Sign Language (DGS) [German Sign Language family] (own competence)  
 BEHIND / Source  
 CORNER [**SUBS:edge-left; left.front-INDEX-left.back**]  
 ‘**from behind** the corner’

In (27), another example is shown with the use of a classifier to express FROM + BEHIND. While it illustrates well the theoretical marking of this configurational spatial relation, it also shows how much the glossing leaves to imagination with sign languages. The Index sign has two locational morphemes and shows a path between these too. However, there can be virtually indefinite realisations of this. The path can show how a Figure is walking very closely to the corner, or the distance between the two locations can be very big, showing that the Figure is going a long way.

However, often enough, Directionality and Configuration are marked on the verb itself. There is a subclass of special spatial verbs in sign languages that have the locational Index marker on the verb, like STAY, GO, COME etc. With the placement of this Index in relation to the Ground sign or classifier sign, Configuration can be marked on the verb, too. (28a) illustrates how a verb, containing a manipulator classifier, encodes both Configuration and Directionality within the verb by having the starting and ending locations within the verb PUT. In example (28b), the directional verb LOOK, which can carry indexical markers for who is looking and what is looked at, encodes TO + BEHIND by a right sided lean of the body to enable an iconically free viewing path to what is behind the tree.

- (28) German Sign Language (DGS) [German Sign Language family] (own competence)
- a. UNDER / Goal  
 1SG TABLE [**SUBS:surface-left;**  
**BEER left.middle-MANIP:container-PUT-left.bottom**]  
 ‘I put a glass of beer **under the table.**’
- b. BEHIND / Goal  
 \_\_\_\_\_ leaning.right  
 TREE [**SUBS:upright.object-left;** 1SG-LOOK-**left.front**]  
 ‘I look **behind** the tree.’

As illustrated, the grammaticalisation of the signing space, combined with classifiers, multi-purpose indexical morphemes, and iconicity can lead to a plethora of different expressions of our investigated configurational spatial relations. In the absence of a spatial or directional verb, the spatial relations can be overtly marked on the Ground. Often enough, however, the Ground is zero-marked whenever the locational markers can be outsourced to the verbal domain.

Of all the investigated languages, DGS is the one with the highest number of morphemes on average. This can have many reasons. Prominently, the signing space might require more exact marking as it contains far more possible parameters than oral languages. Given that sign languages can articulate morphemes simultaneously, this might not lead to a longer duration of the expression. At least, effects of language economy do not seem to have shortened the constructions so far. As mentioned above, we have chosen the pattern with a classifier as the basic context-free one for our investigation, acknowledging that DGS has many more options for the expression of configurational spatial relations, arguably even structures without classifiers like [HOUSE-left left.middle-DIR.ALL-left.side] ‘into the house’, which would still contain four morphemes, however. Eberle (2013: 54–56) also reports that using lexical items only and placing them in relation to each other in the sign space (e.g. by using the dominant and non-dominant hand simultaneously) is a common alternative to the use of classifiers. However, neither construction in DGS can be considered canonical as defined above, since they all show a higher number of morphemes than there are functions (Directionality and Configuration).

**Table 14:** Paradigm of configurational spatial relations in DGS (own competence).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	X(-IX) (INDEX-IX)	X(-IX) [CLASS:___-IX; (INDEX-IX)]	X(-IX) [CLASS:___-IX; (INDEX-IX)]	X(-IX) [CLASS:___-IX; (INDEX-IX)]	X(-IX) [CLASS:___-IX; (INDEX-IX)]	X(-IX) [CLASS:___-IX; (INDEX-IX)]
Goal	X(-IX) (DIR. ALL-IX)	X(-IX) [CLASS:___-IX; (IX-DIR.ALL-IX)]	X(-IX) [CLASS:___-IX; (IX-DIR.ALL-IX)]	X(-IX) [CLASS:___-IX; (IX-DIR.ALL-IX)]	X(-IX) [CLASS:___-IX; (IX-DIR.ALL-IX)]	X(-IX) [CLASS:___-IX; (IX-DIR.ALL-IX)]
Source	X(-IX) (IX-DIR.ABL)	X(-IX) [CLASS:___-IX; (IX-DIR.ABL-IX)]	X(-IX) [CLASS:___-IX; (IX-DIR.ABL-IX)]	X(-IX) [CLASS:___-IX; (IX-DIR.ABL-IX)]	X(-IX) [CLASS:___-IX; (IX-DIR.ABL-IX)]	X(-IX) [CLASS:___-IX; (IX-DIR.ABL-IX)]
Option	I / V	I / V	I / V	I / V	I / V	I / V

Table 14 shows the syncretism pattern for the mentioned classifier construction in DGS. GL as well as all configurational expressions show no syncretism for Directionality. The opposite is true for when the Configuration and/or Directionality is marked on the verb as mentioned above. Then, Option V emerges as all expressions are marked the same, viz. zero.

## 5 The quantitative side of spatial Directionality and Configuration

In this section, we analyse and discuss quantitative aspects of our study. In Section 5.1, we measure the morphological complexity of the constructions under scrutiny to test our Hypotheses 1 and 2. Hypothesis 3 will be addressed in Section 5.2 which offers an analysis of the occurring syncretism patterns.

### 5.1 Morphological complexity

For our quantitative analysis, we focus on the number of morphemes to assess the morphological complexity of the constructions under scrutiny. As Stolz et al. (2017b: 16) note, “[t]he complexity of constructions can be measured against several yardsticks”. The yardsticks chosen in their volume on spatial interrogatives are the number of words (mono-word vs. multi-word constructions), the number of morphs and morphemes, zero-marking, the number of syllables, and the number of segments. Based on these yardsticks, they calculate the complexity scores for Place, Goal, and Source constructions and establish a markedness hierarchy which is replicated in Figure 1.

<WHERE; WHITHER; WHENCE>

**Figure 1:** Markedness hierarchy of spatial interrogatives (Stolz et al. 2017b: 585).

According to this hierarchy, the calculated complexity scores increase from WHERE (= Place) via WHITHER (= Goal) to WHENCE (= Source). This hierarchy was confirmed by Nintemann et al. (2020) for spatial interrogatives on the one hand and spatial adverbs on the other hand. As stated in Hypothesis 1 formulated in (3) above, we expect a similar hierarchy, provided in Figure 2, for GL and configurational constructions in combination with a nominal Ground as well. Since measurements on different linguistic levels as conducted by Stolz et al. (2017b) would exceed the scope of this study, measuring the morphological complexity on the basis of the number of morphemes will suffice at this point.

Place < Goal < Source
-----------------------

**Figure 2:** Markedness hierarchy of spatial relations.

To test Hypothesis 1, we first include all languages with no empty columns for our calculations (13 languages<sup>39</sup>). We calculate the mean number of morphemes for each Directionality across all constructions. In case cells are filled with more than one form, we calculate the mean so that for some cells we get a decimal number. Across the 13 languages this results in an average of 2.15 morphemes for Place, 2.24 for Goal, and 2.45 for Source. Even though the average complexity increases only by 0.09 from Place to Goal, it builds up by 0.21 for Source. Thus, Hypothesis 1 can be confirmed.

Subsequently, we check the hierarchy for GL and all configurational constructions individually. For this we only include languages that have no empty cells in the respective spatial relation. The number of languages that are included in the calculations for each spatial relation are indicated in Table 15. Due to the different languages that are included for the different spatial relations, we exclusively compare Place, Goal and Source in individual Configurations. The results are not suitable for making comparisons between the individual Configurations. As can be seen from Table 15, the hierarchy in Figure 2 can be confirmed for GL and all configurational constructions.

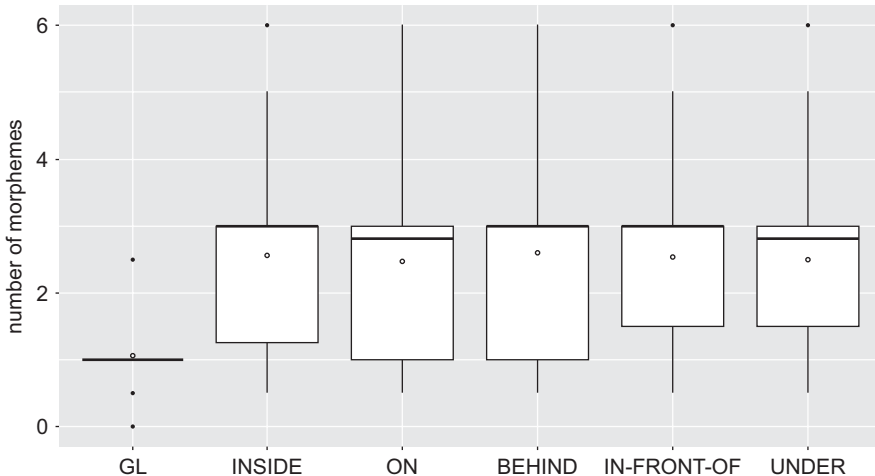
**Table 15:** Average number of morphemes (mean).

Directionality	All spatial constructions	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Number of languages	13	27	30	25	21	16	22
Place	2.15	0.96	2.45	2.29	2.42	2.42	2.33
Goal	2.24	0.96	2.56	2.45	2.54	2.50	2.48
Source	2.45	1.19	2.68	2.68	2.85	2.69	2.68

Furthermore, we expect that in those languages that employ GL constructions, configurational constructions are morphologically more complex than GL constructions. For example, Japanese makes use of the particle *ni* ‘at’, i.e. one morpheme, for GL as presented in (2a) in Section 2, while the ON construction in (2b) consists of three morphemes, viz. *no ue ni* (GEN on at) ‘on’.

<sup>39</sup> Hungarian is not included here as it has no means of expressing GL, cf. Section 4.1.

The sample for the comparison between general and configurational spatial relations includes those 13 languages of our sample that employ GL and have no empty cells for the other configurational constructions. The boxplot in Figure 3 clearly shows that Hypothesis 2 can be confirmed for our small sample. With an average complexity of 1.04, general spatial relations have the lowest score while also showing less variation compared to the other configurational spatial constructions. The complexity for general spatial relations ranges from 0 to 2.5 morphemes while configurational spatial constructions show a range from 0.5 to 6 morphemes. The boxes for the five configurational spatial constructions look almost identical with only minor differences between the means that range from 2.47 morphemes for ON to 2.60 for BEHIND.<sup>40</sup>



**Figure 3:** Boxplot of complexity across spatial relations.

This also shows that our sample languages correspond to what is postulated by the canon in so far as on average, GL is expressed with fewer morphemes than configurational constructions. Although languages are not always in line with the canon by employing one and only one morpheme for each function, the average scores displayed in Figure 3 show quite clearly that constructions encoding both Directionality and Configuration use a larger number of morphemes than constructions encoding only Directionality.

<sup>40</sup> However, due to our small sample size, we do not attempt to make any generalising statements about the differences in complexity of the individual Configurations at this point.

## 5.2 Syncretism

As mentioned in Section 2 above, syncretism is one of the most discussed topics in studies on spatial relations. Previous studies mostly focussed on the distribution of patterns across languages (cf., e.g. Creissels 2006; Pantcheva 2010; Lestrade 2010; Stolz et al. 2017b; Nintemann et al. 2020). In our study, the focus lies on whether languages attest to the same pattern throughout the paradigms of GL and configurational constructions.

We group the 30 languages of our sample into five categories according to the number of filled columns in the paradigms. Table 16 shows the number of languages in each category and the number of languages that can display that same syncretism pattern across all filled columns. In case individual cells were filled with more than one option, we counted whether one of the options was possible across other columns. This is the case in ten languages of our sample. Overall, Hypothesis 3 is confirmed by our data. 11 (= 84.62%) of the 13 languages that have no empty columns are in line with Hypothesis 3. The group of languages with only one empty column is made up of four languages, where three languages show the same syncretism pattern throughout while one does not. All languages in the other three categories with four to two filled columns employ the same pattern throughout. However, it cannot be ruled out that this might be due to the low number of columns that are compared.

**Table 16:** Distribution of syncretism patterns in individual languages.

Number of filled columns	Number of languages	Number of languages that show same patterns	%
6	13	11	84.62
5	4	3	75.00
4	6	6	100.00
3	4	4	100.00
2	3	3	100.00
Total	30	27	

All in all, only three languages (=10%) do not behave in line with Hypothesis 3, while 27 languages (=90%) do. In Shoshoni [Uto-Aztecan], GL employs option I while INSIDE, ON, and UNDER follow option III, and BEHIND and IN FRONT OF show option V. Welsh [Indo-European, Celtic] also attests to three different patterns. GL, INSIDE, and IN FRONT OF follow Option I, with the latter also showing option IV, while ON, BEHIND, and UNDER attest to option II. The third language is Modern Eastern Armenian [Indo-European, Armenic], which follows option I for GL while other configurational constructions show option II.

## 6 Conclusions

Our description of only a selected number of languages in Section 4 shows the varying means of languages to express general and configurational spatial relations. Even though the difficult data collection process limited our sample size to 30 languages, for half of which only incomplete paradigms could be obtained, our hypotheses are confirmed by our data.

Based on the 13 languages that present complete paradigms, our study confirms the markedness hierarchy (viz. Figure 2) observed in former research (cf. Stolz et al. 2017b; Nintemann et al. 2020). Despite only small differences in the average number of morphemes, GL and all configurational constructions confirm our Hypothesis 1 by exhibiting an increase from Place via Goal to Source.

To test Hypothesis 2 that addresses a difference in the morphological complexity of GL vs. configurational constructions, we are again only able to include those 13 languages for which a complete paradigm could be obtained. However, the data clearly reveals that the average GL construction exhibits a distinctly lower morphological complexity than the average configurational construction. In comparison, the five configurational constructions show almost identical complexity in the languages considered. Mizuno's (2024a) findings in his study on coding asymmetries in locational expressions corroborate these results, providing additional support for the hypothesis. He notices that “[m]arkers used for axial [= configurational] locations tend to be more complex than those used for plain [= general] locations.”

27 of our 30 languages, i.e. 90%, show the same syncretism pattern for all configurational spatial relations that are included in our sample for the respective language. Even if only considering those 13 languages with a complete paradigm, the share remains almost the same with 84.62%. We thus conclude that there is a notable tendency to employ one and the same syncretism pattern across GL and configurational constructions and consider our Hypothesis 3 confirmed. Nevertheless, exceptions do occur. Providing explanations for occurring inconsistencies lies beyond the scope of our study and remains a task for the future.

Our study has demonstrated that configurational spatial relations are still an understudied area of research. A substantial number of the grammars that we consulted do not discuss them and only a few cover the whole range of spatial relations. Because of this limitation, we do not have sufficient data for possible generalisations regarding, for example, the complexity of different Configurations or the grammatical categories involved in the expression of spatial relations from a typological perspective. At this point, we want to draw the attention once more to Mizuno's work which seeks to tackle the question of “[w]hat cross-linguistic generalisations can be made regarding the coding of configuration and direction” (Mizuno 2024b). Apart from the shortcomings imposed by our limited data set, we have also

taken a restricted perspective by largely excluding the verbal domain from our analysis. Extending the scope in this regard will certainly offer new insights into the coding of (configurational) spatial relations. While the results of our study need to be tested against a larger and more diverse sample, we hope to have provided some insights into the exploration of configurational spatial relations and their morphological complexity.

## Acknowledgements

We are grateful to Thomas Stolz for helping us develop a topic for a joint project – unaware that it would be published in a collective volume in his honour. Since we all are his students, our gratitude is not limited to his help with this study, but extends to the fact that it is thanks to his support that we were able to pursue this path in linguistics. For this reason, it was important to us to contribute to this collective volume together to show our appreciation for Thomas Stolz's many years of support.

Furthermore, we gratefully acknowledge the comments by Nataliya Levkovich on an earlier version of this study. Our thanks also go to our informants Ani Karapetyan, Stian Lybach, Iwan Rees, and Jingting Ye who provided valuable help with the compilation and understanding of data on Modern Armenian, Kalaallisut, Welsh, and Mandarin Chinese, respectively. We also thank Shogo Mizuno for the exchange about the topic and the current status of his work which helped shape our own perspectives and expectations for this project. The sole responsibility for the contents of this paper including any errors lies with us.

## Abbreviations

1, 2, 3	1st, 2nd, 3rd person
ABL	ablative
ACC	accusative
ALL	allative
ART	article
ATTR	attributive
CA	connective adverbial
CAUS	causative
COLL	collective
DEF	definite
DEST	destinative

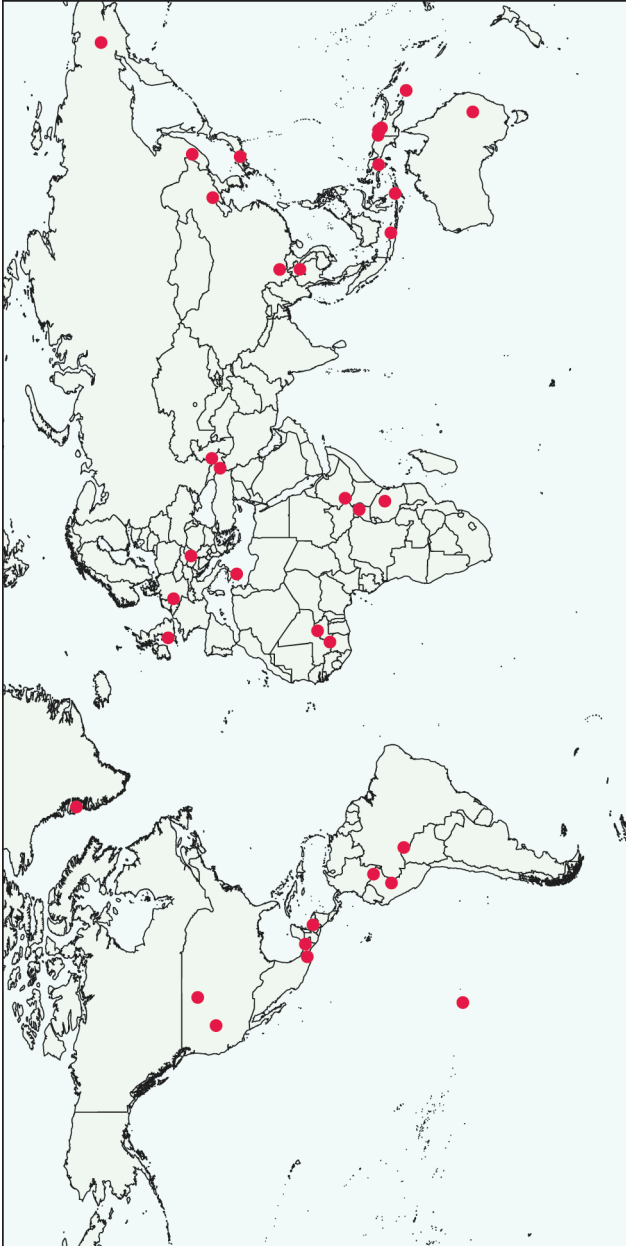
DGS	German Sign Language (Deutsche Gebärdensprache)
DIM	diminutive
DIR	directional
DO	direct object
DS	different-subject
ELA	elative
ERG	ergative
EXCL	exclusive
F	feminine
FOC	focus
FUT	future
FP	future participle
GEN	genitive
GL	general location
GM	Genus-Macroarea
ILL	illative
IMP	imperative
INCL	inclusive
IND	indicative
INE	inessive
INFR	inferential
INST	instrumental
IPFV	imperfective
LAT	lative
LOC	locative
L	lenition
M	masculine
MOV	continuous moving
N	neuter
NEG	negation
NMLZ	nominaliser
NOM	nominative
NTR	neutral aspect (he)
NUM	numeral marker
PL	plural
PFV	perfective
POSS	possessive
PRED	predicative
PROL	prolative
PROM	prominence marker
PROP	proper article
PROX	proximal
PRP	present participle
PRS	present tense
PST	past tense
R	realis
REP	repetitive

RES	resultative
SBJV	subjunctive
SF	S-focus
SG	singular
SING	singulative
SR	spatial relation
SS	same-subject
SUPER	superior
TOP	topic
TR	transitive
V	vowel
VN	verbal noun

## Additional glossing for Sign Languages

[SIGN; SIGN]	Signs in square brackets, divided by a semicolon are signed simultaneously
Xa, Xb	Used when two different loci are established in a similar signing space area
INDEX	Sign that is signed by the pointing finger for referring within the signing space
left	Sign is placed to the left of the signer
left.high	Sign is placed to the left of the signer and slightly higher
left.low	Sign is placed to the left of the signer and slightly lower
left.middle	Sign is placed to the left, but further to the middle of the signer
MANIP:___	Manipulator classifier sign, indicating how an object is handled
SUBS:___	Substitutor classifier sign where the handform replaces the referent

## Appendix 1: Map



**Map 1:** Geographical distribution of sample languages.

## Appendix 2: Paradigms

**Table 17:** Ayutla Mixe [Mixe-Zoque] (Romero-Mendez 2008).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	X=ě'n	X-ojt-py	X-kěx-p			
Goal	X=ě'n	X-ojt-py	X-kěx-p	NA	NA	NA
Source	X=ě'n	X-ojt-py	X-kěx-p			
Option	V	V	V			

**Table 18:** Bora [Boran] (Thiesen and Weber 2012).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place		X pañe	X hallú	X déju		X lliiñé
Goal	NA	X pañe-vu	X hallú-vu	X déju-vu	NA	X lliiñé-vu
Source		X pañe-tu	X hallú-tu	X déju-tu		X lliiñé-tu
Option		I	I	I		I

**Table 19:** Chol [Mayan] (Vázquez Álvarez 2011; Coon 2010).<sup>41</sup>

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	tyi X	tyi i-mali X	tyi i-jol X	tyi i-paty X		tyi y-e'bal X
Goal	tyi X	tyi i-mali X	*tyi i-jol X	*tyi i-paty X	NA	*tyi y-e'bal X
Source	tyi X	tyi i-mali X	*tyi i-jol X	*tyi i-paty X		*tyi y-e'bal X
Option	V	V	V	V		V

**Table 20:** Crow [Siouan] (Graczyk 2007).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	X-n	X (ashk)awúua(-n) X awúualee		X alitchia-n		
Goal	X-ss(aa)	X awúu(a)-s(s) X awúua-ko		X alitchia-s(s)		
Source	X-kaa X-ss(aa) X-n	X awúua(-n)	NA	X alitchia-n	NA	NA
Option	I / III / IV	I / IV		IV		

<sup>41</sup> The forms marked by an asterisk are reconstructed based on Coon (2010: 235–236).

**Table 21:** Garifuna [Arawakan] (Haurholm-Larsen 2016).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X-rugu</i>	<i>t-ídan X</i>				
Goal	<i>X-rugu</i> <i>X-rugu-n</i>	<i>t-íd-on X</i>	NA	NA	NA	NA
Source	<i>X-rugu-giyen</i>	<i>t-ídan-giyen X</i>				
Option	I / II	I				

**Table 22:** Iatmul [Sepik] (Jendraschek 2012).<sup>42</sup>

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X(-ba)</i>	<i>X(-ba)</i>	<i>X(-ba)</i>	<i>X(-ba)</i>	<i>X(-ba)</i>	<i>X(-ba)</i>
Goal	<i>X(-ba)</i>	<i>X(-ba)</i>	<i>X(-ba)</i>	<i>X(-ba)</i>	<i>X(-ba)</i>	<i>X(-ba)</i>
Source	<i>X(-ba)</i> <i>X(-ak)</i>	<i>X(-ba)</i> <i>X(-ak)</i>	<i>X(-ba)</i> <i>X(-ak)</i>	<i>X(-ba)</i> <i>X(-ak)</i>	<i>X(-ba)</i> <i>X(-ak)</i>	<i>X(-ba)</i> <i>X(-ak)</i>
Option	II / V	II / V	II / V	II / V	II / V	II / V

**Table 23:** Imonda [Border] (Seiler 1985).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X-ia</i>	<i>X me-ia</i>	<i>X huls-ia</i>	<i>X mās-ia</i>		
Goal	<i>X-ia-m</i>	<i>X me-ia-m</i>	<i>X huls-ia-m</i>	<i>X mās-ia-m</i>	NA	NA
Source	<i>X-ia-nèi</i>	<i>X me-ia-nèi</i>	<i>X huls-ia-nèi</i>	<i>X mās-ia-nèi</i>		
Option	I	I	I	I		

**Table 24:** Jalkunan [Mande] (Heath 2017).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X b</i>	<i>X dù</i>	<i>X mà</i>			
Goal	<i>X b</i>	<i>X dù</i>	<i>X mà</i>			
Source	<i>X b</i>	<i>X dù</i>	<i>X mà</i>	NA	NA	NA
Option	V	V	V			

<sup>42</sup> Iatmul marks Configuration only on the verb. Directionality can be optionally marked on the noun with a suffix, where there is syncretism either between all three spatial relations (= Option V) or between Place and Goal, while Source can employ a distinct morpheme (= Option II).

**Table 25:** Japanese [Japonic] (Hasegawa 2015; Tanimori 1994).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X ni</i>	<i>X no naka ni</i>	<i>X no ue ni</i>	<i>X no ushiro ni</i>	<i>X no mae ni</i>	<i>X no shita ni</i>
	<i>X de</i>	<i>X no naka de</i>	<i>X no ue de</i>	<i>X no ushiro de</i>	<i>X no mae de</i>	<i>X no shita de</i>
Goal	<i>X ni</i>	<i>X no naka ni</i>	<i>X no ue ni</i>	<i>X no ushiro ni</i>	<i>X no mae ni</i>	<i>X no shita ni</i>
	<i>X e</i>	<i>X no naka e</i>	<i>X no ue e</i>	<i>X no ushiro e</i>	<i>X no mae e</i>	<i>X no shita e</i>
Source	<i>X kara</i>	<i>X no naka kara</i>	<i>X no ue kara</i>	<i>X no ushiro kara</i>	<i>X no mae kara</i>	<i>X no shita kara</i>
Option	I / II	I / II	I / II	I / II	I / II	I / II

**Table 26:** Kalamang [West Bomeraï] (Visser 2022).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X=ko</i>	<i>X(-)ne=ko</i> <i>X nerun=ko</i>				
Goal	<i>X=ka</i> <i>X=ko</i>	<i>X(-)ne=ko</i> <i>X nerun=ko</i>	NA	NA	NA	NA
Source	<i>X=ka</i>	<i>X-ne=ka</i> <i>X nerun=ka</i>				
Option	II / III	II				

**Table 27:** Kolyma Yukaghir [Yukaghir] (Maslova 2003).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X-ge</i>	<i>X molho</i>	<i>X budie</i>			<i>X āl</i>
Goal	<i>X-ge</i> <i>X-ŋin</i> <i>X lanj(n)</i>	<i>X molho-n</i>	<i>X budie-n</i>			<i>X ā-n</i>
Source	<i>X-get</i>	<i>X molho-t</i>	<i>X budie-t</i>			<i>X ā-t</i>
Option	I / II	I	I			I

**Table 28:** Kombio [Torricelli] (Henry 1992).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X Ø</i>	<i>X (pmin)</i>	<i>X (keipm)</i>			<i>X (tapm)</i>
Goal	<i>X Ø</i>	<i>X (pmin)</i>	<i>X (keipm)</i>	NA	NA	<i>X (tapm)</i>
Source	<i>X Ø</i>	<i>X (pmin)</i>	<i>X (keipm)</i>			<i>X (tapm)</i>
Option	V	V	V			V

**Table 29:** Koromfé [Atlantic-Congo] (Rennison 1997).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X nɛ</i>	<i>X joro</i>	<i>X dɔba</i>	<i>X belle</i>	<i>X jika nɛ</i>	<i>X hogo</i>
Goal	<i>X nɛ</i>	<i>X joro</i>	<i>X dɔba</i>	<i>X belle</i>	<i>X jika nɛ</i>	<i>X hogo</i>
Source	<i>X nɛ</i>	<i>X joro</i>	<i>X dɔba</i>	<i>X belle</i>	<i>X jika nɛ</i>	<i>X hogo</i>
Option	V	V	V	V	V	V

**Table 30:** Mehek [Sepik] (Hatfield 2016).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X=k</i>	<i>X siki=k</i>	<i>X ili=k</i>			<i>X nuw=k</i>
Goal	<i>X yoko</i>	<i>X siki yoko</i>	<i>X ili yoko</i>	NA	NA	<i>X nuw yoko</i>
Source	<i>X fenda</i>	<i>X siki fenda</i>	<i>X ili fenda</i>			<i>X nuw fenda</i>
Option	I	I	I			I

**Table 31:** Modern Eastern Armenian [Indo-European, Armenian] (Dum-Tragut 2009; Ani Karapetyan, p.c.).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X-i</i> <i>X-um</i>	<i>X-i meġ</i>	<i>X-i vra</i>		<i>X-i dimacʻ</i>	<i>X-i tak</i>
Goal	<i>X Ø</i>	<i>X-i meġ</i>	<i>X-i vra</i>	NA	<i>X-i dimacʻ</i>	<i>X-i tak</i>
Source	<i>X-icʻ</i> <i>X-yicʻ</i>	<i>X-i meġ-icʻ</i>	<i>X-i vraj-icʻ</i>		<i>X-i dimacʻ-icʻ</i>	<i>X-i tak-icʻ</i>
Option	I	II	II		II	II

**Table 32:** Reta [Alor-Pantar] (Willemsen 2001).<sup>43</sup>

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	Ø	Ø	Ø			
Goal	Ø	Ø	Ø	NA	NA	NA
Source	Ø	Ø	Ø			
Option	V	V	V			

<sup>43</sup> “In Reta such modifiers are all verbal in the sense that they comprise sequences of fully lexical verbal predicates – e.g. oblique participants such as comitatives and allatives are introduced by bivalent verbs, specifications of manner, quality and speed are expressed by means of stative and dynamic verbs in lieu of manner adverbs, etc” (Willemsen 2001: 331).

**Table 33:** Sandawe [Isolate] (Eaton 2010).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X-ts'j</i>	<i>X-tà</i>				
Goal	<i>X-nà</i>	<i>X-tà-nà</i>	NA	NA	NA	NA
Source	<i>X-tfè</i>	<i>X-tà-tfè</i>				
Option	I	I				

**Table 34:** Shipibo-Konibo [Panoan] (Valenzuela 2003).<sup>44</sup>

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X-n(ko)</i> <i>X-ain</i> <i>X-ainko</i>	<i>X meran</i> <i>X chichó</i>	<i>X manaon</i>	<i>X pekáo</i>	<i>X bebon</i>	<i>X naman</i>
Goal	<i>X-n(ko)</i> <i>X-ain</i> <i>X-ainko</i>	<i>X meran</i> <i>X chichó</i>	<i>X manaon</i>	* <i>X pekáo</i>	* <i>X bebon</i>	* <i>X naman</i>
Source	<i>X-nkonía</i> <i>X-ainoa</i> <i>X-ainkoania</i> <i>X-mea</i> <i>X-kea</i>	<i>X mera-mea</i> <i>X meran-oa</i> <i>X chicho-kea</i>	<i>X manaon-kea</i>	* <i>X pekao-kea</i>	* <i>X bebo-mea</i>	* <i>X nama-mea</i>
Option	II	II	II	II	II	II

**Table 35:** Shoshoni [Uto-Aztecan] (Shaul 2012).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X ma</i>	<i>X gupaN</i> <i>X gupandeN</i>	<i>X ba'aN</i> <i>X ba'andeN</i>	<i>X gewayaH</i>	<i>X mamanaí</i>	<i>X dukaN</i> <i>X dukandeN</i>
Goal	<i>X garu</i>	<i>X gupai</i> <i>X gupandi</i>	<i>X ba'ai</i> <i>X ba'andi</i>	<i>X gewayaH</i>	<i>X mamanaí</i>	<i>X dukai</i> <i>X dukandeN</i>
Source	<i>X nai</i>	<i>X gupai</i> <i>X gupandi</i>	<i>X ba'ai</i> <i>X ba'andi</i>	<i>X gewayaH</i>	<i>X mamanaí</i>	<i>X dukai</i> <i>X dukandeN</i>
Option	I	III	III	V	V	III

<sup>44</sup> The forms marked by an asterisk are reconstructed based on Valenzuela (2003: 228, 231).

**Table 36:** Tabasaran [Nakh-Daghestanian] (Babaliyeva 2023).

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place		<i>X-f</i>	<i>X-'il</i>	<i>X-k</i>		<i>X-k</i>
Goal	NA	<i>X-f-na</i>	<i>X-'il-na</i>	<i>X-k-na</i>	NA	<i>X-k-na</i>
Source		<i>X-f-an</i>	<i>X-'il-an</i>	<i>X-k-an</i>		<i>X-k-an</i>
Option		I	I	I		I

**Table 37:** Yuwaalaraay [Pama-Nyungan] (Giacon 2014).<sup>45</sup>

SR	GL	INSIDE	ON	BEHIND	IN FRONT OF	UNDER
Place	<i>X-da</i>	<i>X-da</i> <i>mudhuu-ga</i>	<i>X-da gaburran-da</i>	<i>X-da bawa-ga</i>	<i>X-da bani-dja</i>	<i>X-da ganhaga-dha</i>
Goal	<i>X-gu</i>	<i>X-gu</i> <i>mudhuu-gu</i>	<i>X-gu gaburran-gu</i>	<i>X-gu bawa-guu</i>	<i>X-gu bani-guu</i>	<i>X-gu ganhaga-y</i>
Source	<i>X-dji</i>	<i>X-dji</i> <i>mudhuu-dhi</i>	<i>X-dji gaburran-di</i>	<i>X-dji bawa-di</i>	<i>X-dji bani-dji</i>	<i>X-dji ganhada-dhi</i>
Option	I	I	I	I	I	I

<sup>45</sup> Yuwaalaraay shows a big variety in non-core uses of the locative cases that are mostly driven lexically, as can be seen in example (6). We cannot account for all individual verbs where, for example, the ablative case construction is used for a Goal function or else. Also, verbs that carry configurational meaning do not need to use relational nouns, they then only mark Directionality (Giacon 2014: 45). As we did not focus on verbs, this is left out of our analysis.

## References

- Babaliyeva, Ayten. 2023. Standard Tabasaran: Short grammar sketch. *Languages of the Caucasus* 6(1). 1–51.
- Behrens, Kevin. 2024. De Stellen vun uutwäälte europäasche Gebärdenspraken in'n Standard Average European-Spraakbund. *Journal for EuroLinguistiX* 21. 1–78.
- Borg, Albert J. 1988. To be or not to be a Copula in Maltese. *Journal of Maltese Studies* 17–18. 54–71.
- Borg, Albert & Marie Azzopardi-Alexander. 1997. *Maltese*. London: Routledge.
- Bourdin, Philippe. 1997. On Goal-bias across languages: modal, configurational and orientation parameters. In Bohumil Palek (ed.), *Proceedings of LP'96: Typology, item orderings and universals*. (Proceedings of the Conference held in Prague, August 20–22, 1996), 186–218. Prague: Karolinum - Charles University Press.
- Chao, Yuen Ren. 2011. *A grammar of spoken Chinese*. (Selected works of modern Chinese learning). Beijing: The Commercial Press. [originally published in 1968 by California University Press]
- Comrie, Bernard, Martin Haspelmath & Balthasar Bickel. 2015. *The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses*, <https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf>. [28 August, 2024].
- Coon, Jessica. 2010. *Complementation in Chol (Mayan): A theory of split ergativity*. Cambridge, MA: Massachusetts Institute of Technology dissertation.
- Corbett, Greville G. 2005. The canonical approach in typology. In Zygmunt Frajzyngier (ed.), *Linguistic Diversity and Language Theories, 25–50*. Philadelphia: John Benjamins Publishing Company.
- Creissels, Denis. 2006. Encoding the distinction between location, source and direction: A typological study. In Maya Hickmann & Stéphane Robert (eds.), *Space in Languages: Linguistic Systems and Cognitive Categories*, 19–28. Amsterdam, Philadelphia: John Benjamins.
- Cysouw, Michael & Jeff Good. 2013. Languoid, doculect, glossonym: Formalizing the notion 'language'. *Language Documentation & Conservation* 7. 331–360.
- Du Feu, Verónica. 1996. *Rapanui*. (Descriptive grammars). London: Routledge.
- Dékány, Éva & Veronika Hegedűs. 2021. Chapter 2 Postpositions: formal and semantic classification. In Katalin É. Kiss & Veronika Hegedűs (eds.), *Syntax of Hungarian: Postpositions and postpositional phrases*, 14–191. Amsterdam University Press.
- Dum-Tragut, Jasmine. 2009. *Armenian: Modern Eastern Armenian*. (London Oriental and African language library 14). Amsterdam: Benjamins.
- Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68. 81–138.
- Eaton, Helen. 2010. *A Sandawe grammar*. Dallas: SIL.
- Eberle, Sarah E. 2013. *Locative expressions in signed languages a cross-linguistic comparison*. Grand Forks, ND: University of North Dakota MA thesis.
- Eichmann, Hanna, Martje Hansen & Jens Heßmann. 2012. *Handbuch Deutsche Gebärdensprache – Sprachwissenschaftliche und anwendungsbezogene Perspektiven*. Seedorf: Signum Verlag.
- Everett, Daniel L. & Barbara Kern. 1997. *Wari: the Pacaas Novos Language of Western Brazil*. (Descriptive Grammar Series.) London: Routledge.
- Giacon, John. 2014. *A grammar of Yuwaalaraay and Gamilaraay: a description of two New South Wales languages based on 160 years of records*. Canberra: The Australian National University dissertation.
- Graczyk, Randolph. 2007. *A grammar of Crow: Apsálooke Aliláau*. Lincoln: University of Nebraska Press.
- Hagège, Claude. 2010. *Adpositions*. Oxford: Oxford University Press (Oxford studies in typology and linguistic theory).

- Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2024. Glottolog 5.1. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://doi.org/10.5281/zenodo.10804357>. Available online at <http://glottolog.org> [last accessed 01.11.2024].
- Hasegawa, Yoko. 2015. *Japanese: A linguistic introduction*. Cambridge United Kingdom: Cambridge University Press.
- Haspelmath, Martin. 2019. Differential place marking and differential object marking. *STUF - Language Typology and Universals* 72(3). 313–334.
- Hatfield, Adam. 2016. *A grammar of Mehek*. Buffalo: University of Buffalo PhD dissertation.
- Haurholm-Larsen, Steffen. 2016. *A grammar of Garifuna*. Bern: University of Bern dissertation.
- Heath, Jeffrey. 2017. *A grammar of Jalkunan (Mande, Burkina Faso)*. Ann Arbor: Univ. of Michigan. (Unpublished manuscript).
- Hegedűs, Veronika & Éva Dékány. 2021. Chapter 3 The internal syntax of PPs. In Katalin É. Kiss & Veronika Hegedűs (eds.), *Syntax of Hungarian: Postpositions and postpositional phrases*, 194–249. Amsterdam University Press.
- Henry, Joan. 1992. *Kombio grammar essentials*. Ukarumpa: Summer Institute of Linguistics.
- Holst, Jan Henrik. 2005. *Einführung in die eskimo-aleutischen Sprachen*. Hamburg: Helmut Buske Verlag.
- Ikegami, Yoshiko. 1987. ‘Source’ and ‘Goal’: A case of linguistic dissymmetry. In René Dirven & Günther Radden (eds.), *Concept of case*, 122–146. Tübingen: Gunter Narr Verlag.
- Jendraschek, Gerd. 2012. *A grammar of Iatmul*. Regensburg: Universität Regensburg dissertation.
- Kenesei, Istvan, Robert M. Vago & Anna Fenyevesi. 1998. *Hungarian*. (Descriptive Grammars). Hoboken: Taylor and Francis.
- Kievit, Paulus. 2017. *Grammar of Rapa Nui*. (Studies in Diversity Linguistics 12). Berlin: Language Science Press.
- King, Gareth. 2003. *Modern Welsh: A comprehensive grammar* (Routledge Grammars). London: Routledge.
- Kopecka, Anetta & Marine Vuillermet. 2021. Source-Goal (a)symmetries across languages. *Studies in language* 45(1). 2–35.
- Korpus Malti. 2016. version 3.0. Available online at: <https://mlrs.research.um.edu.mt/index.php?page=corpora>.
- Lestrade, Sander. 2010. *The space of case*. Nijmegen: Radboud Universiteit.
- Li, Charles N. & Sandra A. Thompson. 1989. *Mandarin Chinese: A functional reference grammar*. Berkeley, Los Angeles: University of California Press.
- Lybäck, Stian. 2022. An Introduction to West Greenlandic. 2nd edition. Online manuscript: <https://oqa.dk/assets/aitwg2ED.pdf> [last accessed 08.11.2024]
- Maslova, Elena. 2003. *A grammar of Kolyma Yukaghir* (Mouton Grammar Library 27). Berlin, New York: Walter de Gruyter.
- Miestamo, Matti, Dik Bakker & Antti Arppe. 2016. Sampling for variety. *Linguistic Typology* 20(2). 233–296.
- Mizuno, Shogo. 2024a. Coding asymmetry types in locational expressions and their complexity. Talk at *Vielfaltslinguistik 6*, Graz, 14.–15. June, 2024.
- Mizuno, Shogo. 2024b. Axial forms exhibit stronger cohesion with the ground nominal than directional markers. Talk at *LLcD 2024*, Paris, 9.–11. September, 2024.
- Nielsen, Flemming A. J. 2019. *Vestgrønlandsk grammatik*. København: Books on Demand.
- Nikolaeva, Irina & Maria Tolskaya. 2001. *A grammar of Udihe*. Berlin, New York: Mouton de Gruyter.
- Nintemann, Julia. 2024. Spatial asymmetries and variation: On the role of the Ground in Place, Goal, and Source constructions. Talk at the *57th Annual Meeting of the Societas Linguistica Europaea*, Helsinki, 21.–24. August 2024. (<https://doi.org/10.5281/zenodo.13379881>)

- Nintemann, Julia. In preparation. The morphosyntax of (GOING) HOME in typological perspective. *STUF – Language Typology and Universals*.
- Nintemann, Julia & Nicole Hober. 2023. On the morphosyntax of place names vs. common nouns in pidgins and creoles: The encoding of two types of Grounds in Goal and Source constructions. In Nataliya Levkovych (ed.), *Diversity in Contact* (KPL 21), 195–248. Berlin/ Boston: De Gruyter.
- Nintemann, Julia & Maja Robbers. 2019. Towards a typology of spatial deictic expressions. *STUF – Language Typology and Universals* 72(3), 335–371.
- Nintemann, Julia, Maja Robbers & Nicole Hober. 2020. *Here – Hither – Hence and Related Categories: A Cross-linguistic Study*. (Studia Typologica 26). Berlin, Boston: De Gruyter.
- Pantcheva, Marina. 2010. The syntactic structure of locations, goals, and sources. *Linguistics* 48(5), 1043–1081.
- Po-Ching, Yip & Don Rimmington. 2006. *Chinese: An essential grammar*. 2nd Edition. London/New York: Routledge.
- Quer, Josep, Carlo Cecchetto, Caterina Donati, Carlo Geraci, Meltem Keleşir, Roland Pfau & Markus Steinbach. 2019. *SignGram Blueprint – A Guide to Sign Language Grammar Writing*. Berlin/Boston: De Gruyter.
- Robbers, Maja & Nicole Hober. 2018. Verb-framed spatial deixis in Mesoamerican languages and the increasing complexity of source constructions vis Spanish *de*. *STUF – Language Typology and Universals* 71(3), 397–423.
- Rennison, John R. 1997. *Koromfe*. (Descriptive Grammar Series). London: Routledge.
- Romero-Mendez, Rodrigo. 2008. *A reference grammar of Ayutla Mixe*. Buffalo: University at Buffalo dissertation.
- Ross, Claudia & Jing-heng Sheng Ma. 2006. *Modern Mandarin Chinese grammar: A practical guide*. London: Routledge.
- Sandler, Wendy & Diane Lillo-Martin. 2006. *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press.
- Seiler, Walter. 1985. *Imonda, a Papuan language*. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Shaul, David Leedom. 2012. Survey of Shoshone Grammar with Reference to Eastern Shoshone. Ms. [<https://easternshoshone.org/wp-content/uploads/2021/01/survey-of-shoshone-grammar1.pdf> accessed 14.11.2024]
- Stolz, Thomas. 1990. Flexion und Adpositionen, flektierte Adpositionen, adpositionelle Flexion. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 43(3), 334–354.
- Stolz, Thomas. 1992. *Lokalkasussysteme: Aspekte einer strukturellen Dynamik*. Weinberg: Egert.
- Stolz, Thomas. 2018. Deiktische Antworten auf räumliche Fragen. In Angelika Wöllstein, Peter Gallmann, Mechthild Habermann & Manfred Krifka (eds.), *Grammatiktheorie und Empirie in der germanistischen Linguistik*, 309–330. Berlin/Boston: De Gruyter Mouton.
- Stolz, Thomas. to appear. Erkenntnis durch Widerspruch: Was Sprachen müssen und was sie dürfen. In Andrea Daase & Anna Mattfeldt (eds.), *Mehrsprachigkeit im Widerspruch: (Un-)Doing linguistic and cultural differences – sprachliche Praktiken in der Migrationsgesellschaft*. (Beiträge zur soziokulturellen Theorie der Sprachaneignung). Münster: Waxmann.
- Stolz, Thomas, Sander Lestrade & Christel Stolz. 2014. *The Crosslinguistics of Zero-Marking of spatial relations*. (Studia Typologica 15). Berlin: De Gruyter Mouton.
- Stolz, Thomas, Nataliya Levkovych & Aina Urdze. 2017a. Spatial interrogatives: Typology and dynamics (with special focus on the development from Latin to Romance). In Silvia Luraghi, Tatiana Nikitina & Chiara Zanchi (eds.), *Space in Diachrony*, 207–240. Amsterdam: Benjamins.

- Stolz, Thomas, Nataliya Levkovich, Aina Urdze, Julia Nintemann & Maja Robbers. 2017b. *Spatial interrogatives in Europe and beyond: Where, whither, whence*. (Studia Typologica 20). Berlin/Boston: De Gruyter Mouton.
- Stolz, Thomas, Nataliya Levkovich & Aina Urdze. 2017c. When zero is just enough . . . In support of a special toponymic grammar in Maltese. *Folia Linguistica* 51(2). 453–482.
- Stolz, Thomas, Nataliya Levkovich & Maike Vorholt. 2023. Variable overt marking of Place/Goal with place names as complements: On the competition between *fi*, *ġewwa*, and *ġo*. In Michael Spagnol & Ray Fabri (eds.), *On Maltese Syntax*. [Special issue]. *Journal of Maltese Studies* 30. 177–215.
- Stolz, Thomas & Julia Nintemann. 2024. *Special Onymic Grammar in typological perspective: Cross-linguistic data, recurrent patterns, functional explanations*. (Studia Typologica 34). Berlin/Boston: De Gruyter Mouton.
- Stolz, Thomas & Maike Vorholt. in preparation. *The Grammar of Maltese Prepositions*. (Studia Typologica 35). Berlin/Boston: De Gruyter Mouton.
- Talmy, Leonard. 1985. Lexicalization patterns. Semantic structure in lexical forms. In Timothy Shopen (ed.), *Language Typology and Syntactic Description 3: Grammatical Categories and the Lexicon*, 57–149. Cambridge: Cambridge University Press.
- Talmy, Leonard. 1991. Path to realization: A typology of event conflation. In Christopher Johnson, Laurel A. Sutton & Ruth Shields (eds.), *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistic Society*, 480–519. Berkeley: Berkeley Linguistics Society.
- Tanimori, Masahiro. 1994. *Handbook of Japanese grammar*. Boston: Tuttle Publishing.
- Thiesen, Wesley, David Weber. 2012. *A Grammar of Bora with Special Attention to Tone*. Dallas, Texas: SIL International.
- Ursini, Francesco-Alessio. 2020. General location across languages: On the division of labour between functional and lexical items in spatial categories. *The Linguistic Review* 37(4). 495–542.
- Valenzuela, Pilar M. 2003. *Transitivity in Shipibo-Konibo grammar*. Eugene: University of Oregon dissertation.
- Vázquez Álvarez, Juan J. 2011. *A grammar of Chol: A Mayan language*. Austin, TX: The University of Texas at Austin dissertation.
- Visser, Eline. 2022. *A grammar of Kalamang*. Berlin: Language Science Press.
- Willemsen, Jeroen. 2021. *A Grammar of Reta*. Aarhus University dissertation.



Johanna Mattissen

# Aspectual head/dependent marking in relation to verb/satellite framing

**Abstract:** The present study tries to set structures into relation, viz. head marking, double marking and dependent marking of aspectual values, with a special eye on stacking, conflation, differential lexicalisation and differential voice marking on verbs of consumption, production and goal-attainment, on the one hand, with verb, satellite and serial framing as well as case and adposition inventories on the other. Both aspect languages and languages with an advanced aspectoid system are analysed. Besides polypersonalism, voice/valency and preverbs turn out to be relevant criteria in correlations; however, aspectual head-marking does not cooccur with verb framing in the present sample.

**Keywords:** aspectuality, differential lexicalisation, differential voice marking, goal-attainment verbs, lexicalisation patterns

## 1 Aspect and advanced aspectoid systems

Aspect and aspectuality are most controversial categories in linguistic discussion and have been approached with different frameworks (for an overview cf. Sasse 2002).

The present contribution discusses features beyond aspect and aspectoid markers that may interact with aspectuality in order to find correlations, with a focus on verbs of consumption, production and goal-attainment (typical candidates for the successive-terminative actional class of verbs, cf. Mattissen (2024: 173–175)). Those features are head/double/dependent marking of aspectual values, stacking

---

**Acknowledgements:** I gratefully acknowledge research funding for this paper by the German Research Foundation (DFG), Project-ID 281511265, SFB 1252, and I very much appreciate support from the 2011 Grant-in-Aid for Scientific Research of Sapporo Gakuin University for the International Project ‘Description of the West Sakhalin dialect of Nivkh’, headed by Hidetoshi Shiraiishi. I would like to express my profound gratitude to all those who discussed a considerable battery of questions, tests and examples with me, especially Maria Bardají i Farré, Sevim Genç and Luana Reveriot, as well as Aleksandra Vladimirovna Khuriun and Zoia Ivanovna Liutova for generously sharing their language expertise on Nivkh with me. All errors are, of course, my own.

---

**Johanna Mattissen**, University of Cologne, e-mail: Johanna.Mattissen@uni-koeln.de

of seemingly contradictory aspectual values, differential lexicalisation, differential voice marking, verb/satellite/serial framing, local and directional case or adposition systems. To this end, languages with different formal systems of encoding aspect or aspectual values have been chosen.

The study is based on the descriptive model established in Mattissen (2024), which is applicable cross-linguistically; it assumes the interaction between lexically-inherent actional classes (on the basis of Breu 1992, 1994, 1996, 1997, 2000, 2007, 2009; currently 13 classes, which may be represented in any language independent of an aspect system), a morphosyntactic and binary aspect opposition (perfective – imperfective) and/or aspectoid markers (such as progressive, completive, habitual etc.). The actional classes differ in their design with respect to lexicalised boundaries (initial, final, intermediate) and phases (static, dynamic, accruing, (non-)interruptible, temporary or permanent). They can only be tested language-specifically, and cannot be predicted, especially not by reasoning (linguistic symbols stand for concepts/notions of extra-linguistic realities only).

In a fully grammaticalised, i.e. binary, aspect system, the perfective forms select the inherent boundaries or the temporariness of a verbal meaning (according to the respective actional class), hence they signal boundary transgression or temporal delimitation and are used for sequential states of affairs, as well as for the intruding state of affairs in the incidence taxis constellation. Imperfective forms select the inherent phase(s), hence signal an existing situation “intra-boundaries” (e.g. action in progress) and are particularly used for states of affairs in simultaneous taxis constellations, the backgrounded state of affairs in the incidence constellation as well as for habitual and generic states of affairs, to give a rough orientation. Arguments and adverbials must be compatible with the inherent valence of the verb and with the manner of affectedness of a participant (e.g. holistic, gradual, partial), both inherent in the verbal meaning. They can activate the inherent (and pre-defined) actional class readings, but do not change the actional class of the VP or other syntagmatic expressions (Mattissen 2024).

Besides fully grammaticalised and obligatory aspect systems, for instance in Romance (e.g. French, Spanish, Romanian), Kartvelian (e.g. Laz, Georgian), Slavic (e.g. Bulgarian, Russian) and Semitic (e.g. Modern Standard Arabic, Maltese) languages, Tagalog (Austronesian), Cayuga (Iroquoian) and Ngan’gityemerri (Southern Daly), there are languages with advanced aspectoid systems which do not (yet) have a binary opposition of perfective and imperfective forms, such as English, Hungarian (Finno-Ugric), Japanese, Turkish (Turkic), Ket (Yeniseian) or Nivkh (Paleosiberian).

Among the languages with a fully grammaticalised, i.e. binary aspect opposition, this opposition is encoded, for instance, by different verb stems or suppletion (e.g. Greek), inflectional endings (e.g. Romance and Kartvelian languages), by

affixes on the verb (e.g. Slavic and Kartvelian languages), by reduplication (e.g. Tagalog, Ngan'gityemerri), light verbs (e.g. Ngan'gityemerri, Basque) or combinations of the above. These languages may also have aspectoid markers that combine with the aspect forms and are at different stages of grammaticalisation, e.g. in Maltese (Stolz and Ammann 2008). Among the languages with an advanced aspectoid system, basically the same range of forms can be observed, but forms are not in a binary opposition.

This contribution proceeds as follows: Section 1 introduces the classification of head, dependent and double marking of aspectual values and elaborates on the stacking of seemingly contradictory aspectual values in contrast to conflation. Section 2 deals with differential lexicalisation and section 3 with goal-attainment verbs. Section 4 discusses the relation of verb, satellite and serial framing to goal-attainment verbs, differential lexicalisation and the inventory of local cases and adpositions in languages and works out whether there are correlations between the structural phenomena presented. Section 5 concludes.

## 1.1 Head marking and dependent marking of aspectual values

The first concept to be introduced here is the distinction of head and dependent marking of aspectual values. When aspect languages as well as languages with an advanced aspectoid system mark all aspectually relevant material on the verbal head and do not need participants or adverbials to activate one or the other reading of the inherent actional class this is understood to be head marking of aspectual values here. When languages of either type rely on participants, adverbials or copredicates to activate an inherent intra-boundaries reading or boundary-transgression reading, this strategy is called dependent marking here. Dependent marking takes the form of the presence of a participant or adverbial, of quantifying and individuating morphology on participants, including differential object marking (DOM) and articles, and marking of adverbials. Most languages seem to make use of a mixed strategy, which is called double marking here.

Head, dependent and double marking are analysed together with the distinction of full and partial affectedness and effectedness with verbs of consumption and production in this section.

## 1.2 Dependent marking of aspectual values

Dependent marking of aspectual values may seem an odd concept as “aspect” is a verbal category, represented by binary perfective and imperfective forms. However,

“aspectuality”, the notional distinction of a state of affairs presented as an existing phase “intra-boundaries” (e.g. progressive, stative, partial affectedness/effectiveness) vs. as a boundary transgression (e.g. ingressive or completive readings) or as a state of affairs *en bloc* (delimitative reading) can be expressed by other means than the verb, in interaction with lexically inherent actional values of verbs.

A candidate for (strong) dependent marking of aspectuality, taking the form of dependents activating the verb-inherent readings, is Finnish (Finno-Ugric). The language has no perfective/imperfective verb forms, but a few (head marking) aspectoid markers (Leinonen 1984: 245; Bridgen 1984: 193; Kiparsky 1998: 293), and no article system, but uses DOM for encoding a binary opposition of aspectual values (Kiparsky 1998:267, 272). The actional class of a verb controls whether DOM is possible or whether the verb governs almost only the partitive or only the accusative case (Heinämäki 1984: 154, 163–165; cf. Kiparsky 1998: 281–282; Larsson 1984: 99). If DOM is possible, partitive is chosen for partial affectedness (1b, d), encoding of ongoing situations (1d, f, h) including backgrounded states of affairs in the incidence constellation (1k), conativity or iterativity/habituality ((1h), depending on the concrete actional class of the verb), i.e. shades of non-transgression of the final boundary; accusative case activates a boundary-transgression reading, in particular full affectedness or effectiveness of a participant and goal attainment (1a, c, e, g, i). Compare the following examples:

(1) Finnish (Kiparsky 1998: 267, 272, 273; Leinonen 1984: 246–247, 249; Heinämäki 1984: 154)

- a. *ostin omen-at*  
buy.1SG.PST apple-ACC.PL  
‘I bought (all) the apples.’
- b. *ostin omen-oita*  
buy.1SG.PST apple-PART.PL  
‘I bought (some of the) apples.’
- c. *hän kirjoitti kirje-t*  
3SG write.PST.3SG letter-ACC.PL  
‘s/he wrote the letters’
- d. *hän kirjoitti kirje-itä*  
3SG write.PST.3SG letter-PART.PL  
‘s/he wrote (some) letters’ or ‘s/he was writing (the) letters when. . .’
- e. *kaadoin puun*  
fell.PST.1SG tree-ACC  
‘I felled a tree.’

- f. *kaadoin puuta*  
 fell.PST.1SG tree-PART  
 'I was felling a tree.'
- g. *hän avasi ikkuna-n*  
 3SG open.PST.3SG window-ACC.SG  
 's/he opened the window'
- h. *hän avasi ikkuna-a*  
 3SG open.PST.3SG window-PART.SG  
 's/he was opening the window, when. . .', or 's/he opened the window for a while', or 's/he opened the window a bit/partly', or 's/he opened the window again and again'
- i. *ammuin karhu-n / karhu-t*  
 shoot.PST.1SG bear-ACC.SG bear-ACC.PL  
 'I shot the/a bear / the bears'
- j. *ammuin karhu-a / karhu-ja*  
 shoot.PST.1SG bear-PART.SG bear-PART.PL  
 'I shot at the/a bear / (the) bears'
- k. *Pidin hattu-a käde-ssä-ni*  
 keep.PST.1SG hat-PART.SG hand-INESS-my  
 'I kept the hat in my hand . . .' (e.g. when s.o. entered.)

The Finnish partitive case differs in function from partitives in French (partitive article *du, de la, des*) and Russian (see Section 1.4, examples (11), (12)). In these languages the partitive is restricted to mass and plural nouns. In Finnish, count nouns in the singular or plural and even persons' names (Heinämäki 1984: 171) are marked with partitive. It is the affectedness of the participant that plays the crucial role, and affectedness is the effect of the state of affairs on the participant, controlled by the lexically inherent readings of the verb. Said the other way round, it is not the properties of the noun that control DOM, but the actional class and reading of the verb. Nevertheless, affectedness marking is not on the verb, but on its participant, the dependent. I am not aware of further equally strong dependent-marking aspect systems (cf. also Kiparsky 1998: 296).

### 1.3 Head marking of aspectual values

In contrast to the strategy of having a dependent activate the readings provided for by the actional class of the verb, in pure head-marking aspectual systems dependents are not needed for that activation. Typically, in such languages participants need not be overt and/or marked for definiteness, number/quantification and/or

case (i.e. have individuating or quantifying features). In contrast, in addition to aspect and/or aspectoid markers, predicates tend to have polypersonal and referential inflection (cf. Evans 2002) or inflection for the undergoer (including undergoer oriented voice, as in Tagalog, or other person constellation marking, as in Japanese (Mattissen 2015)), making the presence or absence of an overt nominal participant irrelevant for the aspectual reading.

Aspectually head-marking languages are, for instance, Laz, Tagalog, Ngan'gityemerri, Cayuga and Modern Standard Arabic, as well as Nivkh and Japanese with an advanced aspectoid system.

Laz has fully grammaticalised perfective and imperfective polypersonal verb forms, but no case and referentiality marking on the participants, which need not be overt; number marking is marginal. Its actional classes were discussed in Mattissen (2001, 2003a). Full affectedness or effectedness is additionally signalled by the use of a completive preverb on the perfective form.

(2) Laz (Mattissen 2003a: 263, 2001: 38, 1995: 63)

- a. (*oşkuri*) *şkomi*  
apple eat.PFTV.PST.(1>3)SG  
'I ate (an/the apple/(the) apples)'
- b. (*oşkuri*) *bimxort'i*  
apple eat.IPFV.PST.(1>3)SG  
'I was eating (an apple/(the) apples).'
- c. *mektubi* *nčaru*  
letter write.PFTV.PST.(3>3)SG  
'S/he wrote a letter/letters.'
- d. *mektubi* *do-nčaru*  
letter CPL-write.PFTV.PST.(3>3)SG  
'S/he wrote (and finished) the letter(s).'

In its sister language Georgian, the overt nominal participant of a perfective verb form does not change its case marking in dependence of the presence or absence of the completive preverb on the perfective form. With verbs of consumption and production, it is the presence of the completive preverb that signals full affectedness or effectedness, its absence partial affectedness.

(3) Georgian (Vogt 1936: 238)

- a. *þuri* *v-čame*  
bread 1SG-eat:1SG.PFTV.PST  
'I ate (at the) bread'

- b. *ṗuri še-v-čame*  
 bread CPL-1SG-eat:1SG.PFTV.PST  
 'I ate up the bread'

Ngan'gityemerri, a polysynthetic language, presents a similar picture where participants are concerned. The language has verbs in several actional classes and perfective and imperfective inflection as well as aspectoids (Reid 2011: 176–178, 185–188, 230–232, 240–244). The nominal participants are not marked for any categories.

- (4) Ngan'gityemerri (Reid 2011: 177–178)
- a. *egeningge wubum-da*  
 goose 3SG.Bash.PFTV-shoot  
 'he shot a goose'
- b. *egeningge webe-dada-tye*  
 goose 3SG.Bash.IPFV-shoot.RED-PST  
 'he was shooting geese'

The Tagalog aspect system also makes use of reduplication for the imperfective forms. Tagalog verb forms are apersonal, but distinguish two actor voices and three undergoer voices.

- (5) Tagalog, extract of the paradigm of 'buy' (cf. Latrouite 2011: 31–32)
- |                        |                     |                   |
|------------------------|---------------------|-------------------|
| actor voice realis     | perfective          | imperfective      |
| -um-                   | <i>bumili</i>       | <i>bumibili</i>   |
| undergoer voice realis | "effective voice"   |                   |
| -in                    | <i>binili</i>       | <i>binibili</i>   |
| undergoer voice realis | "superficial voice" |                   |
| -an                    | <i>binilhan</i>     | <i>binibilhan</i> |
| undergoer voice realis | "theme voice"       |                   |
| i-                     | <i>ibinili</i>      | <i>ibinibili</i>  |

The lexical actional class of the verb and the voice affix on the predicate control the affectedness or effectedness of the participant in the *ang*-phrase of a clause. Undergoer voice *-in* signals an effected or totally affected participant or an attained goal, undergoer voice *-an* a superficially or partially affected participant, including path, source and recipient, and undergoer voice *i-* a moved theme, an instrument or beneficiary (cf. Himmelmann 1987: 132–139). Note that the nominal participants are not marked for referentiality or quantification in (6).

## (6) Tagalog (Latrouite 2011: 165–166, 158–159)

- a. *i-sulat mo ang pangalan*  
 TV-write 2SG.GEN NOM name  
 ‘write your name’ (the name is not created by writing)
- b. *sulat-in mo ang liham*  
 write-EV 2SG.GEN NOM letter  
 ‘write a/the letter’
- c. *sulat-an mo ang papel / ang abogado*  
 write-SV 2SG.GEN NOM paper NOM lawyer  
 ‘write on paper / write to the lawyer’
- d. *kain-in mo ang isda.*  
 eat-EV 2SG NOM fish fully affected  
 ‘you eat the fish’
- e. *kain-an mo ang isda.*  
 eat-SV 2SG NOM fish partially affected  
 ‘you eat from the fish’

Not just any verb is compatible with all of the voices (Himmelmann 2008: 499, 1987: 129–145). Typically this “differential voice marking” (DVM), as it is called here, is possible for verbs of production (6a–c), consumption (6d–e) and goal-attaining motion (see Section 3, (42)). As such verbs are characterised by an inherent dynamic and incremental phase which can be interrupted at any point, as well as a final boundary, DVM distinguishes a delimitative reading (end of a state of affairs without final boundary transgression) and a completive reading (with final boundary transgression, when the undergoer is fully affected or effected by the state of affairs).

Insofar, Tagalog with its head-marking DVM is the mirror image of Finnish, where affectedness or effectedness is marked on the participant by DOM, i.e. dependent marking.

Cayuga, a polysynthetic language like Ngan’gityemerri, but of the holophrastic type (Mattissen 2017a: 90), operates with predicate forms (Sasse 2000: 238) and does not have nominal participants as such that could activate one or the other reading.

## (7) Cayuga (Sasse 2000: 231–232)

- |   |                                     |
|---|-------------------------------------|
| imperfective                              | perfective                          |
| <i>aka:trɛ́:no:t</i>                      | <i>akatɾɛ́nó:tɛ́</i>                |
| ‘I am singing’ = ‘the voice stands on me’ | ‘I sang’ = ‘I made the voice stand’ |

Cayuga makes use of voice marking for aspect: the causative forms are perfective, the semireflexive forms are imperfective (intransitive-anticausative, Sasse 2000: 196, 232, 238), as well as of differential lexicalisation (see Section 2) for a finer-grained distinction of actional classes.

Polysynthetic Nivkh with an advanced aspectoid system incorporates the “incremental” participant as well, in the form it has as a free-standing word form.

(8) Nivkh (Mattissen 2012)

- a. *ʃas ñaqr-ux pityə-uru-ɖ*  
hour one-LOC book-read-IND/NML  
‘I read a book for one hour.’
- b. *ʃas ñaqr-ux pityə-uru-ɣər-ʃ*  
hour one-LOC book-read-CPL-IND/NML  
‘I read a book in one hour.’

Another aspectually head-marking language is Russian. In Russian, most aspect pairs of verbs consist of an unmarked imperfective form and a perfective form bearing a grammaticalised preverb (cf. Isačenko 1962). Nominal participants are marked with the same case and number in both aspects. Verbs of consumption and production typically have two perfective forms with two different grammaticalised preverbs, one with a delimitative reading (mostly with the preverb *po-*), signalling that the state of affairs has stopped or ended without transgressing the final boundary, i.e. without fully affecting/effecting the participant (‘V-ed for a while/spent some time V-ing’); the other with a completive reading, i.e. signalling the transgression of the final boundary through full affectedness/effectiveness of the participant (cf. Mattissen 2024: 174).

(9) Russian (Isačenko 1962: 359, 360, 362, 383, 391–392)

infinitive		3SG.M.CPL.PFTV	3SG.M.DEL.PFTV
<i>est’</i>	‘eat’	<i>s”-e-l</i>	<i>po-e-l</i>
<i>čitat’</i>	‘read’	<i>pro-čita-l</i>	<i>po-čita-l</i>
<i>pisat’</i>	‘write’	<i>na-pisa-l</i>	<i>po-pisa-l</i>
<i>tancevat’</i>	‘dance’	<i>s-tanceva-l</i>	<i>po-tanceva-l</i>

Braginsky and Rothstein (2008: 19, 22–23) note that neither the presence or absence of an overt object nor quantification of the object influence the aspectual effect (definiteness is not marked on a Russian noun/NP) and that the compatibility and reading of Russian verbs with adverbials (e.g. of time span, or *postepenno* ‘gradually’) does not change between their unprefixated and prefixated forms, between transitive and intransitive verbs, nor for singular and plural objects (as in (10)).

They conclude that this “indicates that the incrementality is inherent in the verb meaning and not derived from a structure imposed on a plurality or from a complete reading dependent on the definiteness of the direct object” (Braginsky and Rothstein 2008: 19), a clear trait of aspectual head-marking.

(10) Russian (Kiparsky 1998: 7; Braginsky and Rothstein 2008: 28)

- a. *on pisa-l pis'm-a*  
 3SG.M write-PST.M.SG letter-ACC.PL  
 ‘he wrote letters’
- b. *on na-pisa-l pis'm-a*  
 3SG.M CPL-write-PST.M.SG letter-ACC.PL  
 ‘he wrote the letters’ (completely)
- c. *Ivan postepenno e-l syr / pirog-i*  
 I. gradually eat-PST.M.SG cheese / cake-PL  
 ‘Ivan gradually used to eat / was eating cheese / cakes’
- d. *Ivan postepenno s''-e-l syr / pirog-i*  
 I. gradually CPL-eat-PST.M.SG cheese / cake-PL  
 ‘Ivan gradually ate cheese / cakes’ (completely)

The phenomena mentioned in this section are head marking insofar as any dependents present do not distinguish readings of the verb, but marking on the verb controls the readings of the verb and participants. Nevertheless, Russian has some DOM which will be taken up in the next section.

## 1.4 Double marking of aspectual values

Double marking of aspectuality combines aspectual markers on the verb with features marked on overt participants, such as individuation and quantification, or adverbials that activate one of the possible lexically inherent readings according to the actional class of the verb.

In Russian, masculine mass nouns in object function to a perfective verb form are marked either with accusative case for full affectedness or with genitive for partial affectedness (11). As both the verbal head (perfective form) and the dependent (case) interact, this constitutes a niche of double marking. The distinction cannot be made with imperfective verb forms, however (Paducheva 1998); this underlines that it is the verb form which controls affectedness, not the marking of the noun.

## (11) Russian (de Hoop and Malchukov 2007: 1652–1653)

- a. *on vy-pi-l vod-y* completive reading  
 3SG.M CPL-drink-PST.M.SG water-GEN.SG  
 ‘he drank up (some) water’
- b. *on vy-pi-l vod-u* completive reading  
 3SG.M CPL-drink-PST.M.SG water-ACC.SG  
 ‘he drank up the water’
- c. *on po-pi-l vod-y* delimitative reading  
 3SG.M DEL-drink-PST.M.SG water-GEN.SG  
 ‘he drank some water (a little bit/for a while)’
- d. *on po-pi-l vod-u* delimitative reading  
 3SG.M DEL-drink-PST.M.SG water-ACC.SG  
 ‘he drank (the) water (a little bit, for a while)’
- e. \**on pi-l vod-y* imperfective form  
 3SG.M drink-PST.M.SG water-GEN.SG  
 ‘he was drinking from the water’

Whereas in Russian, a partitive form is not compatible with an imperfective, in French, it is a non-partitive marked mass noun (implying full affectedness/effectiveness) that is incompatible with an imperfective form, but fine with a perfective form. Partitive marked mass nouns, signalling partial affectedness/effectiveness, are compatible with both perfective and imperfective forms in French ((12), also in contrast to Russian). Again, it is the verb form that controls affectedness/effectiveness.

## (12) French (Mattissen 2017b)

- a. *il a bu du vin*  
 he has drunk [PFTV.PST] ART.PART wine  
 ‘he drank (from the/some) wine’
- b. *il a bu le vin*  
 he has drunk the wine  
 ‘he drank up the wine’
- c. *il buvait du vin*  
 he drink.IPFV.PST.3SG ART.PART wine  
 ‘he was drinking/used to drink (from the) wine’
- d. \**il buvait le vin*  
 he drink.IPFV.PST.3SG the wine  
 ‘he was drinking up the wine’

In the aspect language Spanish, the perfective and imperfective are inflectional forms of the verb; individuation and quantification are marked on the nominal

participant by articles and quantifiers; and full affectedness/effectiveness is again marked on the verb by the obligatory reflexive proclitic *se* acting as a completive marker. The completive marker interacts with different types of measuring dependents, including cognate objects (measuring a state of affairs by singling out one instance of it), with physical measures (of space or time), and fully effected products (Mattissen 2024: 179–180, 176), while being incompatible with non-quantified patients (cf. Nishida 1994: 431–432, 439) and with imperfective forms ((13), except when the reading is habitual; Nishida 1994: 446).

- (13) Spanish (Mattissen 2017b; Chapado Chorro and García García 1991: 66)
- a. *La semana pasada Carmen se pintó un cuadro.*  
ART week past C. REFL paint.PFTV.PST.3SG a picture  
 ‘Last week, Carmen painted a picture.’
  - b. *Se comió dos/las manzanas / una/la manzana.*  
REFL/CPL eat.PFTV.PST.3SG 2/DEF.PL apples / one/the apple  
 ‘S/he ate two/the apples / a/the apple.’
  - c. *Comió manzanas.*  
eat.PFTV.PST.3SG apples  
 ‘S/he ate apples.’
  - d. *\*Se comió manzanas.*  
REFL/CPL eat.PFTV.PST.3SG apples  
 ‘S/he ate up apples.’
  - e. *\*Comió dos manzanas / la manzana.*  
eat.PFTV.PST.3SG 2 apples / the apple  
 ‘S/he ate two apples / the apple.’

As *se* is a reflexive marker, i.e. a voice or valency marker, Spanish is able to make a partial vs. full affectedness/effectiveness distinction on verbs of consumption, production and goal attainment via valency change in a similar way to Tagalog DVM, whereas French has DOM (although not traditionally acknowledged as such) reminiscent of Finnish.

Bulgarian, similar to the Kartvelian languages, operates with completive preverbs in addition to perfective and imperfective inflection, instead. Like Georgian, Bulgarian usually prefixes the perfective forms; with verbs of consumption and production prefixing is used to distinguish the delimitative reading (without a preverb; the state of affairs ends with partial affectedness/effectiveness of the undergoer) from the completive reading (with the preverb; full affectedness/effectiveness). In Bulgarian, the perfective form of a transitive verb cannot be used without a direct object (Lindstedt 1985: 185; compare Oltenian-Romanian, Mattissen 2024: 177) and the completive preverb triggers the definite marker on the unquantified nominal

participant, i.e. verb marking and participant marking interact, there is double marking.

(14) Bulgarian (Lindstedt 1985: 185)

- a. *jadox*                      *zaxar*  
eat.PFTV.PST.1SG sugar  
'I ate some sugar'
- b. *iz-jadox*                      *zaxar-ta*  
CPL-eat.PFTV.PST.1SG sugar-DEF  
'I ate the sugar'

Tagalog, Georgian, Laz, Spanish, Bulgarian, Russian and Finnish show that in head-marking, double-marking and dependent-marking aspect systems the delimitative (without final boundary transgression) and completive readings of verbs of consumption and production can be distinguished by morphosyntactic means, which are marked on the verbal head in head-marking and double-marking systems.

In Greek, however, the perfective form together with a definite object has a completive reading without additional markers, while in the absence of an object or with a non-definite object the perfective form is neutral with respect to boundary transgression (cf. Stephany 1985: 39–40).

(15) Greek (Stephany 1985: 40; Kiparsky 1998: 263)

- a. *i*    *María*    *ðiávase*                      *tin*    *efimeríða*  
ART M.    read.PFTV.PST.3SG    ART newspaper  
'María read the newspaper.'
- b. *i*    *María*    *ðiávase*                      *efimeríða*  
ART M.    read.PFTV.PST.3SG newspaper  
'María did newspaper-reading.'
- c. *eyrapse*                      *ta*    *yrama-ta*  
write.PFTV.PST.3SG    ART letter-ACC.PL  
'S/he wrote the letters.'
- d. *eyrapse*                      *yrama-ta*  
write.PFTV.PST.3SG letter-ACC.PL  
'S/he wrote letters.'
- e. *eyrafe*                      *ta*    *yrama-ta*  
write.IPFV.PST.3SG    ART letter-ACC.PL  
'S/he was writing letters.'
- f. *eyrafe*                      *yrama-ta*  
write.IPFV.PST.3SG letter-ACC.PL  
'S/he was writing letters.'

Turning to languages with an advanced aspectoid system, English has double marking of aspectual values with a strong tendency towards dependent marking. The grammaticalised progressive and past habitual are marked on the head. Dependent marking is based on the presence, individuation and quantification of nouns (e.g. by articles) as well as on adverbials and copredicates. The readings that are lexically inherent in the actional class of the verb are activated by the properties of a participant so that it acts as an increment, measure, or telos (see (16)). It is especially striking that verbs of consumption, production and goal-attainment are generally valency-labile, i.e. can be used with (transitive) or without an object (intransitive), such as *eat (sth.)*, *draw (sth.)*.

## (16) English

delimitative reading	completive reading	
<i>I ate</i>	<i>I ate an apple</i>	
<i>I ate apples</i>	<i>I ate the apples, I ate three apples</i>	
	<i>I ate up</i>	
<i>I danced</i>	<i>I danced a dance</i>	(cognate object)
<i>I ran</i>	<i>I ran a mile</i>	(physical measure)
	<i>I ran to the park</i>	(telos)

In the absence of an overt object in the clause or when a plural noun without an article realises the object, the delimitative reading (partial affectedness) is activated. An overt object with an article or quantifier sets a telos to the state of affairs and activates the completive reading. For example, eating ends when stopped deliberately at some point (before the plate is empty), yielding a delimitative reading of *ate*, or when the food is completely consumed (boundary transgression, completive reading as in *ate an apple*). Alternatively, the completive reading can be encoded by an overt telos marker (signalling that there is no more supply), as in *eat up*. In order to be measurable, a state of affairs must have an initial and final boundary. A cognate object measures out a state of affairs by singling out one full entity of it *en bloc* (from beginning to end). A cognate object is therefore incompatible with an “intra-boundaries” reading, such as of a progressive.

(17) \**They were dancing a dance when . . .*

In Hungarian with its advanced aspectoid system, full and partial affectedness can be signalled by differential object (dependent) marking on mass nouns, which is in turn mirrored on the verb by (in)definite undergoer marking. Full affectedness/effectiveness is marked by accusative and definite article as well as definite undergoer marking on the verb (18) (Knittel and Forintos-Kosten 2002). Partial affected-

ness is encoded by relative on the noun and indefinite undergoer marking on the verb.

(18) Hungarian (Leinonen 1984: 245)

- a. *et-te*                      *a*    *sütemény-t*  
 eat-3SG.PST.DEF    the    pastry-ACC  
 ‘s/he ate the pastry’
- b. *ev-ett*                      *a*    *sütemény-ből*  
 eat-3SG.PST.INDEF    the    pastry-ELA  
 ‘s/he ate from (some of) the pastry’

Hungarian uses preverbs for the aspectual distinction (Knittel and Forintos-Kosten 2002). They interact with definiteness and indefiniteness of the object, signalled by the definite article, and (in)definite undergoer marking on the verb (19). The indefinite object must be overt, as (19e) shows. When there is a definite object marker on the verb an overt nominal can be optional.

(19) Hungarian (Knittel and Forintos-Kosten 2002: 58, 59, 63–64, 78)

- a. *Mari ette*                      *az almá-t*  
 M.    eat.3SG.PST.DEF    the    apple-ACC  
 ‘Mari was eating the apple (and may still be eating it).’
- b. *Mari meg-ette*                      *az almá-t*  
 M.    PFTV-eat.3SG.PST.DEF    the    apple-ACC  
 ‘Mari ate up the apple.’
- c. *Mari evett*                      *egy almá-t*  
 M.    eat.3SG.PST.INDEF    an    apple-ACC  
 ‘Mari ate an apple.’
- d. *Mari meg-evett*                      *egy almá-t*  
 M.    PFTV-eat.3SG.PST.INDEF    an    apple-ACC  
 ‘Mari did eat an apple indeed.’
- e. \**Mari meg-evett*  
 M.    PFTV-eat.3SG.PST.INDEF  
 ‘Mari ate.’
- f. *Mari meg-ette*  
 M.    PFTV-eat.3SG.PST.DEF  
 ‘Mari ate it.’

Both aspect and advanced aspectoid systems can be head marking or double marking and distinguish marking of partial vs. full affectedness/effectiveness as

well as delimitative and completive readings. Valency and preverbs are the most important means in addition to aspect or aspectoid inflection.

## 1.5 Stacking or conflation?

Besides matching combinations of aspect forms with aspectoids such as imperfective with progressive, aspectual values of the imperfective and the perfective realm can be combined. Typical combinations are habitual sequences of boundary transgressions (perfective in the scope of imperfective) or a temporally extended approach to the final boundary transgression (imperfective in the scope of perfective), e.g. in Bulgarian, Georgian and Spanish. However, not all aspect or advanced aspectoid languages can stack markers in this way, but resort to a conflational strategy instead, as will be shown here. Conflation will be taken up again in the context of verb/satellite framing in Sections 4.1, 4.2 and 4.4.

Turning to Bulgarian and Georgian first, perfective past (*aorist*) forms in these languages usually bear a completive preverb, whereas imperfective forms generally do not (cf. Lindstedt 1985: 169–175; Holisky 1981: 139–141). However, it is possible not to prefix a perfective form (the so-called “imperfective aorist”) in order to signal the durativity or temporal extension of a state of affairs in a sequence (20). On the other hand, prefixation of an imperfective form (so-called “perfective imperfect”) is possible for a habitual sequence of states of affairs (cf. Lindstedt 1985: 169–171, 189–190).

(20) Bulgarian (Lindstedt 1985: 175, 189)

a. *Toj caruva trijset godini.*  
 he reign.PFTV.PST.3SG 30 years  
 ‘He reigned for thirty years.’

b. *vseki pät, kogato iz-lezexme na poljana,*  
 every time when CPL-come.out.IPFV.PST.1PL on meadow  
*viždaxme. . .*  
 see.IPFV.PST.1PL . . .  
 ‘Every time we came out onto a meadow, we saw. . .’

Spanish can combine a perfective past form (*indefinido*) and a progressive (*estar + gerund*) as in (21a) (cf. Chapado Chorro and García García 1991: 64–65; Breu 1994: 38) for the expression of a temporal extension of a completed state of affairs. In addition, Spanish can express a temporally stretched approach to the final boundary transgression or immediacy of the final boundary transgression with a progressive imperfective past form combined with the completive marker *se* and a patient

marked as fully affected by the definite article. Example (21b) expresses taking the last forkful of the meal to the mouth. The imperfective is used because the final stage of eating constitutes the backgrounded state of affairs in the incidence taxis constellation. The intruding state of affairs (in the perfective past) is that a shot is heard.

(21) Spanish (Sanz 2000: 87)

- a. *Mi hermano estuvo leyendo un libro.*  
 my brother be.PFTV.PST.3SG read.GER a book  
 ‘My brother was reading a book.’
- b. *Se estaba comiendo la carne cuando oyó un disparo.*  
 REFL/CPL be.IPFV.PST.3SG eat.GER the meat when  
 hear.PFTV.PST.3SG a shot  
 ‘S/he was just eating up the meat dish when s/he heard a shot.’

In Russian, such a temporally stretched approach to the final boundary transgression can be encoded by the so-called secondary imperfective of (usually prefixed) terminative verbs (depending on the inherent actional class), which can also have a habitual or conative (but no progressive) reading (Isačenko 1962: 366).

(22) Russian (Isačenko 1962: 396; Batiukova 2015: 33)

- do-pit’* ‘drink up’ *dopivat’* ‘to be drinking up’  
*do-pisa-t’* ‘write up’ *do-pis-yva-t’* ‘to be finishing writing’

In polysynthetic Nivkh with its advanced aspectoid system (cf. Mattissen 2003b: 16–18, 2024; Gruzdeva 2012) the completive and the progressive suffix can be stacked on the verb form to encode the stretched approach to the final boundary transgression, e.g. in (23).

(23) Nivkh (Mattissen 2012)

- ra-γət-ivu-ŋan...*  
 drink-CPL-PROG-CV  
 ‘when almost having finished drinking it . . .’

Similar stacking is possible in the polysynthetic aspect language Ngan’gityemerri (Reid 2011: 174–177).

In the aspect languages French and Italian, however, such a stacking is not possible (cf. Bertinetto and Delfitto 2000: 197; Breu 1994: 39). The temporally stretched approach to the final boundary transgression cannot be encoded by combining

an imperfective past form (*imparfait*, *imperfetto*) or past progressive form with a participant marked for full affectedness (as in (24a–c)). Instead, a different verb featuring an inherent final boundary with a preceding conative phase (gradual terminative actional class, Mattissen 2024: Ch. 2.6.1) has to be used in its imperfective past (progressive) form, viz. *finir* and *finire* ‘finish’ (as in (24d, e)).

## (24) French (Mattissen 2017b)

- a. \**il buvait le vin quand...*  
 he drink.IPFV.PST.3SG the wine when  
 ‘he was drinking up the wine when ...’
- b. \**il était en train de boire le vin quand...*  
 he PROG(he.IPFV.PST).3SG drink the wine when  
 ‘he was drinking up the wine when ...’
- c. \**il fut en train de boire le vin quand...*  
 he PROG(he.PFTV.PST).3SG drink the wine when  
 ‘he was drinking up the wine when ...’
- d. *il finissait le vin quand...*  
 he end.IPFV.PST.3SG the wine when  
 ‘he was drinking up the wine when ...’
- e. Italian (Haßler 2016: 256)  
*Francesca stava giusto finendo di scrivere una*  
 F. he.IPFV.PST.3SG just finish.GER of write a  
*lettera, ...*  
 letter  
 ‘Francesca was just finishing writing a letter, ...’

The verb *finir* then conflates the manner of consumption or production, e.g. *manger* ‘eat’ or *boire* ‘drink’ etc., or has the manner verb as its satellite as Italian *finire*. Conflation of manner reminds of the lexicalisation patterns of path and manner (Talmy 1985, 2000; Slobin 2004), as verb framing construes manner, if at all, as a satellite; one reason for searching for a correlation between aspectual marking and verb/satellite framing in Section 4.

In sum, preverbs are a strategy of distinguishing delimitative and completive readings with verbs of consumption, production and goal-attainment as well as of expressing a stretched approach to final boundary transgression. Languages that do not operate with preverbs may use other forms of stacking, e.g. combinations of aspectoids or of aspect and aspectoid markers. Stacking is possible in both aspectually head-marking and double-marking systems and in aspect and advanced aspectoid systems, especially polysynthetic ones, but not a necessary feature.

Aspect languages that do not allow stacking are, for instance, French and Italian. These languages resort to conflating manner of action by a general boundary transgression verb.

Manner conflation can be seen as a contrasting strategy to differential lexicalisation, to be discussed in the next section.

## 2 Differential lexicalisation

Differential lexicalisation is understood here in the sense of a finer semantic granularity featured by verbs with respect to valence and actional class. A language with differential lexicalisation tends to have more specific verbs.

With respect to valency, differential lexicalisation means that verbs are not labile, but in all their uses are either atransitive or intransitive or monotransitive or ditransitive etc., with fixed semantic roles of their participants. Further arguments may not be added without overt derivation, and typically, separate verbal lexemes exist for affine concepts with different valencies, in a derivational or suppletive relation.

Laz, a polypersonal aspect language, is a pertinent example. All participants are encoded on the verb by inflection and version (valency) vowel and there are separate verbs formed from the same root by different conjugation class markers, version vowels, preverbs and voice markers (such “building-blocks” are separated by a dot in (25)).

(25) Laz (Mattissen 2003a: 265–266)

intransitive		transitive	
<i>b.i.ncir.am</i>	‘I go to bed/to sleep’	<i>b.o.ncir.am</i>	‘I bring s.o. to bed’
<i>b.i.bir.am</i>	‘I play’	<i>b.u.bir.am</i>	‘I play with sth. of s.o.’

Likewise, Nivkh (with an advanced aspectoid system) clearly distinguishes transitive and intransitive verbs

(26) Nivkh (Mattissen 2003b: 29)

intransitive verb		transitive verb	
<i>ʔo-ɖ</i>	‘sth. bends’	<i>zo-ɖ</i>	‘s.o. bends sth.’
<i>oz-ɖ</i>	‘s.o. rises’	<i>j-ozu-ɖ</i>	‘s.o. wakes s.o. up’

Ngan’gityemerri distinguishes single and multiple undergoer verbs.

## (27) Ngan'gityemmerri (Reid 2011: 187)

single undergoer		multiple undergoers (iteration)
- <i>tyat</i> -	'place one thing in a position'	- <i>tyuk</i> - 'place multiple things in a position'
- <i>ket</i> -	'chop'	- <i>gerrgirr</i> - 'chop into multiple parts'

Compare this to languages like English (with an advanced aspectoid system) with its broad range of labile verbs, such as *eat (sth.)*, *draw (sth.)*, *sing (sth.)*.

Differential lexicalisation also yields semantically affine verbs that belong to different actional classes (cf. also Sasse 2000: Ch. 5.3.1, 5.3.2). For motion verbs, for instance, this means that they are fixed as to whether goal-attainment is lexically inherent or not, for telic verbs, whether the resultant state is lexically inherent or not. Compare, for instance, Cayuga (cf. Sasse 2000: 211, 223) and English.

## (28) Cayuga (Sasse 2000: 237, 232)

terminative		aterminative
- <i>ya'k</i> -	'cut' ('sever')	- <i>hren</i> - 'cut' ('incise')
- <i>k</i> -	'eat sth., eat sth. up'	- <i>atekhoni</i> - 'eat' ('have a meal', intr.)

In Laz, derivationally related n-tuples pervade the system of verbal lexemes.

## (29) Laz (Mattissen 1994–2000, 2003a: 265–266)

aterminative		terminative
<i>mcar</i>	'I sleep'	<i>bi.ncir.am</i> 'I go to bed / go to sleep'
<i>p.xe.r</i>	'I sit'	<i>dop.xed.ur</i> 'I sit down'
<i>bi.gza.r</i>	'I walk'	<i>bi.gzal.am</i> 'I depart'
<i>ma.zir.en</i>	'I see sth.'	<i>b.zir.am</i> 'I find sth.'

Its sister language Georgian derives intransitive aterminative verbs from manner-of-motion-verbs (Holisky 1981: 73–78).

## (30) Georgian (Holisky 1981: 74–75)

terminative		aterminative
<i>goravs</i>	'roll'	<i>goraobs</i> 'roll around'
<i>curavs</i>	'swim'	<i>curaobs</i> 'swim around'
<i>xṭunavs</i>	'jump'	<i>xṭunaobs</i> 'jump around'

In addition, the preverb *da-* derives non-goal-attaining path verbs with meanings such as “motion back and forth, here and there, frequent or regular coming and

going” (Holisky 1981: 78–79). The non-goal-attaining verbs are neither compatible with a goal phrase nor a direct object and they do not take completive preverbs (Holisky 1981: 76–78). On the other hand, ingressive derivations of manner-of-motion verbs add a path component and always have a preverb in their perfective forms (Holisky 1981: 78, 42).

(31) Georgian (Holisky 1981: 78)

- |                       |                          |
|-----------------------|--------------------------|
| <i>ga-gor-d-eba</i>   | ‘roll out’               |
| <i>šemo-gor-d-eba</i> | ‘roll in’                |
| <i>mi-gor-d-eba</i>   | ‘roll away from speaker’ |

Himmelman (2008: Ch. 4.2) describes what classifies as differential lexicalisation for Tagalog. For instance, dynamic and stative verbs derived from the same root belong to different paradigms.

(32) Tagalog (Panganiban 1972)

- |                 |                         |                 |                          |
|-----------------|-------------------------|-----------------|--------------------------|
| stative         |                         | dynamic         |                          |
| <i>ma-buo</i>   | ‘become whole/complete’ | <i>buu-in</i>   | ‘to complete/make whole’ |
| <i>ma-lalim</i> | ‘deep’                  | <i>lalim-an</i> | ‘make deep’              |
| <i>ma-itim</i>  | ‘black’                 | <i>itim-an</i>  | ‘wear black clothes’     |

In Japanese (advanced aspectoid system), differential lexicalisation tends to yield both valence and actional class distinction (cf. Jacobsen 1982: 92).

(33) Japanese (Jacobsen 1982: 93)

- |    |   |                       |
|----|---|-----------------------|
| a. | <i>Tokei-o</i>  | <i>naoshi-te iru.</i> |
|    | clock-ACC   | repair-CV be.IND      |
|    | ‘Someone is fixing the clock.’ ( <i>naosu</i> ‘repair’ exoactive, progressive form) |                       |
| b. | <i>Tokei-wa</i>   | <i>naot-te iru.</i>   |
|    | clock-TOP   | repair-CV be.IND      |
|    | ‘The clock has been fixed.’ ( <i>naoru</i> ‘repair’ endoactive, resultative form)   |                       |

In Maltese differential lexicalisation can be observed in the grammaticalisation of specialised continuatives from several different verbs (Stolz and Ammann 2008: 175–175) as well as in the grammaticalisation of verbs of beginning (Stolz and Ammann 2007: 150), which could be in line with the existence of several types of inceptive verbs in Semitic languages (cf. Mattissen 2024: 162).

Results are lexically inherent in verbs in Greek and Romance languages and therefore not overtly expressed (cf. Horrocks and Stavrou 2003) as in (34a–d), but not in English or Finnish as shown in (35a–b), where the result is encoded verb-externally.

- (34) a. Greek (Giannakidou and Merchant 1999: 123)  
*O Janis skupise to piato tu (\*katharo).*  
 ART J. wipe.PFTV.PST.3SG ART plate his clean  
 ‘Janis wiped his plate clean.’
- b. French (Clark 2005: 176)  
*Il a essuyé la table (\*propre).*  
 he has wiped the table clean  
 ‘He wiped the table clean.’ (grammatical as ‘wiped the clean table’)
- c. Spanish (Sanz 2000: 101)  
*El río se congeló (\*solido).*  
 the river REFL/CPL froze solid  
 ‘The river froze solid.’
- d. Spanish (Sanz 2000: 102)  
*Juan limpió la mesa (\*impeccable).*  
 J. cleaned the table flawless  
 ‘Juan wiped the table clean.’
- (35) English
- a. *wipe sth. clean, freeze solid, hammer sth. flat, sing s.o. asleep*
- b. Finnish (Heinämäki 1984: 157)  
*metsästäjä ampui lehmän kuoliaa-ksi*  
 hunter shoot.PST.3SG cow.ACC dead-TRNSL  
 ‘the hunter shot the cow dead’

Both English and Finnish are languages that rely heavily on dependent marking of aspectual values and therefore can afford verbs that allow a broader range of readings. The possible readings are then activated by participants, adverbials and copredicates.

Differential lexicalisation comes in handy when nominal participants are not marked for individuation or are not obligatory in a language, which is the case in aspectual head-marking systems. Differential lexicalisation is independent of polypersonalism, however, as apersonal Tagalog and monopersonal Russian show. On the other hand, both Tagalog and Nivkh mark the undergoer on the predicate,

the former by voice, the latter by person marking; Japanese has a more complex system of undergoer encoding (Mattissen 2015). Thus, it is languages with undergoer marking and polypersonal languages in my sample that have non-valency-labile verbs, are aspectually head marking and prefer differential lexicalisation with respect to actional class.

In order to build a bridge to verb/satellite framing, we will now turn to verbs of goal attainment and analyse their properties with respect to head/dependent marking, full vs. partial affectedness, differential lexicalisation, DVM and DOM.

### 3 Goal-attainment

Besides verbs of consumption and production, a third ontological group of verbs which frequently belong to the same actional class (successive-terminative verbs, see Mattissen 2024: 173–175) are directed motion verbs with a lexically inherent final boundary that is transgressed when the goal is attained (completive reading). They are called goal-attainment verbs here. Nevertheless, goal-attainment verbs also have a lexically inherent dynamic phase of motion leading towards the goal, and that motion can end at any point without attaining the goal (progressive or delimitative readings; cf. Mattissen 2024: 183–184). Motion verbs tend to be differentially lexicalised with respect to goal-attainment in aspectual head-marking languages, and this differential lexicalisation is all the more relevant in the context of goal prominence (cf. Himmelmann and Primus 2015).

Among the head-marking aspect languages, Laz distinguishes between goal-attainment and non-goal-attainment motion verbs by differential lexicalisation. Goal-attainment verbs are only compatible with adverbials in the motative, a case that signals motion to a goal or from a source; non-goal-attainment verbs are not compatible with adverbials in the motative, but only with case-unmarked local nouns. Compare this with English dependent marking (in the right column of (36)), where the delimitative and completive readings of the same less specific motion verb (*walk*) are activated by adverbials encoding a direction or space on the one hand (36f, g) and a goal attained (36e) on the other.

- (36) Laz (Mattissen 2003a: 264) English
- goal-attaining**
- a. *noya-şa bulur* e. *I walk to the market.*  
 market-MOT go.1SG.PRS  
 'I go to the market'
- non-goal-attaining**
- b. *noya go-bulur* f. *I walk around the market.*  
 market PRV-go.1SG.PRS  
 'I go around the market'  
 counter check
- c. *\*noya bulur*  
 market go.1SG.PRS  
 'I go to the market'
- d. *\*noya-şa go-bulur* g. *I walk in the direction of the market.*  
 market-MOT PRV-go.1SG.PRS  
 '\*I go around to the market'

In Russian, a whole range of motion verbs exhibit differential lexicalisation with respect to goal-attainment (cf. Isačenko 1962: 419–424).

- (37) Russian (Isačenko 1962: 423–424)
- |                           |                                       |         |
|---------------------------|---------------------------------------|---------|
| inherently goal-attaining | inherently non-directed or round trip |         |
| <i>idti</i>               | <i>xodit'</i>                         | 'go'    |
| <i>bežat'</i>             | <i>begat'</i>                         | 'run'   |
| <i>nesti</i>              | <i>nosit'</i>                         | 'carry' |
| <i>letet'</i>             | <i>letat'</i>                         | 'fly'   |
| <i>plyt'</i>              | <i>plavat'</i>                        | 'swim'  |

Ket, a polysynthetic language with a head-marked advanced aspectoid system, has an even finer-grained differential lexicalisation with roots for motion away (38a), goal incorporation (38b), atterminative motion (38c), round trip (38d–g), and specific path verbs oriented at the local system (38h) (cf. Werner 1997: 233–236, 206–210, 227–250; Drossard 2002: 241). The round trip verbs can incorporate goals, too.

(38) Ket (Werner 1997: 200, 233, 235; Drossard 2002: 241)

- a. *ɔyatn'* 'he goes away'
- b. *d-igd-ajdaq* 'he goes into the woods' (*-igda-* lexical affix 'wood')
- c. *t-tajga* 'he goes around'
- d. *du-jaq* 'he runs there and back again for some days/weeks'
- e. *d-igda-ksaq* 'he runs to the riverbank and back again for some days/weeks'
- f. *d-butseq* 'he runs there and back again for some hours'
- g. *d-igda-butseq* 'he runs to the riverbank and back again for some hours'
- h. *d-aRar-o-l-vet*  
1SG.S-go.from.wood.to.river-PST-PERF-do  
'I went from the woods (Taiga) to the river'

Nivkh, polysynthetic with a head-marked advanced aspectoid system as well, has a whole system of very specific path verbs oriented at its locational system, among them goal-attaining and non-goal-attaining verbs, intransitive and transitive ones (Krejnovič 1960; Mattissen 2003b: 26–27, 138, 2006: 308).

(39) Nivkh (intransitive) (Krejnovič 1960: 80)

- qaɖ* 'go downriver'
- tuɖ* 'go upriver'
- qoɖ* 'go into the water, go across the river'
- mayɖ* 'moor on the river bank, go from river bank into the interior'
- məɖ* 'go down from the mountain, down to the river bank'
- mərɖ* 'go up (a mountain)'

On the other hand, in dependent-marking Finnish, as in double-marking English, a general motion verb is compatible with both goal-attaining and non-goal-attaining adverbials.

(40) Finnish

- a. <https://amara-malik.com/fi/issues/16011-10-examples-of-dialogue-paragraphs>  
... *ja meni saare-lle*  
and go.PST.3SG island-ALL  
'... and went to the island'
- b. <https://www.ess.fi/uutissuomalainen/3132187>  
*monenlaista muuta väkeä meni edestakaisin*  
all.sorts.of other public go.PST.3SG back&forth  
'all sorts of other people went back and forth'

In advanced aspectoid and double-marking Hungarian, goal-attainment is signalled by the presence and position of a preverb on a motion verb while a participant is not marked for goal-attainment (head marking). Note the expression of the stretched approach to final boundary transgression in (41c), with the completive preverb in postverbal position (with a progressive reading in the incidence constellation, Knittel and Forintos-Kosten 2002: 53–54).

## (41) Hungarian (de Groot 1984: 135)

- a. *a fiú ment a szobá-ba*  
 the boy go.3SG.PST.INDEF the room-ILL  
 ‘the boy was going toward the room’
- b. *a fiú be-ment a szobá-ba*  
 the boy PFTV-go.3SG.PST.INDEF the room-ILL  
 ‘the boy entered the room’
- c. *a fiú ment be a szobá-ba amikor...*  
 the boy go.3SG.PST.INDEF PFTV the room-ILL when  
 ‘the boy was entering the room when . . .’

Head-marking Tagalog does not rely on differential lexicalisation for goal-attainment, but, again, on voice. Relevant manner-of-motion verbs are compatible with at least two voices, *-in* signals goal-attainment, i.e. boundary transgression, whereas *-an* signals non-attainment (cf. Latrouite 2011: 159–161; see Section 1.3). The pivotal *ang*-phrase is identical in both versions and does not signal goal-attainment on the dependent.

## (42) Tagalog (Latrouite 2011: 160)

- a. *akyat-in mo ang puno.*  
 go\_up-EV 2SG.GEN NOM tree  
 ‘you climb the tree’ (on top)
- b. *akyat-an mo ang puno.*  
 go\_up-SV 2SG.GEN NOM tree  
 ‘you go up on the tree’
- c. *Lakar-in mo ang Luneta.*  
 walk-EV 2SG.GEN NOM Luneta  
 ‘Walk up to Luneta.’
- d. *Lakar-an mo ang Luneta.*  
 walk-SV 2SG.GEN NOM Luneta  
 ‘Walk in Luneta.’

Signalling of goal-attainment follows the same strategies as signalling of full or partial affectedness and effectedness, another reason for classifying verbs of consumption, production and goal-attainment as one actional class (cf. Mattissen 2024: 169, 185).

Further on, goal-attainment constitutes a relation between aspect/aspectuality and the lexicalisation patterns of verb and satellite framing. Therefore any possible correlations between head/dependent marking of aspectual values and verb/satellite framing will be discussed in the next section.

## 4 Aspect and verb/satellite/serial framing

Goal attainment is related to the way direction of motion is encoded in a language. The most influential approach in this respect are Talmy's (1985, 2000) and Slobin's (2004) verb-framing and satellite-framing categorisation, even if it has been criticised in many ways (cf. Flecken et al. 2014: 49; Alhamdan et al. 2018 for an overview). Basically, verb-framing languages prefer encoding the direction of motion ("path") in the finite verb if a goal is expressed in the clause, whereas the manner of motion may be encoded as a dependent verb form, the satellite. Satellite-framed languages encode the manner of motion in the finite verb together with a goal expression, and the direction of motion by a satellite, e.g. a preverb (as in German or Slavic languages) or a particle ("postverb", as in English). Further possibilities of encoding are the use of both schemes, serialisation of manner and path (cf. Ameka and Essegbey 2013), as in Chinese (with serial verbs) or Ngan'gityemerri (in a polysynthetic verb form), or completely different structures.

(43) Ngan'gityemerri (Reid 2011: 240)

*ninymunggurr wanniny-pap*  
 escarpment 3PL.S.go.PERF-climb  
 'they climbed up onto the escarpment'

Flecken et al. (2014: 53, 71) studied attendance to possible endpoints, both verbally and in eye-tracking experiments using pictures of entities in motion underway to a goal visible afar, and found a correlation to aspect languages (e.g. Arabic, without attention to possible endpoints) and aspectless languages (e.g. German, with attention to possible endpoints), but no clearcut preferences in correlation to path and manner framing (verb and satellite-framed) languages. Motivated by their findings, this section explores whether there are correlations between the framing types and

the way aspectual values are encoded, with a special look at conflation of directional and local cases or adpositions.

#### 4.1 Aspectual head marking and verb/satellite/serial framing

On the basis of the situation in Greek, Horrocks and Stavrou (2007: 625–626, 633) draw the conclusion that (formulated in the terminology used here) in languages with a grammaticalised binary aspect category which is encoded by different perfective and imperfective stems (not preverbs) in a verbal paradigm, verbs are either telic or atelic, i.e. in the terms used here, verbs are differentially lexicalised as to actional class. As a consequence, non-goal-attainment verbs are not compatible with goal PPs. Modern Greek is aspectually head-marking, verb-framed and combines goal phrases with path verbs only. Its prepositions are used for location and goal-marking indiscriminately, e.g. *se* ‘in(to)’ (44), i.e. only the verbal semantics determine the reading (Horrocks and Stavrou 2007: 620). Therefore, a locational phrase modifying a manner-of-motion verb is interpreted as location, not as direction (Horrocks and Stavrou 2007: 611, 616).

- (44) Greek (Horrocks and Stavrou 2007: 611, 617)
- a. *O Janis pije sto parko.*  
 DEF J. go.PFTV.PST.3SG in.the park  
 ‘Janis went into the park’
- b. *O Orestis kolibise sto nisi.*  
 ART O. swim.PFTV.PST.3SG in.the island  
 ‘Orestes swam on/near the island’ \*‘to the island’

Reformulated in the terms of the present approach the reason for this would be that goal-attainment is lexically inherent in the verb by differential lexicalisation and adverbials are not needed in order to activate that reading. Ancient Greek, on the contrary, used to be satellite-framed, could use manner-of-movement verbs with goal PPs (Horrocks and Stavrou 2007: 613) and had more prepositions that, in addition, distinguished between location and goal marking (Horrocks and Stavrou 2007: 619–620).

In our sample, further languages fit into the scenario set up by Horrocks and Stavrou (2007). Tagalog is aspectually head-marking with reduplication marking the imperfective. It has less specific cases or adpositions, in particular a multifunctional phrase marker *sa* that conflates several peripheral relations, including path, goal, location and even source. The language shows both verb and satellite framedness (see (42) and (45)).

## (45) Tagalog (Latrouite 2011: 119)

- a. *P<um>asok      ako      sa      bahay.*  
 <AV>[REAL]enter 1SG.NOM DAT house  
 'I entered the house.'
- b. *L<um>abas      ako      sa      bahay.*  
 <AV>[REAL]leave 1SG.NOM DAT house  
 'I left the house.'
- c. *Ilang beses ko      rin      sinubukan na      lakar-in      ang*  
 many times 1SG.GEN also try(SV) CMPL walk-EV NOM  
*kagubatan*  
 forest  
*at      ma-abot      ko      ang      hangganan      nito.*  
 and POT-reach(EV) 1SG.GEN NOM border its  
 'Many times I tried to walk through the forest and actually reached its border.'

In Tagalog, however, a location or goal may also occupy the pivotal *ang*-phrase while being reflected on the predicate by a corresponding undergoer voice affix as in (45c); cf. Latrouite (2011: 11–12). As Tagalog verbs are not generally compatible with all of the actor and undergoer voices but with their individual choice (Latrouite 2011: 39), not just any verb may promote a location and/or goal into the *ang*-phrase.

The aspectually head-marking and serial-framed language Ngan'gityemerri (Reid 2011: 160, 178) is of the composite-stem layout type of polysynthesis (Mattissen 2017a: 90) and has only a few local postpositions and case enclitics (Reid 2011: Ch. 6). It differentiates between perfective and imperfective light verbs, derives atelic verbs by reduplication and uses serialised posture and motion auxiliaries to denote the progressive (Reid 2011: 160, 180, 183), thus aspectual and framing markers are morphologically entangled.

## (46) Ngan'gityemerri (Reid 2011: 183, 184)

- a. *yerim-fityi-pefi-yaganim*  
 2SG.S.hands.PRS-roll-THITHER-2SG.S.go.PRS  
 'you can drive it [the car] away'
- b. *a-yaga menyirr nganimuy-nide*  
 A-DEM sand loose-LOC  
*demem-dundem-yenim*  
 3SG.S.hands.DTR.PRS-bury.RED-3SG.S.go.PRS  
 'that one who always buries himself in the loose sand'

Ngan'gityemerri, as an aspect language, can incorporate goals and sources (the same is true of Nivkh, see below), so there is a structure that does without local cases or adpositions, as in Tagalog.

- (47) Ngan'gityemerri (Reid 2011: 190)
- a. *dangim-firr-pawal*  
3SG.S.Poke.PERF-foot-spear  
'he speared him in the foot'
  - b. *mimenem ngerrminy-ba-ket*  
billygoat.plum 1PEX.S.Hands.PERF-arm-pick  
'we picked plums from the branches'

Cayuga is aspectually head-marking, too, with perfective causative stems and imperfective semireflexive stems, but also polysynthetic of the holophrastic type (Mattissen 2017a: 90), which is equally responsible for its lack of case or adpositions. Cayuga has a dislocative (Sasse 2000: 237) in an originally serial-framing strategy.

In sum, the conflation of direction and location in case/adpositions is borne out to correlate with head marking of aspectual values, not with the framing type.

However, the counter-check yields a more differentiated picture. On the one hand, head-marking and satellite-framed Laz has case marking only for genitive and motative, with the latter case conflating direction and source, and no marking on locational phrases (see also (36)).

- (48) Laz (Mattissen 1994–2000)
- |                 |                    |                |              |
|-----------------|--------------------|----------------|--------------|
| <i>oxori-ša</i> | <i>gamo-bulur,</i> | <i>noya-ša</i> | <i>bulur</i> |
| house-MOT       | out-go.1SG.PRS     | market-MOT     | go.1SG.PRS   |
- 'I leave the house, go to the market'

Russian, on the other hand, is also aspectually head-marking and satellite-framed like the Kartvelian languages Laz and Georgian, but has a considerable inventory of prepositions for either local or directional relations, with only three prepositions (*v* 'in(to)', *na* 'on(to)', *pod* 'under') that distinguish location and direction via case government (prepositional or instrumental case for location, accusative for direction). The difference between Laz and Russian, however, is that in Russian the prepositions are formally identical to aspectual and derivational preverbs in a grammaticalisation relation whereas in Laz, there are no adpositions identical in form to preverbs.

So, among the aspect languages those having preverbs with an aspectual function are also satellite-framed; if there are formally identical adpositions, case/adposition inventories are large (e.g. Slavic languages). In Greek, verbal prefixes are not involved in aspect encoding, in Laz, preverbs are not formally identical to adpositions.

Turning to languages with an advanced aspectoid system, in aspectually head-marking and verb-framing Japanese a directional phrase can only be combined with a manner-of-motion verb in the clause when the latter is in a converb construction with a path verb as the predicate.

(49) Japanese (Horrocks and Stavrou 2007: 610)

\**kishi=e oyo-da* → *kishi=e oyo-de it-ta*  
 shore=to swim-PST shore=to swim-CV go-PST  
 ‘s/he swam to the shore’

Japanese distinguishes locational and goal-attaining postpositions/case markers, with only *ni* being polyfunctional (cf. Ikegami 1982: 123).

Head-marking, verb-framed and polysynthetic Nivkh uses converb constructions similar to Japanese.

(50) Nivkh (Panfilov 1965: 88; Beffa 1982: 87)

a. *tə-eri maŋ-gu-r ver-la-ʁa ŋi řəm-t*  
 this-river strong-CST-CV:3SG wide-PERM-CV:if 1SG swim-CV:1SG  
*tozə-jiki-nə-ǰ-ra*  
 cross-cannot-FUT-IND/NML-HILI

‘If this river is very wide I will not be able to swim across.’

b. *tʰu-in ʁoŋǰi-ror vəyi-ř məy-ǰ.*  
 sledge-LOC load-ACV:3SG drag-CV:3SG descend-IND/NML

‘He loaded it onto the sled and dragged it downhill (dragging it along, descended).’

Nivkh has five local cases and, in addition, incorporates relational morphemes (“postpositions”) as well as location, goal or source into transitive verbs according to a secundative scheme (Mattissen 2003b: Ch. 4.2.2–4.2.2.2).

- (51) Nivkh (Savel'eva and Taksami 1970: 253; Panfilov 1965: 46–47; Otaina 1978: 34)
- a. *p'-rəu-oblə*      *parta-řiv-đ*  
REFL-teach-child desk-sit(down)-IND/NML  
'the pupil sits at a desk'
  - b. *ʃ'o-ŋəŋ-ñivx*      *k'e toqo-xro-đ*  
fish-hunt-person net fence-hang-IND/NML  
'the fisherman hung the net over the fence'
  - c. *objezdʃ'ik*      *k'e atak-asqam-đ*  
bay\_watcher net grandfather-take\_away-IND/NML  
'the bay watcher took the net away from grandfather'  
(lit. 'away-took grandfather the net')

As both Japanese and Nivkh do not have a binary aspect system, they still fit into Horrocks and Stavrou's (2007) scenario with their differentiation of locational and directional adpositions.

## 4.2 Aspectual double marking and verb/satellite/serial framing

Aspectual double-marking co-occurs with both verb framing and satellite framing.

In verb-framed and aspectually stacking Spanish, manner-of-motion verbs are not compatible with goal-attainment PPs (cf. Talmy 1985, 2000). Spanish distinguishes the prepositions *a* 'to' and *hasta* 'up to', encoding mostly goal-attainment with path verbs, *hacia* 'in the direction of' that encodes a direction without goal-attainment and can be used with manner-of-motion verbs (52), and *en* 'in' for local relations.

- (52) Spanish (Mattissen 2017b)
- a. *ir a la isla*      vs.      *\*nadar a la isla*  
'go to the island'      'swim to the island'
  - b. *nadar en la playa*  
'swim at the beach'
  - c. *nadar hacia la isla*  
'swim towards the island'
  - d. (Sanz 2000: 105)  
*Fue a la costa nadando.*  
go.PFTV.PST.3SG to the coast swimming  
'S/he swam to the coast.'

Verb-framed, aspectually double-marking and non-stacking French, however, has a large inventory of prepositions, e.g. *à*, *en* ‘to, in’, *dans* ‘in(to)’, *devant* ‘in front of’, *chez* ‘at, to’, that encode both directional and local relations. So, French is “highly conflating”, concerning manner of consumption (instead of stacking, see Section 1.5), manner of motion (verb framing) and the distinction of direction and location in prepositions.

Thus for aspectually double-marking aspect languages, there is no correlation to case/adposition inventories.

Turning to advanced aspectoid systems, double-marking and satellite-framed English is well known for its manner-of-motion verbs being compatible with both goal-attainment and non-goal-attainment adverbials; the degree of conflation in prepositions differs with the variety.

(53) English

non-goal-attaining			goal-attainment
<i>walk</i> , <i>walk in the park</i> ,			
<i>walk in the direction of the park</i>	vs.		<i>walk to the park</i>
<i>swim</i>			<i>swim to the island</i>
<i>swim in the direction of the island</i>			
<i>swim around the island</i>			
<i>swim near the island</i>			

Double-marking and satellite-framed Hungarian is known for its extensive differentiation of locational and directional cases. So again, there is no correlation with case/adposition inventories, but a strong correlation of aspectually relevant pre-verbs (and English particles) and satellite framing.

### 4.3 Aspectual dependent marking and verb/satellite/serial framing

Aspectual dependent marking can arguably only describe the Finnish system. As Finnish is satellite-framed, there are no aspectual dependent-marking and verb-framed nor serial-framed languages to my knowledge. Indeed, aspectual dependent-marking and verb-framing do not match in the sense that dependent marking relies on satellites for the activation of the verbal readings, among them direction and goal attainment, whereas direction is expressed by the head verb in verb-framed languages.

- (54) Finnish (<https://www.helsinginuutiset.fi/paikalliset/1699716>)  
*Kadonnut hevonen oli ilmeisesti uinut saare-lle.*  
 lost horse was apparently swim.PTC island-ALL  
 ‘The lost horse had apparently swum to the island.’

In distinction to Laz and Russian, Finnish (as English) motion and transfer verbs are not differentially lexicalised and thus allow their phase and boundary readings to be activated by participants and adverbials. In addition, path verbs like *mennä* ‘go’ (cf. (40)), *kuljettaa* ‘carry’ (55a, b) and manner-of-motion verbs (54, 55c, d) can be used with both adverbials encoding a goal and adverbials cancelling the goal-attainment, e.g. by encoding a round trip (‘back and forth’).

- (55) Finnish
- a. (<https://launokorpi.com/fi/matka/kiertomatka-viron-saarilla-muhu-saarenmaa-hiidenmaa-vormsi>)  
*... joka kuljettaa meidät Muhun saare-lle*  
 which carry 1PL.ACC M. island-ALL  
 ‘... who takes us to Muhu Island’
- b. (<https://rtech.fi/rtech/yritys/blogi/miten-aloittaa-tyomatkapyyoraily/>)  
*... ettei niitä tarvitse joka päivä kuljettaa*  
 that.not 3PL.PART need.2SG every day carry  
*edestakaisin.*  
 back&forth  
 ‘so they don’t have to be carried back and forth every day’
- c. (Heinämäki 1984: 160)  
*Tiina heitti keihästä*  
 T. throw.PST.3SG javelin.PART  
 ‘Tiina threw the javelin.’
- d. *Tiina heitti keihään metsään*  
 T. throw.PST.3SG javelin.ACC forest.ILL  
 ‘Tiina threw the javelin into the forest.’

Manner-of-motion verbs are also compatible with physical measures either activating a goal-attainment (completive) reading or a non-goal-attaining progressive/delimitative reading.

(56) Finnish (Heinämäki 1984: 174)

- a. *Kirsi ui kilometri-n*  
 K. swim.PST.3SG km-ACC  
 'Kirsi swam a kilometer'
- b. *Kirsi ui vielä kilometri-ä*  
 K. swim.PST.3SG still km-PART  
 'Kirsi is still swimming a kilometer'

In sum, Finnish constitutes a balanced system in which aspectual values, goal-attainment and path are encoded by dependents, and dependents are marked by a rich inventory of strictly distinguished locational and directional cases. This is also in line with Horrocks and Stavrou's (2007) prediction. Bridgen (1984: 197) formulates a relation between aspectual dependent marking and satellite framing:

There may be some connection between the principally affected argument and the application of a direction satellite in Finnish. Predicates of movement often demand the addition of a direction satellite if the goal is the principally affected argument in the accusative case [ . . . ].

#### 4.4 Are there correlations?

Among the features analysed here, it is first and foremost polypersonalism or undergoer marking on the predicate that goes hand in hand with aspectual head marking, but this bears no relation to verb/satellite/serial framing. Polypersonal languages (Cayuga, Ngan'gityemerri, Ket) and languages with undergoer marking on the predicate in the form of voice or person marking (Tagalog, Nivkh) do not need overt encoding of nominal participants. The same is true of Japanese with differential lexicalisation as to valency and syntactic encoding of the person constellation (Mattissen 2015). These languages have non-valency-labile verbs and prefer differential lexicalisation with respect to actional class.

As the framing types are more of a preference or a tendency of lexicalisation patterns, there are languages that use both patterns, serialise path and manner or cannot be described in that framework, as well as languages that have local valency, incorporation or pivot promotion of local elements, e.g. Tagalog, Ngan'gityemerri, Ket and Nivkh. Keeping this in mind, in the present sample there are (advanced aspectoid systems underlined)

- aspectually head-marking and verb-framed languages (Nivkh, Japanese, Turkish),
- aspectually head-marking and satellite-framed languages (Laz, Georgian, Russian, Maltese (Stolz and Ammann 2008: 176; Ebert 1995: 770; Ket),

- aspectually head marking and both verb and satellite-framed languages (Tagalog, Modern Standard Arabic, Lezgian (Northeast Caucasian)),
- aspectually head-marking and serial-framed languages (Cayuga, Ngan'gityem-erri),
- aspectually double-marking and verb-framed languages (Greek, Spanish, French),
- aspectually double-marking and satellite-framed languages (Bulgarian, Hungarian, English) and
- aspectually dependent-marking and satellite-framed languages (Finnish).

Aspectually dependent-marking and verb-framed languages have not been found, neither among aspect nor among the advanced aspectoid systems. Aspectually double-marking languages seem to be well represented and either verb-framed or satellite-framed. Of course, there are languages of any framing type that do not have an aspect or advanced aspectoid category, e.g. Greenlandic (verb-framed), Ainu (serial-framed) and German (satellite-framed).

What is most striking about the sample is that there are no **aspect** languages that are aspectually head-marking and verb-framed, but only languages with an advanced aspectoid system. This may be due to its non-representativeness, of course. Aspectual head-marking and verb-framing would match insofar as aspectual head-marking operates with inherently goal-attaining verbs that express path. Such verbs do not need the activation of the goal-attainment (boundary transgression) reading by adverbials or overt participants nor differentiated case/adposition inventories. The languages closest to this pattern show properties of both verb and satellite framing (for Modern Standard Arabic cf. Alhamdan et al. 2018; for Lezgian cf. examples in Haspelmath 1993: 94, 192, 376, 378). Advanced aspectoid systems may combine aspectual head-marking and verb framing, but may have case inventories without conflation of direction and location.

On the other hand, aspectual dependent-marking and satellite framing match insofar as a dependent activates the goal-attainment (boundary transgression) reading of less specific or manner-of-motion verbs, for which the dependents need differentiated inventories of case/adpositions. This type is represented only by Finnish here.

The polysynthetic languages are aspectually head marking, and the path component is a non-root bound morpheme in Cayuga and Ket and a bound light verb in Ngan'gityemerri.

The most striking trait are prefixes or preverbs, which are used both for expressing path in satellite-framed languages and for encoding aspect including degree of affectedness, e.g. in Kartvelian, Slavic languages and Hungarian. Those languages are satellite-framed and aspectually head or double marking. Where

the preverbs/particles have been grammaticalised from case/adpositions, as in Slavic, Germanic languages and Lezgian (Haspelmath 1993: 169–170, Ch. 12), the adposition/case inventory is large. In the aspect language Lezgian, however, the preverbs are no longer productive and they do not have aspectual functions (cf. Haspelmath 1993: 167) as in Greek and Ket; the language shows both verb and satellite-framed patterns (compare examples in Haspelmath 1993: 94, 192, 376, 378). A large inventory of local and directional cases and postpositions is also found in Hungarian, although aspectual and path preverbs have only a small intersection.

Double-marking and verb-framed languages without preverbs tend towards direction–location conflation in adpositions (especially Greek and French, but not Spanish). Said the other way round, local and directional case/adposition distinction is low for aspectually double-marking languages without preverbs, and those languages tend to be verb-framed as well.

Vice versa, verb-framed languages seem to prefer head-marking advanced aspectoid systems or double-marking aspect systems, whereas satellite-framed languages tend to be aspectually dependent marking, but are found among all marking types.

In view of the weak co-occurrence preferences, correlations between aspectual systems and framing systems cannot be assumed. However, strong synergies can be observed in the languages that employ preverbs. They are a multifunctional tool for local relations, differential lexicalisation of actional classes, satellite framing, stacking of aspectual values and aspect. On the other hand, valency is used for encoding aspectual values particularly in languages without preverbs (e.g. Cayuga, Tagalog, Spanish, Japanese). Differential lexicalisation with respect to valency is found in aspectually head-marking languages; valence-lability in dependent-marking languages.

## 5 Conclusion

The present study has a look at structure in the realm of head, double and dependent marking of aspectual values in aspect languages and languages with an advanced aspectoid system, at the stacking of aspectual values, differential lexicalisation, DVM, verb, satellite and serial framing and conflation of local and directional case or adpositions in order to find possible correlations. It underlines the parallels of verbs of consumption, production and goal-attainment in encoding partial vs. full affection; valency is a means of encoding aspectual values particularly in languages without preverbs.

Differential lexicalisation with respect to goal attainment seems to be typical of aspectually head-marking languages, and not typical of dependent-marking languages (e.g. Finnish), as well as less pronounced in verb-framed double-marking languages (e.g. Spanish, French). It is generally independent of the use of aspectual preverbs, as Russian shows, which has differential lexicalisation for goal attainment, but preverbs elsewhere.

A correlation of head/dependent marking of aspectual values and verb/satellite framing is not borne out, but several preferences can be found. In the present sample, it is striking that there are no aspect languages that are aspectually head marking and verb-framed, in particular in view of the fact that aspectual head marking is compatible with the broadest range of framing types.

In French and Italian, we observed a conflation of manner of consumption and production as “replacement” strategy for stacking of aspectual values, and French, as a verb-framed language, conflates manner of motion, as well, in the sense that path verbs can be considered less specific in meaning than verbs which encode manner and path on the same verb form, be they satellite framing or serial framing. However, regarding verb framing as a conflation strategy seems to be in contrast to the use of differential lexicalisation in verb-framed languages, especially when taking into account specific path verbs oriented at the local systems, as in Ket and Nivkh (cf. Mattissen 2006).

Verb-framed Spanish stands out for its stacking strategy in view of most stacking aspect languages (such as Kartvelian and Slavic languages) being satellite-framed.

## Abbreviations

ACC	accusative
ACV	anterior converb
ADE	adessive
ALL	allative
ART	article
AV	actor voice
CMPL	complementiser
CPL	completive
CST	causative
CV	converb
DAT	dative
DEF	definite
DEL	delimitative
DEM	demonstrative
DOM	differential object marking

DTR	detransitivised
DVM	differential voice marking
ELA	elative
EV	effective voice
FUT	future
GEN	genitive
GER	gerund
HILI	highlighting focus
ILL	illative
IND	indicative
INDEF	indefinite
INESS	inessive
INS	instrumental
IPFV	imperfective
LOC	locative
M	masculine
MOT	motative
NML	nominalisation
NOM	nominative
NP	nominal phrase
PART	partitive
PERF	perfect
PERM	permanent
PEX	plural exclusive
PFTV	perfective
PL	plural
PP	prepositional phrase
POSS	possessor
POT	potential
PROG	progressive
PRS	present
PRV	preverb
PST	past
PTC	participle
RED	reduplication
REFL	reflexive
S	subject
SG	singular
SM	singular masculine
SV	superficial voice
TOP	topic
TRNSL	translative
TV	theme voice
V	verb
VP	verb phrase

## References

- Alhamdan, Bandar, Oudah Alenazi & Zouheir A. Maalej. 2018. Motion verbs in Modern Standard Arabic and their implications for Talmy's lexicalization patterns. *Language Sciences* 2018. 1–14.
- Ameka, Felix & James Essegbey. 2013. Serialising languages: Satellite-framed, verb-framed or neither. *Linguistics*. DOI: 10.4314/GJL.V2I1.34.
- Batiukova, Olga. 2015. Event structure and lexical semantics in a scalar approach to actionality. In Elisa Barrajón López, José Luis Cifuentes Honrubia & Susana Rodríguez Rosique (eds.), *Verb Classes and Aspect*, 21–59. Amsterdam: Benjamins. doi.org/10.1075/ivitra.9.02bat.
- Beffa, Marie-Lise. 1982. L'organisation de l'espace et son parcours. *Études mongoles et sibériennes* 13. 117–137.
- Bertinetto, Pier Marco & Denis Delfitto. 2000. Aspect vs. Actionality: Why they should be kept apart. In Östen Dahl (ed.), *Tense and Aspect in the Languages of Europe*, 189–225. Berlin: Mouton de Gruyter.
- Braginsky, Pavel & Susan Rothstein. 2008. Vendler classes and the Russian aspectual system. *Journal of Slavic Linguistics* 16(1). 3–55.
- Breu, Walter. 1992. Zur Frage des Verbalaspekts im Italienischen und Italoalbanischen. In Gabriele Birken-Silverman, & Gerda Rössler (eds.), *Beiträge zur sprachlichen, literarischen und kulturellen Vielfalt in den Philologien*, 158–172. Stuttgart: Steiner.
- Breu, Walter. 1994. Interactions between lexical, temporal and aspectual meanings. *Studies in Language* 18. 23–44.
- Breu, Walter. 1996. Komponentenmodell der Interaktion von Lexik und Aspect. In Wolfgang Girke (ed.), *Slavistische Linguistik 1995*, 37–73. München: Sagner.
- Breu, Walter. 1997. Komplexe aktionale Verbklassen, insbesondere Inchoativa. In Tilman Berger & Jochen Raecke (eds.), *Slavistische Linguistik 1997*, 55–80. München: Sagner.
- Breu, Walter. 2000. Zur Position des Slavischen in einer Typologie des Verbalaspekts. In Walter Breu (ed.), *Probleme der Interaktion von Lexik und Aspekt (ILA)*, 21–54. Tübingen: Niemeyer.
- Breu, Walter. 2007. Der Verbalaspekt im Spannungsfeld zwischen Grammatik und Lexik. *Sprachwissenschaft* 32(2). 123–166.
- Breu, Walter. 2009. Verbale Kategorien: Aspekt und Aktionsart. In Sebastian Kempgen, Peter Kosta, Tilman Berger & Karl Gutschmidt (eds.), *Die slavischen Sprachen / The Slavic Languages*. HSK 32.1, Vol. I, 209–225. Berlin: Mouton de Gruyter.
- Bridgen, Nigel. 1984. Towards a functional grammar of aspect in Finnish. In Caspar de Groot & Hannu Tommola (eds.), *Aspect Bound: A Voyage into the Realm of Germanic, Slavonic and Finno-Ugrian Aspectology*, 179–198. Dordrecht: Foris.
- Chapado Chorro, M. Olga & Luisa García García. 1991. Spanisch. In Hans-Jürgen Sasse (ed.), *Aspektsysteme*. Arbeitspapier Nr. 14 (Neue Folge), 47–68. Köln: Institut für Sprachwissenschaft.
- Clark, Eve V. 2005. Resultant states in early language acquisition. In Dorit Ravid & Hava Bat-Zeev Shyldkrot (eds.), *Perspectives on Language and Language Development: Essays in honor of Ruth A. Berman*, 175–190. Berlin: Springer.
- de Groot, Caspar. 1984. Totally affected. Aspect and three-place predicates in Hungarian. In Caspar de Groot & Hannu Tommola (eds.), *Aspect Bound: A Voyage into the Realm of Germanic, Slavonic and Finno-Ugrian Aspectology*, 133–151. Dordrecht: Foris.
- de Hoop, Helen & Andrej Malchukov. 2007. On fluid differential case marking: A bidirectional OT approach. *Lingua* 117. 1636–1656.

- Drossard, Werner. 2002. Paleosiberian Ket as a polysynthetic language (with special reference to complex verbs). In Nicholas Evans & Hans-Jürgen Sasse (eds.), *Problems of Polysynthesis* (STUF Studia Typologica 4), 223–256. Berlin: Akademie.
- Ebert, Karen H. 1995. Aspect in Maltese. In Pier Marco Bertinetto, Valentina Bianchi, James Higginbotham & Mario Squartini (eds.), *Temporal Reference, Aspect, and Actionality*, Vol. 2, 753–785. Torino: Rosenberg & Sellier.
- Evans, Nicholas. 2002. The true status of grammatical object affixes: evidence from Bininj Gun-wok. In Nicholas Evans & Hans-Jürgen Sasse (eds.), *Problems of Polysynthesis*, 15–50. (STUF Studia Typologica 4). Berlin: Akademie.
- Flecken, Monique, Christiane von Stutterheim & Mary Carroll. 2014. Grammatical aspect influences motion event perception: findings from a cross-linguistic non-verbal recognition task. *Language and Cognition* 6(1). 45–78.
- Giannakidou, Anastasia & Jason Merchant. 1999. Why Giannis can't scrub his plate clean: On the absence of resultative secondary predication in Greek. In Amalia Moser (ed.), *Greek Linguistics: Proceedings of the 3<sup>rd</sup> International Conference on Greek Linguistics*, 122–134. Athens: Ellinika Grámmata.
- Gruzdeva, Ekatarina Ju. 2012. Aspektual'nye klassy glagolov v nivxskom jazyke. [Aspectual classes in Nivkh language] In Nikolai N. Kazanskij & Vladimir A. Plungjan (eds.), *Tipologija aspektual'nyh sistem i kategorij [Typology of aspectual systems and categories]*, 458–533. Sankt-Peterburg: Nauka.
- Haspelmath, Martin. 1993. *A Grammar of Lezgian*. Berlin: De Gruyter Mouton.
- Haßler, Gerda. 2016. *Temporalität, Aspektualität und Modalität in romanischen Sprachen*. Berlin: De Gruyter.
- Heinäsmäki, Orvokki. 1984. Aspect in Finnish. In Caspar de Groot & Hannu Tammola (eds.), *Aspect Bound: A Voyage into the Realm of Germanic, Slavonic and Finno-Ugrian Aspectology*, 153–177. Dordrecht: Foris.
- Himmelman, Nikolaus P. 1987. *Morphosyntax und Morphologie. Die Ausrichtungsaffixe im Tagalog*. München: Fink.
- Himmelman, Nikolaus P. 2008. Lexical categories and voice in Tagalog. In Peter Austin & Simon Musgrave (eds.), *Voice and Grammatical Relations in Austronesian Languages*, 247–293. Stanford: CSLI.
- Himmelman, Nikolaus P. & Beatrice Primus. 2015. Prominence Beyond Prosody – A First Approximation. In Amedeo De Dominicis (ed.), *pS-prominenceS: Prominences in Linguistics. Proceedings of the International Conference*, 38–58. Viterbo: DISUCOM Press.
- Holisky, Dee Ann. 1981. Aspect theory and Georgian aspect. In Philip Tedeschi & Annie Zaenen (eds.), *Syntax and Semantics 14: Tense and Aspect*, 127–144. New York: Academic Press.
- Horrocks, Geoffrey & Melita Stavrou. 2003. Actions and their results in Greek and English: The complementarity of morphologically encoded (viewpoint) aspect and syntactic resultative predication. *Journal of Semantics* 20. 297–327.
- Horrocks, Geoffrey & Melita Stavrou. 2007. Grammaticalized aspect and spatio-temporal culmination. *Lingua* 117. 605–644.
- Ikegami, Yoshihiko. 1982. 'Source' vs. 'Goal': a Case of Linguistic Dissymmetry. In Robert N. St. Clair & Walburga von Raffler-Engel (eds.), *Language and Cognitive Styles: Patterns of Neurolinguistic and Psycholinguistic Development*, 292–308. Lisse: Swets & Zeitlinger.
- Isačenko, Alexander V. 1962. *Die russische Sprache der Gegenwart. Teil 1 – Formenlehre*, Halle: Niemeyer.
- Jacobsen, Wesley. 1982. *Transitivity in the Japanese verbal system*. Chicago: University of Chicago dissertation.

- Kiparsky, Paul. 1998. Partitive case and aspect. In Miriam Butt & Wilhelm Geuder (eds.), *The Projection of Arguments: Lexical and Compositional Factors*, 265–307. Stanford: CSLI Publications.
- Knittel, Marie-Laurence & Gyöngyvér Forintos-Kosten. 2002. Préverbes et aspect en Hongrois. In Brenda Laca (ed.), *Temps et Aspect. De la morphologie à l'interprétation*, 49–79. St.-Denis: Presses Universitaires de Vincennes.
- Krejnovič, Eruxim A. 1960. Vyraženie prostranstvennoj orientacii v nivxskom jazyke (k istorii orientacii v prostranstve) [The expression of spatial orientation in Nivkh (toward a history of orientation in space)]. *Voprosy Jazykoznanija* 9(1). 78–89.
- Larsson, Lars-Gunnar. 1984. The role of Baltic influence in the aspectual system of Finnish. In Caspar de Groot & Hannu Tommola (eds.), *Aspect Bound: A Voyage into the Realm of Germanic, Slavonic and Finno-Ugrian Aspectology*, 97–109. Dordrecht: Foris.
- Latrouite, Anja. 2011. *Voice and Case in Tagalog: The coding of prominence and orientation*. Düsseldorf: HHU.
- Leinonen, Marja. 1984. Narrative implications of aspect in Russian and in Finnish. In Caspar de Groot & Hannu Tommola (eds.), *Aspect Bound: A Voyage into the Realm of Germanic, Slavonic and Finno-Ugrian Aspectology*, 239–255. Dordrecht: Foris.
- Lindstedt, Jouko. 1985. *On the Semantics of Tense and Aspect in Bulgarian*. Helsinki: University of Helsinki Press.
- Mattissen, Johanna. 1994–2000. Field data Laz, unpublished field notes from work with native speakers. University of Cologne.
- Mattissen, Johanna. 1995. Verbmorphologie. In Silvia Kutscher, Johanna Mattissen, & Anke Wodarg (eds.), *Das Mutafi-Lazische. Arbeitspapier Nr. 24 (Neue Folge)*, 45–81. Köln: Institut für Sprachwissenschaft.
- Mattissen, Johanna. 2001. Tense and Aspect in Laz. In Karen H. Ebert & Fernando Zúñiga (eds.), *Arbeiten des Seminars für Allgemeine Sprachwissenschaft (ASAS) 16: Aktionsart und Aspectotemporalität in Non-European Languages*, 15–48. Zürich: Seminar für Allgemeine Sprachwissenschaft.
- Mattissen, Johanna. 2003a. Aktionale Klassen in polypersonalen Aspektsprachen – am Beispiel Lasisch. In Winfried Boeder (ed.), *Kaukasische Sprachprobleme*. 247–274. Oldenburg: Carl-von-Ossietzky-Universität Oldenburg.
- Mattissen, Johanna. 2003b. *Dependent-Head Synthesis in Nivkh. A Contribution to a Typology of Polysynthesis*. Amsterdam: Benjamins.
- Mattissen, Johanna. 2006. The ontology and diachrony of polysynthesis. In Dieter Wunderlich (ed.), *Advances in the Theory of the Lexicon*, 287–353. Berlin: Mouton de Gruyter.
- Mattissen, Johanna. 2012. Field data Nivkh, unpublished field notes from work with native speakers. Sapporo Gakuin University.
- Mattissen, Johanna. 2015. Mit Empathie gegen die Ellipse – Wie im Japanischen doch Person markiert wird. In Yoko Nishina (ed.), *Linguistische Berichte, Sonderheft 20: Sprachwissenschaft des Japanischen*, 111–141. Hamburg: Buske.
- Mattissen, Johanna. 2017a. Sub-types of polysynthesis. In Nicholas Evans, Marianne Mithun & Michael Fortescue (eds.), *The Oxford Handbook of Polysynthesis*, 70–98. Oxford: Oxford University Press.
- Mattissen, Johanna. 2017b. Field data French and Spanish, unpublished field notes from work with native speakers. University of Cologne.
- Mattissen, Johanna. 2024. Lexical actional classes in Romance languages in interaction with aspectually relevant material in the clause. In Jakob Egetenmeyer, Sarah Dessi Schmid & Martin G. Becker (eds.), *Tense, Aspect and Discourse Structure*, 153–187. Berlin: De Gruyter. DOI: 10.1515/9783111453897-006.

- Nishida, Chiyo. 1994. The Spanish reflexive clitic *se* as an aspectual class marker. *Linguistics* 32. 425–458.
- Paducheva, Elena V. 1998. On non-compatibility of partitive and imperfective in Russian. *Theoretical Linguistics* 24(1). 73–82. doi.10.1515/thli.1998.24.1.73.
- Panfilov, Vladimir Z. 1965. *Grammatika nivxskogo jazyka [Grammar of Nivkh language]*, Vol. 2. Moscow/Leningrad: Nauka.
- Panganiban, José Villa 1972. *Diksiunaryo-Tesaurus Pilipino-Ingles*. Lungsod Quezon: Manlapaz.
- Reid, Nicholas J. 2011. *Ngan'gityemerri*. München: LINCOM.
- Sanz, Montserrat. 2000. *Events and Predication*. Amsterdam: Benjamins.
- Sasse, Hans-Jürgen. 2000. Aspektsemantik und Lexikonorganisation im Cayuga. In Walter Breu (ed.), *Probleme der Interaktion von Lexik und Aspekt (ILA)* 193–246. Tübingen: Niemeyer.
- Sasse, Hans-Jürgen. 2002. Recent Activity in the Theory of Aspect: Accomplishments, Achievements, or just Non-progressive State?. *Linguistic Typology* 6 (3). 199–271.
- Savel'eva, V.N. & Taksami, Č.M. 1970. *Nivxsko-Russkij Slovar'* [Nivkh-Russian dictionary]. Moscow: Sovetskaja Enciklopedija.
- Slobin, Dan I. 2004. The many ways to search for a frog: Linguistic typology and the expression of motion events. In Sven Strömquist & Ludo Verhoeven (eds.), *Relating events in narrative*. Vol. 2 – *Typological and contextual perspectives*, 219–257. Mahwah: Lawrence Erlbaum Associates.
- Stephany, Ursula. 1985. *Aspekt, Tempus und Modalität*. Tübingen: Narr.
- Stolz, Thomas & Andreas Ammann. 2007. *Beda u Qabad*: The Maltese inchoative/ingressive. *Romano-Arabica* 6/7. 149–158.
- Stolz, Thomas & Andreas Ammann. 2008. The Maltese continuative: A grammaticalization borderline. In Elisabeth Verhoeven, Stavros Skopeteas, Yong-Min Shin, Yoko Nishina & Johannes Helmbrecht (eds.), *Studies on grammaticalization*, 169–184. (Trends in Linguistics 205). Berlin/New York: Mouton de Gruyter.
- Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical forms. In Timothy Shopen (ed.), *Language typology and syntactic description*. Vol. III: *Grammatical categories and the lexicon*, 225–282. Cambridge: Cambridge University Press.
- Talmy, Leonard. 2000. *Toward a cognitive semantics*. Vol. II – *Typology and process in concept structuring*. Cambridge: The MIT Press.
- Vogt, Hans. 1936. *Esquisse d'une Grammaire du Géorgien Moderne*. Oslo: Brøgger.
- Werner, Heinrich. 1997. *Die ketische Sprache*. Wiesbaden: Harrassowitz.



Silvia Luraghi

# The origin of agent markers – Metaphors underlying the extension of spatial prepositions in European languages

**Abstract:** Space provides the source domain for the conceptualization of more abstract domains. This can be shown in the extension of space prepositions to the encoding of non-spatial semantic roles. In the case of the domain of agency, the languages of Europe offer a wide variety of usages of space prepositions for the encoding of the semantic role agent with passive verbs. Metaphors discussed in the paper involve the spatial semantic roles of origin, source, path and various types of locations.

**Keywords:** semantic roles, agent, semantic extension, prepositions, metaphor

## 1 Introduction

The encoding of semantic roles is a fascinating issue that can shed light on the cognitive mechanisms through which we understand reality in terms of the metaphorical mapping between concrete and abstract domains. A well known example of metaphor that explains the extension of a specific encoding from a source to a target semantic role is the Companion Metaphor (Lakoff and Johnson 1980), extensively investigated by Thomas Stolz in a number of important publications (see e.g. Stolz 2001a, b and Stolz et al. 2006).

In my paper, I will explore some semantic extensions that take space as their source domain and, through metaphorical mapping, lead to the target domain of agency. To do so I will focus on space prepositions that occur in the expression of passive agent in some ancient and modern languages of Europe.<sup>1</sup> The idea that a semantic role such as agent can be expressed through a spatial metaphor is based on the assumption that human beings use more concrete categories to understand

---

<sup>1</sup> Agents are usually chosen as subjects in nominative-accusative languages, such as the languages analyzed in this paper. Thus, they are assigned the nominative case, that marks a grammatical relation, rather than a specific semantic role. However, passive constructions can provide evidence on the conceptualization of agents when agent phrases are not included in the verbal valency.

---

**Silvia Luraghi**, University of Pavia, e-mail: [silvia.luraghi@unipv.it](mailto:silvia.luraghi@unipv.it)

less concrete, abstract ones; in particular that, as space is the first accessible experience for human beings, spatial concepts are often extended in order to conceptualize non-spatial relations.

The paper is organized as follows. After a short discussion of the agent role and its connection with syntactic and semantic transitivity in the languages inspected (Section 2), I will proceed to investigate metaphors based on the spatial relations of origin (Section 3), source (Section 4), path (Section 5) and location (Section 6). In Section 7 I summarize the findings. The languages from which I will draw my data comprise Ancient Greek, Latin, Germanic, Romance and Slavic languages, including some historical stages. I take as a starting point findings from Luraghi (2000, 2003 and 2014). Among the languages mentioned above, Ancient Greek is of special interest because, as remarked in Chantraine (1953: 180), the passive was not yet fully grammaticalized at its earliest stages, and constructions that include an agent phrase show a wide variety of spatial prepositions. For this reason, I will often focus on Ancient Greek in Sections 3–6.<sup>2</sup>

## 2 The semantic role agent

Agents are prototypically capable of acting intentionally: as a consequence, they are necessarily animate, or, more precisely, human. Agents control the event that they bring about, that is, they can not only initiate it but also bring it to its end. Hence, the agent role is assigned to the participant who bears primary responsibility for voluntarily initiating (intentionality) and actually performing (controlling) the action.

Prototypically, this role appears with events that denote a change of state in another entity, which is assigned the patient role. Such events conform to the basic form of the Conceptual Archetype described as *billiard ball model* by

---

<sup>2</sup> While this paper specifically concentrates on the role of metaphors in semantic extension, it must be pointed out that metaphors do not account for all cases of such process. Different types of metonymy also have a role: a likely instance is the extension of instrumental markers to the encoding of the agent. In Luraghi (2001b) it is suggested that this type of extension should be accounted for in terms of metonymy, as agent and instrument are contiguous concepts, and an instrument can be seen as a part of an agent. In addition, it is frequently the case that inanimate entities substitute for animate ones in cases of metonymy, precisely when a part substitutes the whole. The occurrence of a metonymic extension here would also explain why this is virtually the only frequent semantic extension that contradicts Heine' scale of increasing grammaticalization (see Heine et al. 1991: 159). Accordingly, I will not discuss the instrumental of agent, which is consistently used for passive agents in several European languages, such as most Slavic languages.

Langacker (1991: 283–286) in its variant called *canonical event model* represented in Figure 1.



**Figure 1:** The canonical event model.

A prototypical agent and patient appear for example in (1a) and its passive variant (1b).

- (1) a. *John ate the apple.*  
 b. *The apple has been eaten **by John**.*

The two sentences denote a change of state: the apple, which functions as patient, exists before the event denoted by the verb *eat* and does not exist after it, and eating is an act that one typically performs intentionally, hence John can be said to take the agent role.

English, as several other languages including those that I will discuss in this paper, shows transitivity prominence, that is, the tendency for the transitive construction to extend to verbs that do not indicate change of state, such as experiential verbs (see Haspelmath 2015). This often implies that, no matter their degree of semantic transitivity, such verbs can also passivize. Consider for example perception verbs, as in (2a) and (2b).

- (2) a. *Mary saw the apple.*  
 b. *The apple has been seen **by Mary**.*

The sentences in (2) do not involve change of state: the apple exists before and after the event denoted by the verb *see*. The semantic role of Mary is actually experiencer; however, this participant is treated as an agent syntactically, and can be encoded as passive agent as well. From the point of view of verbal semantics, we can say that the sentence in (2a) is less transitive than of the sentence in (1a), precisely because the verb does not denote a change of state. As we will see, in Ancient Greek the degree of transitivity of the verb is relevant to the choice of the preposition with which the agent of the passive can be expressed.

Moreover, in English the subject of a transitive verb need not have the properties of a prototypical agent, and it can also be a noun denoting an inanimate

entity. Such an entity may also be encoded as an agent in the passive, as in (3a) and (3b).

- (3) a. *Fog prevented driving on the highway.*  
 b. *Driving on the highway was prevented **by fog**.*

In some cases, whether or not we are in the presence of a prototypical agent also conditions the distribution of spatial metaphors in the languages that we are going to survey. The semantic role assigned to non-prototypical (i.e., more specifically inanimate) agents is called force. It must be emphasized that non-prototypical agents, although they cannot act intentionally, are equally presented as ultimately responsible for a given event: hence, the role of non-prototypical agent differs from the semantic role of cause, because a state of affairs involving a cause can nevertheless also involve an agent, as in (4a) and (4b).

- (4) a. *Because of the fog, the driver could not see the road.*  
 b. *The road could not be seen (**by the driver**) because of the fog.*

Events that conform to the canonical event model can contain additional participants that facilitate the flow of energy from agent to patient. Such participants may have the semantic role instrument and intermediary, as in (5a) and (5b).

- (5) a. *I cut the apple **with a knife**.*  
 b. *The two heads of state communicate **through an interpreter**.*

In (5a) we have an expression of instrument, which denotes the concrete instrument used by the agent to carry out the action of cutting. Notably, while the inanimate entity cannot itself initiate and carry out the action, it is nevertheless necessary for the agent, who otherwise could not carry out a particular action. For this reason, depending on the language, even the instrument can in some cases be presented as responsible for an action, and be encoded as an agent. The intermediary, exemplified in (5b), is a human being that performs an action on behalf of another human being. In the sentences in which intermediary expressions appear, prototypical features of the agent role are divided between a primary agent, who is presented as exerting intentionality and control, and an intermediary, who, while lacking these features, is the participant who actually carries out the action. The instrument and the intermediary have an effect on the unfolding of the action in its final segment: in the canonical event model shown in Figure 1 they are located at the end of the transfer of energy from the agent to the patient.

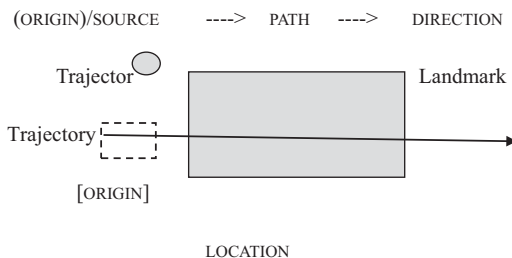
A more complex representation of semantic roles that also includes instrument and the spatial roles discussed in Sections 3–6 is shown in Figure 2 (based on Luraghi 2001a: 50 and Croft 2012: 280).

	ANTECEDENT	CONCOMITANT		SUBSEQUENT	
INTENTION	reason			purpose	
	ergative	comitative			
CAUSAL	cause	instrument			
	agent	means		recipient	beneficiary
		manner			
SPATIAL	ablative	path/locative		allative	

**Figure 2:** Semantic roles and the causal chain.

Figure 2 groups semantic roles into three groups, i.e. antecedent, concomitant and subsequent. It predicts that semantic roles in the antecedent and in the subsequent group cannot normally be encoded in the same way. In fact, there are complex relations between antecedent and subsequent roles, on which see Luraghi (2001a, 2014) and Croft (2012); for the purposes of this paper, however, it suffices noting that among spatial semantic roles only allative remains separated from the area in which the agent role is located. Other spatial roles, including source/origin, here tagged as “ablative”, location and path are located in the same area. Indeed, as I will show in Sections 3–6, these three roles provide the source domain for metaphors of agency in the languages reviewed, while no allative markers serve as source domain for the conceptualization of the agent role.

Figure 3 represents the prototypical motion event. The dashed figure represents the origin, which is focalized in cases of relative motion but often remains on the background; the gray areas represent the trajector (moving entity) and the landmark (reference entity).



**Figure 3:** A prototypical motion event.

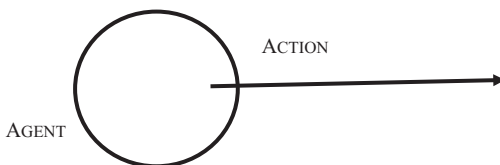
In the mapping of the space domain onto the domain of agency, the prototypical motion event, represented in Figure 3, provides the source for the canonical event schema, represented in Figure 1.

### 3 Agents as origins

The metaphor by which the agent is the origin of an action reflects the transfer of energy represented in Figure 1, in the meantime profiling the point of origin.

Origin expressions with the preposition *ek* ‘out of’ occur in Homeric Greek, and sporadically also in Classical Greek, only with prototypical agents and with verbs that do not indicate change of state. Elsewhere, this preposition indicates motion from the interior of a landmark (see Luraghi 2003, 2009), which is conceived as a container.

Metaphors based on containment are numerous. According to Lakoff and Johnson (1980: 29) “each of us is a container, with a bounded surface and an in-out orientation”: the notion of containment is connected with the first perception of a human being as such. Furthermore, as Johnson (1987) remarks, in our daily experience we also constantly experience our bodies as things in containers: for examples, we are contained in buildings, or vehicles, and so on. For this reason, the Container Schema may apply to many different aspects of reality, including abstract ones such as states and activities. Structural elements of this schema are, as remarked in Lakoff (1987: 272), interior, boundary, exterior. The Origin Metaphor for agents, then, is based on two separate metaphors: (a) ORIGINS ARE CONTAINERS and (b) AGENTS ARE ORIGINS. Hence, the event is conceptualized as coming out of an enclosed space, the agent, as in Figure 4.



**Figure 4:** The Origin Metaphor.

Remarkably, in an origin relationship the landmark exerts some degree of control over the trajector. In fact, the notion of control is implied by the Container Schema. As Vandelois (1994) has shown in his study of prepositions equivalent to *in* in French, English and German, the fact that a landmark is conceived as a container

implies that it exerts “dynamic control” over the trajector; that is, it conditions its behavior to some extent (see further Tyler and Evans 2003: 181; Luraghi 2004).

In Ancient Greek, two origin prepositions can encode passive agents. The first one is the already mentioned *ek* ‘out of’, as shown in (6) and (7).

- (6) Ancient Greek (*Il.* 2.668–669)

*ephilēthen*            *ek*        *Diós*  
 love.AOR.PASS.3PL    out.of    Z..GEN  
 ‘They were loved by Zeus.’

- (7) Ancient Greek (*Il.* 18.74–75)

*tà*                    *mèn*    *dě*    *toi*            *tetélestai*  
 DEM.NOM.PL    PTC    PTC    2SG.DAT    accomplish.PRF.M/P.3SG  
*ek*        *Diós*  
 out.of    Z..GEN  
 ‘Those things have been accomplished for you by Zeus.’

After Homer *ek* occurs in expressions of agent virtually only in Herodotus, and, as in Homer, remains limited to prototypical agents with verbs that have a generic meaning (‘do’, ‘accomplish’) or show a low degree of transitivity, as shown in (8).

- (8) Ancient Greek (Hdt. 3.404)

*trópōi*            *tōi*            *ex*            *emeū*  
 manner.DAT    DEM.DAT    out\_of    1SG.GEN  
*hupokeiménō*                    *akéo*  
 establish.PTCP.PRS.M/P.DAT    strive.PRS.IMP.2SG  
 ‘Try to settle the matter as advised by me.’

An Origin Metaphor also underlies the use of the preposition *prós*+GEN, Homeric Greek *potí*+GEN. In Homeric Greek this preposition does not indicate a concrete motion with human landmarks, but rather the origin, and in some passages it may refer to the human being that, though not performing an action, controls it, as in (9).

- (9) Ancient Greek (*Il.* 6.456)

*pròs*    *állēs*            *històn*    *hupháinois*  
 toward    other.GEN    loom.ACC    weave.OPT.PRS.2SG  
 ‘You will ply the loom at the orders of somebody else.’

Consistent with the Container Schema, the landmark in (9) is conceived as exerting some control on the trajector which performs the action: it is conceptualized as the origin of control. The preposition may occur in agent expressions, as in (10). Again, here and in other occurrences of the Origin Metaphor, verbs involved show a low degree of transitivity and agents are always prototypical (see Luraghi 2000).

- (10) Ancient Greek (*Il.* 11.831)

*tá se protí phasin Akhillêos*

DEM.NOM.PL 2SG.ACC toward say.PRS.3PL A..GEN

*dedidákthai*

teach.PRF.INF.M/P

‘The things that people say you have been taught by Achilles.’

In post-Homeric Greek, *prós*+GEN indicates the agent mostly limited to action nouns, as in (11).

- (11) Ancient Greek (Hdt. 3.19)

*Karkhédónioi mén nun doulosúnēn diéphugon*

Carchedonian.NOM.PL PTC PTC slavery.ACC escape.AOR.3PL

*pròs Perséōn*

toward Persian.GEN.PL

‘Thus the Carchedonians escaped being enslaved by the Persians.’ (Hdt. 3.19)

An Origin Metaphor also occurs with the use of *fram* in Old English, as argued by Fraser (1987). An example is (12).

- (12) Old English (Or. 95.17)

*Hasterbal wearð ofslagen from his agnum*

H.NOM AUX killed.PTCP from POSS.3SG OWN.DAT

*folce*

people.DAT

‘Asterbale was killed by his own people.’

In Old English, the two most used prepositions in agent phrases were *fram/from* and *of* (Green 1914). Fraser (1987: 245) argues that the former profiles the starting point, while *of* rather profiles the trajectory. This is shown by the greater frequency of *fram* with the “dynamic” passive (auxiliary *weorðan*) and of *of* with the “static” passive (auxiliary *beon/wesan*), noted in some texts by Fraser, as I discuss in Section 4.

## 4 Agents as sources

The Source Metaphor is similar to the Origin Metaphor in that it conceptualizes agents as the starting point of the transfer of agency that brings about the event, but it does not rely on the Container Schema. It rather profiles the trajectory: in this case, the metaphor at work is the ontological metaphor EVENTS ARE MOVING ENTITIES.

The difference in profiling between an origin and a source preposition has been observed by Fraser (1987) for Old English *of*. Let us consider example (13).

(13) Old English (Or. 84.2)

*seo burg wæs getimbred of Læcedemonium*  
 DEM city AUX found.PTCP by Spartan.DAT.PL  
 ‘The city had been founded by the Spartans.’

As highlighted by Fraser, we find the “static” passive with auxiliary *beon/wesan* ‘be’: the verb does not indicate a transfer, hence it is the ablative value of the preposition itself that supports the source meaning.

It needs to be remarked that cognates of Old English *fram* and *of* occur in agent phrases in other ancient Germanic languages as well (cf. Green 1914). In Gothic, for example, we find (14) with *fram* and (15) with *af*. As the former preposition occurs with the auxiliary *was* it is not clear that Fraser’s analysis of Old English can also extend to Gothic.

(14) Gothic (Mark 1.9)

*daupips was fram Iohanne in Iaurdane*  
 baptize.PTCP AUX by John in Jordan.DAT  
 ‘He was baptized by John in the Jordan river.’ (Mk. 1.9)

(15) Gothic (*Rom.* 12.21)

*ni gajiucaizau af unþiuþa*  
 not overwhelm.OPT.PASS.2SG by evil.DAT  
 ‘Be not overwhelmed by evil.’

The Source Metaphor is widely attested in modern and ancient languages of Europe, and it is instantiated for example by German *von* and Italian *da*, both meaning ‘from’. It also occurs in Bulgarian, which, contrary to most Slavic languages, does not feature inflectional case. Notably, passive agent is generally encoded by the instrumental case in Slavic languages (see fn. 2), but in Bulgarian loss of inflectional case did not result in an instrumental expression being extended to agent. Rather, we find the preposition *ot* ‘from’, as in (16).

- (16) Bulgarian (personal knowledge)

*Tazi kŭŝhta e postroena ot brat mi*  
 DEM house AUX built.PTCP from brother POSS.1SG  
 ‘This house has been built by my brother.’

Interestingly, in Old Russian one finds the preposition *otŭ* with prototypical agents of passive verbs, as shown in (17).

- (17) Old Russian (from Schmalstieg 2002: 48)

*Poskĕpani sabljami kalenymi ŝelomi*  
 cleave.PTCP.NOM.PL saber.INSTR.PL tempered.INSTR.PL helmet.NOM.PL  
*Ovarŭskii otŭ tebe, Jarŭ ture Vŭsevolode!*  
 Avar.NOM.PL from 2SG.GEN fierce.VOC bull.VOC V.VOC  
 ‘Cleft with tempered sabers are [their] Avar helmets by you, Fierce Bull Vsevolod!’

In Old Church Slavonic prepositional phrases with *otŭ*+GEN are also frequent for passive agents (see Bräuer 1952), as in (18).

- (18) Old Church Slavonic (Luke 21.17)

*bŕdete nenavidimi otŭ vcĕxŭ imene*  
 be.FUT.3PL hate.PTCP.PRS.PASS.NOM.PL from all.GEN.PL name.GEN  
*moego radi*  
 POSS.1SG.GEN for  
 ‘You will be hated by all for my name.’

The Source Metaphor was common with prototypical agents in Latin, as shown in (19).

- (19) Latin (Caes. G. 7.72)

*fit deinde senatus consultum ut ad*  
 happen.PRS.3SG then senate.GEN decision.NOM that to  
*bellum Parthicum legio una a Gn.*  
 war.ACC Parthian.ACC legion.NOM one.NOM by G.  
*Pompeio altera a G. Caesare*  
 Pompeius.ABL other.NOM by G. Caesar.ABL  
*mitteretur*  
 send.SUBJ.IMP.F.PASS.3SG  
 ‘Then there was a decision made by the senate, that one legion should be sent into the Parthian war by Pompeius, another one by Caesar.’

On the other hand, in Classical Greek the Source Metaphor is only sporadically attested, mostly with the preposition *pará*+GEN. This preposition indicated vicinity to a landmark and could occur with the dative, the genitive and the accusative, encoding different spatial relations. As it indicated vicinity, rather than containment, *pará* was especially suitable to indicate spatial relations relative to human landmarks. In Homer, other types of landmark occurred as well, with all three cases, while later on, in Attic-Ionic prose, the dative and the genitive virtually only occurred with human landmarks, and the accusative was also limited to human landmarks when denoting direction.

Agent expressions with *pará*+GEN occur in (20) and (21).

(20) Ancient Greek (Pl. *Phaedr.* 245b)

*ep' eutukhíai tēi megístēi parà theôn*  
 on happiness.DAT ART.DAT greatest.DAT from god.GEN.PL  
*hē toiaútē manía dídotai*  
 ART.NOM INDEF.NOM madness.NOM give:PRS.M/P.3SG  
 'Such madness is given by the gods for our greatest happiness.'

(21) Ancient Greek (Xen. *An.* 1.9.1)

*hōs parà pántōn homologeítai*  
 as from all.GEN.PL agree.PRS.M/P.3SG  
 'As everybody agrees.'

In addition, the preposition *apó*, which is the most frequent way of encoding source with concrete motion, also occurs in agent phrases, especially in Thucydides, as in (22).

(22) Ancient Greek (Th. 1.17)

*eprákhthē de oudèn ap' autōn érgon*  
 do.AOR.PASS.3SG PTC INDEF.ACC from DEM.GEN action.NOM  
*axiólogon*  
 noteworthy.NOM  
 'Nothing worth of being told was accomplished by them.'

Similar to origin prepositions analyzed thus far, both *pará*+GEN and *apó* are sporadically used for passive agents in Classical Greek and remain limited to generic or low transitivity verbs and prototypical agents (Luraghi 2000); however, post-Classical authors employ *apó* it with increasing frequency. In Modern Greek, *apó* is the preposition that routinely encodes the agent of passive verbs, with any degree of transitivity and both with prototypical and non-prototypical agents, as shown in (23) and (24).

- (23) Modern Greek (personal knowledge)  
*skotóthikan* *apó* *átaktus* *stratiótes*  
 killed.PASS.3PL from irregular.ACC.PL soldier.ACC.PL  
 ‘They were killed by irregular soldiers.’
- (24) Modern Greek (personal knowledge)  
*ta* *ḑéntra* *kserizóthikan* *apó* *ton*  
 ART.NOM.PL tree.NOM.PL uproot.PASS.3PL from ART.ACC  
*ánemo*  
 wind.ACC  
 ‘The trees were uprooted by the wind.’

## 5 Agents as paths

In the mapping of the space domain onto the domain of agency, the path of motion can be understood as the channel through which energy is transferred from the agent to the patient. Notably, however, the channel for the transfer of agency is not by itself the entity that controls the energy: hence, the extension to agent needs to be mediated by some other semantic role. This semantic role, as I will show by reviewing diachronic developments in Greek and Latin, is the role of the intermediary. For this reason, we start by discussing a two-step metaphor: CHANNELS ARE PATHS and INTERMEDIARIES ARE CHANNELS.

In Classical Greek, the intermediary role is encoded by *diá*+GEN with human landmarks, as shown in (25).

- (25) Ancient Greek (Hdt. 1.113.3)  
*pémpsas* *dè* *ho* *Hárpagos* *tôn*  
 sent.PTCP.AOR.NOM PTC ART.NOM H..NOM ART.GEN.PL  
*heōutoû* *doruphórôn* *toûs* *pistotátous*  
 REFL.GEN bodyguard.GEN.PL ART.ACC.PL trusty.SUP.ACC.PL  
*eíde* *te* *dià* *toútôn* *kai* *éthapse*  
 see.AOR.3SG and through DEM.GEN.PL and bury.AOR.3SG  
*toû* *boukólou* *tò* *paidíon*  
 ART.GEN cowherd.GEN ART.ACC SON.ACC  
 ‘Harpagus sent the most trusty of his bodyguard and he saw through them and buried the cowherd’s child.’

As has already been remarked in Section 2, the intermediary is a participant who acts at the instigation of an agent: it is to this second entity that not only intentionality, but also control are ascribed. The intermediary, similarly to an instrument, transfers the agent's energy to the patient. Crucially, the intermediary's role in this transfer is more active than that of an inanimate instrument, since in reality the intermediary acts, even if on behalf of the agent, using its own energies, which obviously an instrument cannot do.

In some cases in Classical Greek prose we find *diá*+GEN with human landmarks and passive verbs. These are contexts in which the authors want to indicate a reduced degree of agency, for various reasons. An example is (26).

(26) Ancient Greek (Pl. *Phlb.* 16c)

<i>theôn</i>	<i>mèn</i>	<i>eis</i>	<i>anthrôpous</i>	<i>dósis</i> , . . .	<i>pothèn</i>
god.GEN.PL	PTC	to	man.ACC.PL	gift.NOM	from_somewhere
<i>ek</i>	<i>theôn</i>	<i>erríphē</i>	<i>diá</i>	<i>tinós</i>	
out_of	god.GEN.PL	grasp.AOR.PASS.3SG	through	INDEF.GEN	
<b><i>Promēthēōs</i></b>	<i>háma</i>	<i>phanotátōi</i>	<i>tinì</i>	<i>purì</i>	
P.GEN	together	bright.SUP.DAT	INDEF.DAT	fire.DAT	

'It is a gift of the gods for mankind, grasped from some place from the gods through some Prometheus together with a gleam of fire.'

As remarked in Luraghi (2003), the gods are presented as bearing the responsibility of a certain state of affairs, and *diá tinós Promēthēōs* denotes an indefinite entity, whose volitionality is certainly not an important feature (other occurrences of this type are discussed in Luraghi 2000 and 2003).

While in Ancient Greek intermediary expressions never extended further to passive agents, Latin offers a different picture. In Latin the intermediary was encoded by *per*, which, similar to Ancient Greek *diá*+GEN, means 'through', and indicates path in its spatial usage. An example is (27).

(27) Latin (Caes. G. 1.44.12)

<i>id</i>	<i>se</i>	<i>ab</i>	<i>ipsis</i>	<i>per</i>
DEM.ACC	3SG.ACC	from	DEM.ABL.PL	through
<b><i>eorum</i></b>	<b><i>nuntios</i></b>		<i>compertum</i>	<i>habere</i>
3PL.GEN	messenger.ACC.PL		learn.PTCP.PRF.ACC	AUX

'He had learned this from them through messengers.'

This type of prepositional phrase can also come close to encoding the passive agent in Classical Latin, albeit sporadically, as in (28), in which a primary agent is not specified, or does not exist.

(28) Latin (Caes. G. 1.23.2)

*ea res per fugitivos [ . . . ]*  
 DEM.NOM thing.NOM through fugitive.ACC.PL  
*hostibus nuntiatur*  
 enemy.DAT.PL report.PRS.PASS.3SG

‘This was reported to the enemies by some fugitives [coming from the decuria of Gallic knights under the command of L. Aemilius].’

In Vulgar Latin, the preposition *ab* ‘from’, which indicated the passive agent in Latin and relied on the Source Metaphor, was replaced with *de* ‘from’, which had a similar meaning. In the Romance languages, we also find *de* to indicate the agent of the passive; as mentioned in Section 4, this is the case with Modern Italian *da*.

In some Romance languages, notably Spanish and French, the Source Metaphor was replaced by the extension of intermediary encoding, using prepositions deriving from the Latin *per* for the passive Agent, such as French *par* and Spanish *por*. This change is not brought about by a metaphorical shift: rather, it results from differences in profiling. In this case, the focus is placed on the final segment of the energy transfer. During the early stages of the Romance languages the two ways of encoding the passive agent competed, resulting in different outcomes, but intermediary expressions also occur in Old Italian as in (29).

(29) Old Italian (*Purg.* vii 6)

*fur l' ossa mie per Ottavian*  
 be.PST.3PL ART bone.PL POSS.1SG.PL through O.  
*sepolte*  
 bury.PTCP.PL

‘My bones were buried by Octavian.’

Finally, we may note that in German, while the agent of the passive is routinely expressed with the preposition *von* ‘from’ as remarked in Section 4, in the case of action nouns we find the preposition *durch* ‘through’. Compare (30) and (31).

(30) German (personal knowledge)

*Der Dieb wurde von der Polizei verhaftet*  
 ART.NOM thief AUX from ART.DAT police capture.PTCP.PST

‘The thief was caught by the police’;

- (31) German (personal knowledge)  
*Die Verhaftung des Diebes durch*  
 ART.NOM capture ART.GEN thief.GEN through  
*die Polizei*  
 ART.ACC police  
 ‘The capture of the thief by the police’.

## 6 Location Metaphors

The semantic role of location does not usually extend to agent in the languages under investigation here. Indeed, as shown in Figure 2, locative is a semantic role which is placed further toward the end of the causal chain than agent and related roles; in addition, not being dynamic, it does not easily lend itself to indicate a transfer of energy, as path does. The most clear example of a Location Metaphor occurring in the ancient and modern languages of Europe is based on the notion of superiority, and is instantiated by Ancient Greek *hupó*+GEN, as I discuss in Section 6.1. In Section 6.2, I briefly discuss the extension of English *by* to passive agent. As we will see, it is not clear whether this should be seen as based on the location meaning of the preposition or as based on its perlocative meaning, which also explains the extension to intermediary encoding.

### 6.1 Agents as superiorly located entities

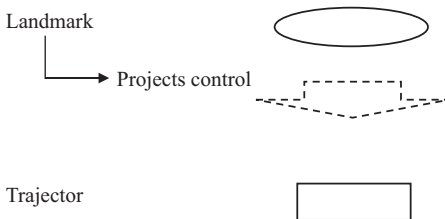
The most frequent way to encode the passive agent in Classical Greek, already frequently employed in Homer, involves prepositional phrases with *hupó* ‘under’ and the genitive case, as in (32).

- (32) Ancient Greek (Hdt. 6.29.2)  
*hōs pheúgōn te katelambáneto hupò*  
 as flee.PTCP.PRS.NOM PTC overtake.IMP.F.M/P.3SG under  
*andròs Pérsēō*  
 man.GEN Persian.GEN  
 ‘was overtaken in his flight by a Persian.’

The metaphor underlying this expression is related to the notion of physical superiority, and has its basis in the bodily experience of human beings. As Lakoff and

Johnson (1980: 34–35) note, “[h]aving control or strength is up, being subject to control or strength is down. . . Physical basis: physical size is generally correlated with physical strength, and the winner in the fight is generally above the vanquished.” The preposition *hupó* indicates a relationship of superiority of the landmark over the trajector: the latter, coming to be “under” the landmark, is subject to its control, being in a condition of physical inferiority.

The metaphor that accounts for this way to conceptualize the agent is PHYSICAL SUPERIORITY IS CONTROL. I will call this the Control Metaphor; it can be represented as in Figure 5.



**Figure 5:** The Control Metaphor.

Since the metaphor is based primarily on the concept of control, and not intentionality, it lends itself to transferring to the landmark both the characteristics of the agent and those of any kind of cause, even inanimate and unintentional. Indeed, inanimate nouns can occur in passive agent phrases with *hupó*+GEN. Remarkably, this prepositional phrase lost its spatial meaning after Homer, hence remaining limited to encoding agent and cause in Classical Greek. This is the opposite of the developments shown by other prepositions, including origin, source and path ones that I have analyzed in the preceding sections. As I have remarked, such prepositions can encode the passive agent with generic and low transitivity verbs or with nominalizations, and require prototypical agents. On the other hand, *hupó*+GEN can occur with all types of verb, regardless of their degree of transitivity, and it is the only prepositional phrase that occurs with change of state verbs; in addition, it also allows non-prototypical agents (see further Luraghi 2000).<sup>3</sup>

<sup>3</sup> In Homer passive agent can also be encoded by *hupó*+DAT. In addition, both with the dative and with the genitive *hupó* can occur in cause expressions and, only with the genitive and after Homer, even in intermediary expressions. See the extensive discussion in Luraghi (2000 and 2003).

## 6.2 Agents as places?

In standard English, the semantic role of agent with passives is routinely encoded by the preposition *by*. Sporadic examples already occur in Old English, as in (33).

(33) Old English (*Bl. H.* 163.27)

*þæt be þære cennendra gewyrhtum þæs bearnes*  
 that by DEM merits parent DEM child  
*weorþe ongyten wære be <þyson> eallum*  
 value understood AUX by DEM all  
*oðrum mannum*  
 other man

‘So that, through the merits of the parents, the value of the child might be understood by all these other men.’

The Old English preposition *be/bi* encoded location, as Modern English *by*, and, remarkably, this was its etymological meaning. However, it had already also extended to various other semantic roles, including perlocative, cause as in the first occurrence in (33), instrument and intermediary. For this reason, Fraser (1987: 247) assumes that the extension to agent was supported by what he calls “an operation of ‘translation’ or ‘transference’”, largely accounted for by the function of the verbal prefix *be-*, which had an applicative function in Old English.

According to Green (1914: 522), the preposition *be/bi* competed in Middle English not only with *of*, but also with *with* and *to*. Note that *with*, besides the current meaning, also indicated a location nearby someone, that is, it had a spatial meaning similar to *be/bi*. The possible occurrence of *to* in agent phrases can look surprising if one considers its allative meaning, but it must be noted that this preposition also indicated location, like Modern English *at*. In other words, all three prepositions that competed with *of* had a similar locative meaning. Green (1914: 529) assumes that the extension to the agent role was based on the locative meaning of *be/bi*, and highlights the fact that the agent can be conceived as the location at which an event takes place. The possible Location Metaphor, if this interpretation is correct, would then be AGENTS ARE PLACES WHERE EVENTS HAPPEN.

Put in this way, the idea that the extension of *be/bi* to the agent role is based on its locative meaning does not look so unlikely: ultimately, the Location Metaphor is similar to the Origin Metaphor, as the place where an event comes about is in fact its place of origin.

## 7 Conclusion

In this paper, I have discussed ways to encode the agent with passive verbs in a number of ancient and modern languages of Europe, concentrating on prepositional phrases that feature prepositions whose basic meaning expresses various spatial semantic roles. I have shown that such prepositions extend to agent through metaphorical processes that map the domain of space onto the domain of agency. Both the Origin and the Source Metaphor are based on spatial relations that involve roles located at the onset of the causal chain as represented in Figure 2. Especially the Source Metaphor is widely attested in agent phrases in the languages discussed in the paper, such as in Latin, Old English, Gothic, Italian, German, Bulgarian, Modern Greek, Old Russian and Old Church Slavonic.

The Origin Metaphor occurs in Ancient Greek and Old English. Path prepositions extend to agent in a two-step process: in the first place, through the Channel Metaphor, they extend to the role of intermediary. Then they may extend to the agent through restriction of the focus on the final segment of the energy transfer. This metaphor is especially present in some Romance languages, such as French and Spanish. Metaphors based on different aspects of the location role are not frequent. I have then described the Control Metaphor, that accounts for the extension of *hupó*+GEN to agent in Ancient Greek, and the Location Metaphor, that likely accounts for the extension of *be/bi* to agent in historical stages of English.

Remarkably, among the languages from which I drew my examples, the ones that display the widest variety of agent expressions are Ancient Greek and Old/Middle English. In Ancient Greek, one can detect a difference that underlies the use of different encoding strategies: while source and origin prepositions can occur only with low transitivity verb, verb with generic meaning and prototypical agents, only *hupó*+GEN occurs with change of state verbs and with all types of agents, including non-prototypical ones. In Old English there may also be a difference between the preposition *fram* that profiles the origin and the preposition *of* that profiles the trajectory, in that the latter preferably occurs with the “static” passive with auxiliary *beon/wesan* ‘be’.

Another difference that has emerged is between agents of passive verbs and agents of action nouns. In Ancient Greek the latter are encoded by *prós*+GEN, an origin prepositional phrase that also encodes behalf beneficiary. In German, passive agent is most often encoded via the source preposition *von* ‘from’ and occasionally by the path preposition *durch* ‘through’ which also encodes intermediary. Only the latter preposition occurs with action nouns. It is worth mentioning that Italian, too, features a special way of encoding the agent of action nouns, with the complex preposition *da parte di*, literally ‘from the side of’, which, similar to Ancient Greek *prós*+GEN, occurs in behalf beneficiary expressions.

## Abbreviations

ABL	ablative
ACC	accusative
AOR	aorist
ART	article
AUX	auxiliary
DAT	dative
DEM	demonstrative
FUT	future
GEN	genitive
IMP	imperative
IMPF	imperfect
INDEF	indefinite
INF	infinitive
INSTR	instrumental
M/P	medio-passive
NOM	nominative
OPT	optative
PASS	passive
PL	plural
POSS	possessive
PRF	perfect
PRS	present
PST	past
PTC	particle
PTCP	participle
REFL	reflexive
SG	singular
SUBJ	subjunctive
SUP	superlative
VOC	vocative

## Primary sources

*Bl. H.* = *Blickling Homilies*

*Caes. G.* = Caesar, *De Bello Gallico*

*Hdt.* = Herodotus, *Histories*

*Il.* = Homer, *Iliad*

*Luke* = *New Testament, Luke's Gospel*

*Mark* = *New Testament, Mark's Gospel*

*Or.* = Orosius

*Pl. Phaedr.* = Plato, *Phaedrus*

Pl. *Phlb.* = Plato, *Philebus*

*Purg.* = Dante Alighieri, *Purgatorio*

*Rom.* = *New Testament, Epistle to the Romans*

Th. = Thucydides, *History of the Peloponnesian War*

Xen. *An.* = Xenophon, *Anabasis*

## References

- Bräuer, Herbert. 1952. *Der persönliche Agens beim Passiv im Altbulgarischen. Eine syntaktischen Untersuchung*, Abhandlungen der Akademie der Wissenschaft und der Literatur in Mainz, Wiesbaden.
- Chantraine, Pierre. 1953. *Grammaire homérique. Tome 2. Syntaxe*. Paris: Klincksieck.
- Croft, William. 2012. *Verbs*. Oxford: Oxford University Press.
- Fraser Thomas. 1987. The establishment of “by” to denote agency in English passive constructions. In Anna Giacalone Ramat, Onofrio Carruba & Giuliano Bernini (eds.), *Papers from the VIIIth International Conference on Historical Linguistics*, 239–249. Amsterdam/Philadelphia: Benjamins.
- Green, Alexander. 1914. The analytic agent in Germanic. *Journal of English and Germanic Philology* 13. 514–552.
- Haspelmath, Martin. 2015. Transitivity prominence. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages: A comparative handbook*, Vol. 1, 131–147. Berlin/New York: De Gruyter Mouton.
- Heine, Bernd, Ulrike Claudi & Friederike Hünemeyer. 1991. *Grammaticalization: a conceptual framework*. Chicago: The University of Chicago Press.
- Johnson, Mark. 1987. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: The University of Chicago Press.
- Lakoff, George. 1987. *Women, fire, and dangerous things. What categories reveal about the mind*. Chicago: The University of Chicago Press.
- Lakoff, George & Mark Johnson. 1980. *Metaphors we live by*. Chicago: The University of Chicago Press.
- Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar*, Vol. 2: Descriptive Application. Stanford: Stanford University Press.
- Luraghi, Silvia. 2000. *Spatial metaphors and agenthood in Ancient Greek*. In Christian Zinko & Michaela Offisch (eds.), *125 Jahre Indogermanistik in Graz*, 283–298. Graz: Leykam.
- Luraghi, Silvia. 2001a. Syncretism and the classification of semantic roles. *Sprachtypologie und Universalienforschung* 54(1). 35–51.
- Luraghi, Silvia. 2001b. Some remarks on Instrument, Comitative, and Agent in Indo-European. *Sprachtypologie und Universalienforschung* 54(4). 385–401.
- Luraghi, Silvia. 2003. *On the Meaning of Prepositions and Cases. A Study of the Expression of Semantic Roles in Ancient Greek*. Amsterdam/Philadelphia: Benjamins.
- Luraghi, Silvia. 2004. The container schema in Homeric Greek. In Augusto Soares de Silva, Amadeu Torres & Miguel Gonçalves (eds.), *Linguagem, cultura e cognição: Estudos de Linguística Cognitiva*, 25–41. Braga: Almedina.
- Luraghi, Silvia. 2009. The evolution of local cases and their grammatical equivalent in Greek and Latin. In Johanna Barðdal & Shobhana Celliah (eds.), *The Role of Semantics and Pragmatics in the Development of Case*, 283–305. Amsterdam/Philadelphia: Benjamins.

- Luraghi, Silvia. 2014. Plotting diachronic semantic maps. The role of metaphors. In Silvia Luraghi & Heiko Narrog (eds.), *Perspectives on Semantic Roles*, 99–150. Amsterdam/Philadelphia: Benjamins.
- Schmalstieg, William. 2002. The Indo-European genitive of agent with finite verbs and participles. *Acta Linguistica Lithuanica* 47. 41–54.
- Stolz, Thomas. 2001a. Comitatives vs. instrumentals vs. agents. In Walter Bisang (ed.), *Aspects of Typology and Universals*, 153–174. Berlin: Akademie Verlag.
- Stolz, Thomas. 2001b. To be with X is to have X: Comitatives, instrumentals, locative and predicative possession. *Linguistics* 39(2): 321–350. DOI: 10.1515/ling.2001.013
- Stolz, Thomas, Cornelia Stroh & Aina Udrze. 2006. *On Comitatives and Related Categories. A Typological Study with Special Focus on the Languages of Europe*. Berlin/New York: Mouton de Gruyter.
- Tyler, Andrea & Vyvyan Evans 2003. *The Semantics of English Prepositions: Spatial Scenes, Embodied Meaning and Cognition*. Cambridge: Cambridge University Press.
- Vandelois, Claude. 1994. Methodology and analysis of the preposition *in*. *Cognitive Linguistics* 5(2). 157–184.



Paolo Ramat

# What(ever) will be, will be

**Abstract:** The paper discusses from the perspective of the “Multi Word Expression” of the constructionist approach the particular nature of (apparently) tautological constructions, comparing these with other types of reduplication or iteration, generalized constructions and stereotyped idiomatic expressions such as *to kick the bucket*. Stereotyped idiomatic expressions are the result of a usually long linguistic tradition, while tautological constructions may be created in every particular (pragmatic) linguistic situation.

**Keywords:** word-iteration, reduplication, tautology, Multi Word Expressions

## 1 Different types of reduplication

The title of this article intentionally alludes to a topic Thomas Stolz has deeply studied in many insightful contributions: word-iteration. Apparently, *what(ever) will be, will be*<sup>1</sup> seems to be a tautological sentence of the “A is A”-type meaning “things are what they are”, i.e., a statement following the “law of identity” that Heger (1985: 469) dubbed as “absence d’information” since, logically speaking, “A is A” is necessarily always true. However, any English speaker will intend the sentence to express a kind of “cheerful fatalism” beyond human control.<sup>2</sup> Obviously, there is a

---

1 As for the history of the sentence, see Hartman (2013) where the author shows that the sentence was originally a heraldic motto of an aristocratic English family, perhaps firstly used by John Russell, first Earl of Bedford, after his experience at the battle of Pavia (1525), in the “Romance” form *quy serra serra*. Hartman’s conclusion is as follows: “Although it is difficult to prove a negative, my searches in authentic collections of traditional proverbs in French, Spanish, and Italian have produced no trace of KSS [abbreviation for the formula *Que sera sera*: P.Rt.]” (Hartman 2013: 63). In fact, its Spanish and Portuguese, French (and Italian) corresponding forms (*que será será*, *que sera sera*, *che sarà sarà*) are, strictly speaking, ungrammatical: it should be *Lo que será*, *O que será*, *Ce qui sera* (and *Ciò che sarà*), respectively. The motto seems a quasi-Romance form, invented when the upper classes of England were discovering the Renaissance. From the construction point of view, it does not matter that the iterated form (*will be*) occurs first in a free relative (or interrogative) sentence introduced by *what*. Indeed, we find the same iteration in affirmative sentences as *boys will be boys*, *business is business*, etc.

2 Quine (1995: 199) speaks of “tautological fatalism”.

---

Paolo Ramat, Accademia dei Lincei – Roma, e-mail: paoram@unipv.it

big difference between *What will be, will be* and other types of reduplication (henceforth RED) such as emphasizing REDs in *a very very old man*, studied by Stolz on many occasions (Stolz 2003, 2006a, 2006b, 2008, etc.). The iteration / reduplication (on the terminology, see below, fn. 3) can also have morphological functions such as plural formation (= ex. (1)), a distributive function (= ex. (2)), an intensifying function (= ex. (3) and (4)), or have particular semantic values such as the so-called “echo reduplication”, as in (5), etc.<sup>3</sup> A good overview of the multifarious functions of RED can be obtained when looking into the subject index of van Lier (2023: 1087, s.v. “reduplication”). Generally speaking, we can say that RED is a widespread strategy in the world’s languages to express many related functions (see Fedriani 2017: 147).

- (1) Indonesian (Stolz 2003: 18)  
*buku buku* ‘books’
- (2) Malti (Stolz 2006a: 119)  
*bieb bieb* lit. ‘door door’, i.e. ‘from door to door’
- (3) Italian (Stolz 2003: 36; Ramat 2019: 151)  
*un caffè caffè* ‘a real, good coffee’
- (4) Emmi (Western Daly) (Louagie 2023: 555)  
*dukandji* ‘big’ ~ *dukduk* ‘very big’
- (5) Turkish (Stolz 2018: 248)  
*dergi mergi okumuyor*  
newspapers m:ECHO read:NEG:PRES  
‘(s)he does not read newspapers and the like’

---

3 Grandi (2017: 77) writes: “According to Medici (1959: 84), the aim of a reduplicated construction such as *caffè caffè* (‘coffee coffee’ [ex. (3)], P.Rt.) is “affermare [. . .] la pienezza delle qualità inerenti [alla base]” (‘to state the fullness / completeness of the qualities of the base word’). In other words, the primary function of reduplication would be to identify the prototype of a concept, its default value. If evaluative morphology (which includes intensification) encodes a deviation from a standard or from a default meaning, if its peculiar function is that of assigning a value which is different from that of the “standard / default” to a concept, reduplication primarily expresses a full identification of an item with its standard image. It identifies the best exemplar of a class”. Interestingly, in a mail diffused via Lingtyp (11.07.2024), Camil Staps notes that RED is used not only as a means expressing authenticity / prototypicality but also diminutive, inauthenticity, and atypicality as in the case of Makasar (Indonesia, Austronesian): *tau* ‘person’ → *tau-tau* ‘statue, doll’, *lima* ‘hand’ → *lima-lima* ‘hand-like object’ (such as a stick with twigs looking like fingers), *balla* ‘house’ → *balla’-balla* ‘hut’ (data from Jukes (2020).

In the following example the RED is slightly modified in its morphological construction: the genitive RED (*zemiu*) is used to indicate a considerable quantity.

- (6) Latvian (Holvoet 2011: 16)

*Vīnš izstaigāja zemiu zemes*  
 3SG.M roam\_through.PST.3 country.GEN.PL country.ACC.PL  
 ‘He roamed through many countries’ (lit. country of countries).

Example (7) shows the insertion of morphological material between the reduplicating verbal lexeme.

- (7) Wandala (Central Chadic) (Frajzyngier 2016: 265)

*Nóo và tə́-n kɪni njà-n-í-njà á-t wáfk-à*  
 PRES time T-DEM C.FOC sit.1SG-EPENTH-sit PRED-TARGET face-GEN  
*ordinateur yá ndà ə'lv Wándàlā*  
 computer 1SG speak language Wandala  
 ‘Here I am sitting in front of the computer speaking the Wandala language.’

The last two examples show that, as is the case with many concepts of linguistics, beyond the strict RED definition, there may also be gradient stages of this phenomenon.

To note that (3) has a completely different meaning than *il caffè è caffè*, whereby the tautology (henceforth, TTL) “A = A” has very poor information (“absence d’information”), unless we imagine a special contextual situation (for instance, two people discussing the difference between coffee and tea, in which case Grice’s implicatures play a crucial role). In what follows I shall deal just with the iteration type exemplified in the title of this paper and not discuss other iteration/reduplication types such as those exemplified in (1)–(7).<sup>4</sup>

---

<sup>4</sup> Stolz (2007) has dedicated an important article to discussing the difference between *Wiederholung* (iteration / repetition) and *Reduplikation*. *Wiederholungen* cause changes at the connotative layer. *Reduplikationen* have to do with changes at the semantic level (Stolz 2007: 57). The definition provided of *Reduplikation* is as follows (Stolz 2007: 64): “der Begriff Reduplikation [trifft] nur dann voll zu[. . .], wenn mit der Gleichheit der Ausdrucksseite auch eine Gleichheit auf der Inhaltsseite einhergeht dergestalt, daß der Gesamtkonstruktion eine andere Bedeutung zukommt als den jeweiligen Einzelbestandteilen” (“the concept of reduplication applies completely only if to an identity at the expression level there corresponds identity at the content level, in such a way that the construction in its totality acquires a meaning which is different from the meaning of its components’, my translation). This seems to apply to our *what will be, will be*-type, where the construct means more than the simple sum of the two parts. However, TTLs are not the object of Stolz’s article, in which he also states that it is difficult to draw a divide between the connotative *Wiederholungen*

## 2 Iteration, tautology and reduplication

However, both tautological sentences and reduplication make use of iteration, and it can be stated that word iteration can be found in every language, though with different scopes and different results. To my knowledge, no language has a rule forbidding word iteration / reduplication (see Stolz 2006b: 291–292).

So, the question is: what is the difference between reduplication and tautology? There exist TTLs such as *business is business* (= German *Geschäft ist Geschäft*) or *boys will be boys* which, like *what will be, will be*, signify much more than their literal meaning. As stated by Hagège (1985: 148), “Face à ces formules, toute analyse en termes de logique propositionnelle ne peut échapper au constat d’identité, et, partant, d’inanité” (‘All analyses based on proposition logic, when confronted with such formules, are bound to constate identity and, consequently, uselessness’, my translation). However, speakers, based on their encyclopaedic knowledge which is the result of the experiences done in their socio-cultural world, will give a particular sense to such sentences, which, if taken literally, are completely deprived of information (see Heger 1985, quoted above).<sup>5</sup>

TTLs of this kind are generalized expressions. It is worth noting that these TTLs are perfectly conforming to the grammatical rules of English and many other languages (cf. French *les affaires sont les affaires*; Italian *i soldi son soldi* ‘money is money’). New non-stereotyped TTLs can always be created: *a car is a car*, *apes will be apes*, etc. – and even *un soldo è un soldo* lit. ‘a penny is a penny’, though the stereotyped, usual TTL takes the plural: *i soldi* (see Ramat 2019: 147–148). As we shall see below, tense, mood, number, and other morphological grams do not matter in the TTL and we can have *cars are cars* and *apes were apes* without any change in the construction and value of the TTL.

Other stereotyped constructions too, like *to kick the bucket*, French *tomber dans les pommes* (lit. ‘to fall among the apples’) ‘to faint’ or Italian *essere al verde* (lit. ‘to be in the green’) ‘to be completely without money’ are used in non-literal meaning. The reason why ‘being completely without money, being broke’ can be

---

and the semantically significant *Reduplikationen*. Consequently, when dealing with TTLs, it is correct to speak both of iteration/repetition and reduplication, as I have done in this paper. On the difference between syntactic repetition and morphological reduplication see, however, Barotto and Mattioli (2020). They suggest a continuum between the two poles of reduplication and repetition / iteration, much in the sense adopted in this article.

<sup>5</sup> I will not initiate here a general discussion of Wierzbicka’s (1987) Radical Semantics, which affirms that *Boys will be boys* cannot be translated into French by *Les garçons sont les garçons*, since French people have an encyclopaedic knowledge of the world different from that of English people. Such a discussion does not concern the topic of the present paper.

expressed by *essere al verde* is not understandable even to the native speaker; nevertheless (s)he will make use of this construction without problems, since (s) he has already heard it. Only those who study the history of the Italian language can explain the origin of the expression,<sup>6</sup> and foreigners who have perfectly learned Italian and know both *essere* and *verde* will ask “what the hell does it mean?”.

### 3 Tautologies and idioms / generalized constructions

TTLs and idiomatic expressions such as *to kick the bucket* make up part of the “generalized constructions” (henceforth GCONs) illustrated by Goldberg’s (2006) constructionist approach as conventionalized pairings of form and function based on an input formed by general cognitive, pragmatic, and processing constraints.<sup>7</sup> However, there is a difference between TTLs and other generalized constructions, idioms included: the (non-native) speaker, faced with the (unknown to him/her) TTL *What will be, will be*, might ask “OK, and so what?”. On the other hand, when faced with the GCON *going great guns*, her/his reaction will be one of total not-understanding, as in the case of *to kick the bucket*.

Speaking of GCONs leads also to a consideration of Multi Word expressions (MWEs). Hüning and Schlücker’s abstract of their chapter 24 (“Multi-word expressions”) in Müller et al.’s (2015) handbook of word formation defines MWEs as “complex lexical units, for example verbal idioms (*bite the bullet*) or frozen adverbials (*all at once*). Others, such as particle verbs (*stick out*) or complex nominals (*day-care center*), indicate a close relationship between MWEs and word-formation units”.

Crucially, MWEs refer to a unitary concept and by default cannot be interrupted by the insertion of other lexemes: the expressions \**day-good-care center* or French \**chemin-long-de-fer* ‘long railway’ instead of *good day-care center* and *long*

---

6 To satisfy the reader’s curiosity, the explanation is as follows. Time ago candles used to be green coloured at their base, thus when the fire reached the green part, the candle was near to being extinguished. Metaphorically the *essere al verde* came to signify ‘to be at the end of money, be broke’.

7 Among the first examples of GCONs given by Goldberg (2006: 5) there is the idiomatic construction *going great guns*, i.e., ‘proceed vigorously or successfully’, which is as idiomatic as *kick the bucket*. Note, however, that Goldberg’s GCONs include morphemes (e.g. *pre-*, *-ing-*), words (*anacanda*), complex words (*daredevil*), etc. However, these GCONs need not concern us in the present context.

*chemin-de-fer* are not admitted (for some possible exceptions see below). We are dealing with both morphology and syntax, as there exists the plural form *chemins de fer* and *chemin* and *fer* are also autonomous words, bound by the morpheme *de* in a NP. Such MWEs show the same structure as *(the) battle of Waterloo* and even *(the) King of England* which also admits a plural *(the) Kings of England* when we allude to the series of kings England has had throughout its history. Even totally non-stereotyped NPs such as *(the) doctor's office* and *office of the doctor* have the same nominal structure, though they cannot be considered as MWEs in Hüning and Schlücker's sense.

Moreover, stereotyped expressions such as *face to face*, French *face à face* (or *vis-à-vis*), Italian *faccia a faccia*, Spanish and Portuguese *frente a frente*, German *Auge in Auge*, Dutch *oog in oog*, Russian *licom k licu*, Irish *aghaidh ar aghaidh*, etc., with iteration of the noun in the construction {N<sub>1</sub> – PREP – N<sub>1</sub>}, also belong to the MWEs. Sometimes MWEs are written with a hyphen, as in *rail-ticket* (and *vis-à-vis*), and sometimes they are not, as in *railway*, *sometime(s)*. At times, it is not easy to distinguish between MWEs and compounds, many of which may be idiomatic: *weekend*, *second-hand*. In fact, MWEs can be compound-like whose meaning goes beyond the sum of its components such as *parking ticket* 'official notice put on your vehicle that you must pay a fine'; or they may be formed by phrasal units with a non-compositional meaning that has to be learned as part of the lexicon. As in the example of *parking ticket*, their meaning is not composed of the meaning of their words, so that it cannot be derived directly from the words that constitute the phrasal unit. Accordingly, *going great guns*, *kick the bucket*, *tomber dans les pommes* and *essere al verde* can be considered MWEs (though sometimes they can be interrupted by the insertion of lexical material as in *essere completamente al verde* 'to be totally without money', the difference being that *chemin de fer* is an NP, while *essere al verde* is a sentence).

## 4 Tautologies vs idioms

After having introduced notions such as GCONs and MWEs, let us go back to the crucial point of this article. Strictly speaking, the constructions *que sera, sera* and *what will be, will be* can also be considered MWEs, inasmuch as the three/five words constitute a semantic unit, as well as *to kick the bucket*: pronouncing just the relative part *what will be* without its correlated second part, i.e., *will be*, does not make much sense. The difference between MWEs in their largest sense and TTLs of the *what will be, will be*-type lies in the fact that the latter are predicative nominal sentences, while the former constitute a unity, endowed with

a specific semantic value.<sup>8</sup> Moreover, MWEs are idioms which, as we mentioned above, are part of the lexicon – more precisely, of the phraseology of a language. As they are fixed expressions, they need a certain lapse of time to become stereotyped.<sup>9</sup> On the contrary, TTLs can be created at every moment and along with *what will be, will be* we can create, if needed by the context situation, *what has been, has been; what should be, should be* or *a dog is a dog, cats are cats* etc. In other words, TTL is an open possibility offered by the (constructional) grammar of the language.

## 5 The nature of Multi Word Expressions

To sum up, we may conclude that the label Multi Word Expression, which may include proverbs, stereotyped expressions like *as good as gold, by and large*, is a cover term both for tautologies of the type dealt with in the present article and generalized constructions.<sup>10</sup> Compounds and MWEs as well as TTLs have the common status of complex lexical constructs, whose meaning is more than the sum of their components. While the constructionist approach, as a critique of generativism, is particularly concerned with syntax, the discussed TTLs underline more the semantic aspect. The MWE perspective considers both the semantic and the syntactic side of the sentence, at the interface of grammar and lexicon (see Sag et al. 2002).

---

<sup>8</sup> See Mathieu et al. (2017), Schlücker (2018).

<sup>9</sup> In political language as in the pub and other specialized languages it is frequent nowadays to find slogans whose aim is to be remembered as fixed expressions. It is, however, rare that such slogans become real MWEs that permanently enter the lexicon. Usually, they last for very short time, *l'espace d'un matin*.

<sup>10</sup> See Masini (2005: 145): “the term multi-word expression is used as a pre-theoretical label to include the range of phenomena that goes from collocations to fixed expressions” For other types of TTL and RED, such as *step by step*, Afrikaans *een-een* ‘one by one’, see Ramat (2019). Foschi Albert (2017) examines the German and Italian type x’x such as *mehr und mehr* ‘more and more’ *parla e parla* ‘talk and talk’, as well as the RED without conjunction (e.g. German *lang lang* and Italian *lungo lungo* ‘very long’). In the same volume Bonacchi (2017) studies from a pragmatic point of view the Italian intensifying RED (e.g. *un cuore grande grande* ‘very generous, lit. a very big heart’, *Maria è una donna donna* ‘Maria is a real woman’) and its German equivalents which often are not fixed expressions as the MWEs are (*ein wahrhaft großes Herz* ‘a really big heart’, *Maria ist eine echte Frau* ‘Maria is a real woman’).

## Abbreviations

ACC	accusative
C.FOC	contrastive focus
DEM	demonstrative
ECHO	image in echo-word
EPENTH	epenthesis
GCON	generalized construction
GEN	genitive
M	masculine
MWE	Multi Word Expression
N	noun
NEG	negation
NP	noun phrase
PL	plural
PREP	preposition
PRES	present tense
PRED	predicator
PST	past tense
RED	reduplication
SG	singular
TTL	tautology
T	target

## References

- Barotto, Alessandra & Simone Mattioli. 2020. Reduplicazione e ripetizione in tipologia: due strategie separate o espressione dello stesso fenomeno?, CLUB: Circolo Linguistico dell'Università di Bologna. *Working Papers in Linguistics* 4. 47–66.
- Bonacchi, Silvia. 2017. What does reduplication intensify? The semantics and pragmatics of reduplicated forms in Italian and their equivalents in German. In Maria Napoli & Miriam Ravetto (eds.), *Exploring Intensification. Synchronic, diachronic and cross-linguistic perspectives*, 289–303. Amsterdam/Philadelphia: Benjamins
- Fedriani, Chiara. 2017. *Nulla sum, nulla sum. Tota, tota occidi. Repetition as a (rare) strategy of intensification in Latin*. In Maria Napoli & Miriam Ravetto (eds.), *Exploring Intensification. Synchronic, diachronic and cross-linguistic perspectives*, 146–169. Amsterdam/Philadelphia: Benjamins.
- Foschi Albert, Marina. 2017. The coordination of identical conjuncts as a means of strengthening expressions in German and Italian. In Maria Napoli & Miriam Ravetto (eds.), *Exploring Intensification. Synchronic, diachronic and cross-linguistic perspectives*, 265–287. Amsterdam/Philadelphia: Benjamins.
- Frajzyngier, Zygmunt. 2016. Modality and mood in Chadic. In Jan Nuyts & Johan van der Auwera (eds.), *The Oxford Handbook of Modality and Mood*, 238–295. Oxford: Oxford University Press.

- Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Grandi, Nicola. 2017. Intensification processes in Italian. In Maria Napoli & Miriam Ravetto (eds.), *Exploring Intensification. Synchronic, diachronic and cross-linguistic perspectives*, 55–75. Amsterdam & Philadelphia: Benjamins.
- Hagège, Claude. 1985. *L'homme de paroles. Contribution linguistique aux sciences humaines*. Paris: Fayard.
- Hartman, Lee. 2013. “Que sera sera”: The English Roots of a Pseudo-Spanish Proverb. *Proverbium* 30. 51–104.
- Heger, Klaus. 1985. Acceptabilité sémantique? In Herman Parret & Hans-Georg Ruprecht (eds.), *Exigences et perspectives de la sémiotique. Recueil d'hommages pour A.J. Greimas*, 463–474. Amsterdam/Philadelphia: Benjamins.
- Holvoet, Axel. 2011. The Baltic languages. In Bernd Kortmann & Johan van der Auwera (eds.), *The Languages and Linguistics of Europe. A Comprehensive Guide*, 3–29. Berlin/New York: Mouton de Gruyter
- Hünig, Matthias & Barbara Schlücker. 2015. ‘Multi-word expressions’. In Müller, Peter O., Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word Formation. An International Handbook of the Languages of Europe --vol 1*, 450–467. Berlin/Boston: De Gruyter Mouton.
- Jukes, Anthony. 2020. A grammar of Makasar. Leiden: Brill.
- Louagie, Dana. 2023. Word Classes in Australian Languages. In Eva Van Lier (ed.), *The Oxford Handbook of Word Classes*, 544–571. Oxford: Oxford University Press.
- Masini, Francesca. 2005. Multi-word expressions between syntax and the lexicon: the case of Italian verb-particle constructions. *SKY, Journal of Linguistics* 18. 145–173.
- Mathieu, Constant, Gülşen Eryiğit, Johanna Monti, Lonneke van der Plas, Carlos Ramisch, Michael Rosner & Amalia Todirascu. 2017. Multiword Expression Processing: A Survey. *Computational Linguistics* 43(4). 837–892.
- Medici, Mario. 1959. Il tipo “caffè caffè”. *Lingua Nostra* 20. 84.
- Müller, Peter O., Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.). 2015. *Word Formation. An International Handbook of the Languages of Europe*, vol 1. Berlin/Boston: De Gruyter Mouton.
- Napoli, Maria & Miriam Ravetto (eds.). 2017. *Exploring Intensification. Synchronic, diachronic and cross-linguistic perspectives*. Amsterdam/Philadelphia: Benjamins.
- Quine, Willard Van Orman. 1995. Things and their Place in Theories. In Paul K. Moser & J.D. Trout (eds.), *Contemporary Materialism, A Reader*, 195–210. London/New York: Routledge.
- Ramat, Paolo, 2019. Tautologia e/o duplicazione. *Rivista Italiana di Linguistica e Dialettologia* 21. 147–156.
- Sag Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. 2002. Multiword expressions: a pain in the neck for NLP. *Proceedings of Computational Linguistics and Intelligent Text Processing: Third International Conference: CICLing-2002, Lecture Notes in Computer Science* 2276. 1–15. [<https://hdl.handle.net/10356/79581>].
- Schlücker, Barbara (ed.). 2018. *Complex Lexical Units. Compounds and Multi-Word Expressions*. Berlin/Boston: De Gruyter.
- Stolz, Thomas. 2003. A new Mediterraneanism. Word iteration in areal perspective. *Mediterranean Language Review* 15. 1–47.
- Stolz, Thomas. 2006a. (Wort-)Iteration: (k)eine universelle Konstruktion. In Kerstin Fischer & Anatol Stefanowitsch (eds.), *Konstruktionsgrammatik. Von der Anwendung zur Theorie*, 105–132. Tübingen: Stauffenburg.

- Stolz, Thomas. 2006b. Europe as a Linguistic Area. In Keith Brown (ed. in chief), *Encyclopedia of Language and Linguistics*, 278–294. Amsterdam: Elsevier.
- Stolz, Thomas. 2007. Das ist doch keine Reduplikation! Über falsche Freunde bei der Suche nach richtigen Beispielen. In Andreas Amman & Aina Urdze (eds.), *Wiederholung, Parallelismus, Reduplikation. Strategien der multiplen Strukturwendung*, 47–80. Brockmeyer: Bochum.
- Stolz, Thomas. 2008. Grammatikalisierung ex nihilo. Totale Reduplikation – ein potentielles Universale und sein Verhältnis zur Grammatikalisierung. In Thomas Stolz (eds.), *Grammatikalisierung und grammatische Kategorien*, 83–109. Brockmeyer: Bochum.
- Stolz, Thomas. 2018. (Non-)Canonical reduplication. In Aina Urdze (ed.), *Non-Prototypical Reduplication*, 201–277. Berlin/Boston: De Gruyter Mouton.
- Van Lier, Eva (ed.). 2023. *The Oxford Handbook of Word Classes*. Oxford: Oxford University Press.
- Wierzbicka, Anna. 1987. Boys Will Be Boys: ‘Radical Semantics’ vs. ‘Radical Pragmatics’, *Language* 63. 95–114.

Bernard Comrie

# Uilta (Orok) attributive adjective order: Grammar, text, typology

**Abstract:** Northern Eurasia, including the Transeurasian (Altaic) languages, is known for its near-exclusive dominant Adjective-Noun constituent order. WALs claims that Uilta, alone among the Transeurasian languages in its sample, has dominant Noun-Adjective order. This article aims to show that Uilta in fact has Adjective-Noun order, like its genealogical and areal neighbors. The article aims not only to correct a questionable claim in the literature, but also to explore a methodology for investigating such claims, by examining the behavior of attributive adjectives in a corpus of Uilta texts, concentrating on adjectives that express descriptive properties.

**Keywords:** Uilta (Orok) language, languages of Northern Eurasia, attributive adjective order, areal typology, corpus linguistics

## 1 Introduction

Northern Eurasia is well-known as the world's largest solid area of languages with dominant constituent order Adjective-Noun (AN) in the noun phrase; see, for instance, Dryer (2005, 2013b), on whose map the presentation in this paragraph is based. The core of this area can be considered to be the languages of the Transeurasian (Altaic) families: Turkic, Mongolic, Tungusic, Koreanic, and Japonic. However, it also includes the small families and language isolates of northern Asia: Yeniseian, Yukaghir, Nivkh, and Ainu; only Chukotko-Kamchatkan on the eastern periphery of the area departs from this pattern by including languages with no dominant order (NDO), a point to which I return in Section 6. To the west, it also includes Uralic, and the Balto-Slavic and Germanic branches of Indo-European. In the west, the boundary between Balto-Slavic and Germanic on the one hand with AN, and Celtic and Romance on the other with Noun-Adjective (NA), is sharp. Elsewhere, one moves from the solid AN area of northern Eurasia into zones with more mixed representation of AN, NA, and NDO before reaching solidly NA areas in Southwest Asia and North Africa and in Southeast Asia, and another solidly AN area in South

---

**Bernard Comrie**, University of California, Santa Barbara, e-mail: comrie@ucsb.edu

Asia south of the Himalayas. It is surprising, then, to find a single NA language, Uilta (Orok), in the middle of Sakhalin island, with AN languages to the north, west, and south (and the Pacific Ocean to the east). All the more surprising when one realizes that Uilta is a Tungusic language, i.e. belongs to one of the Transeurasian families usually considered to form the core of the northern Eurasian AN area. This areal-typological outlier certainly merits further investigation.

Dryer gives as the source of his information on Uilta attributive adjective order Piłsudski (2011).<sup>1</sup> Piłsudski actually prepared two versions of his Uilta grammar sketch, one in Russian and one in Polish. I will start with the Russian version (1) (even though this is not the one to which Dryer 2005, 2013b refers), because it is quite unequivocal.

(1) *Прилагательные*

*Всегда ставится впереди существительного.* (Piłsudski 2011: 688)

Adjectives

Always is placed in front of the noun. [My translation and minor orthographic correction of the original]

Dryer actually refers to the Polish version, or rather to the volume editors' English translation of Piłsudski's Polish text. The original Polish is given in (2).

(2) *Przymiotniki stoją zawsze (?) przed rodzajnikami.* (Piłsudski 2011: 656)

My translation of (2) is given in (3).

(3) Adjectives always (?) stand in front of “rodzajniki”.

Two comments are in order. First, Piłsudski hedges the absolute statement of (1) with a question mark, suggesting that there might also be some instances of adjectives following the noun. But more importantly, the question arises of what Piłsudski means by a “rodzajnik” (plural: *rodzajniki*). This is not the usual Polish word for ‘noun’, which is *rzeczownik*, a word that is used elsewhere by Piłsudski in this sense. In modern Polish linguistic terminology *rodzajnik* means ‘article’ (as in *(in)definite article*), a category that is, however, quite alien to Uilta (as it is to Polish and Russian). The editors of the 2011 volume translate *rodzajniki* as “the words (nouns) they [i.e. the adjectives

---

<sup>1</sup> Dryer actually refers to a 1987 preprint that circulated before the 2011 publication. I have verified that there are no significant relevant differences between the two versions. Note that Piłsudski's original manuscripts are from the early 20th century; see Section 3.

– BCJ determine”, and I accept this interpretation. It means that *rodzajnik*, while not actually denoting ‘noun’, is to be interpreted as ‘noun’ in this context. Now, the full translation of (2) given by the editors is as in (4).

- (4) Adjectives always (?) follow the words (nouns) they determine. (Piłsudski 2011: 656)

It will be noted that while the Polish original (2), cf. my translation in (3), unequivocally says that adjectives precede their “rodzajnik”, i.e. their head noun, the published translation says that they follow. I can only interpret this as an unfortunate translation error, in a work that is undeniably a monumental contribution to Uilta studies. Piłsudski (2011: 656, 688), incidentally, provides an illustrative example ((5) below, in the transcription used in this article), which clearly shows AN order, with the attributive adjective *bərəmi* and the head noun *ulaa*.

- (5) *bərəmi*            *ulaa*  
 well-behaved    reindeer  
 ‘well-behaved reindeer’

In other words, Piłsudski actually claimed that Uilta’s dominant order is AN.

The published version Piłsudski (2011) thus contains an unfortunate mistranslation of the Polish original, and one might simply take note of this, correct the mistranslation, and correct the datapoint in Dryer (2013b). All the more so given that a number of other grammars and grammar sketches of Uilta also identify AN as the dominant order (with none known to me that claim NA), including: Nakanome (1928: 34), Tsumagari (2009: 10), Ozoliņa (2013: 246), Tsumagari and Yamada (2024: 451). Only Ozoliņa (2018: 22), while repeating this general statement, tempers it somewhat by saying that proprietive adjectives usually follow their head; see Section 5.5. However, in correspondence on this Dryer (p.c.) emphasized to me, quite correctly, that the final arbiter of what constitutes the order of attributive adjective and head noun in Uilta should not be what a grammar or grammars say, but rather what documentation of the language shows. Whence the present article.

## 2 The language and its speakers

The language name “Uilta” is based on the indigenous name (*uilta*, variant *ul̥ta*) of the language that in earlier literature is usually referred to as “Orok”. In this, I follow Tsumagari and Yamada (2024), the most recent survey article on the language

and on which this section is based. Uilta is a Tungusic language, most closely related to Ulcha, and more distantly to other members of the Nanaic subgroup; for further information on subgrouping of Tungusic languages, see Janhunen (2024: 7–10).

The traditional home of the Uilta is the central horizontal belt of Sakhalin, with the unrelated (also to each other) languages Nivkh to the north and Ainu to the south. There is a shallow but clear division into Northern and Southern dialects of Uilta. From 1905 to 1945 Sakhalin was divided politically between Russia/the USSR (northern half) and Japan (southern half), with this boundary running between the areas where Northern and Southern Uilta are spoken. The southern half of the island reverted to the USSR/Russian Federation after World War II, and in the following period many speakers of Southern Uilta relocated to Hokkaido (Japan). In the historical period, ethnic Uilta have numbered in the hundreds, although the language is now heavily endangered; Tsumagari and Yamada (2024: 437) suggest that “[t]oday, only a couple of fully fluent speakers survive”.

Ozoliņa (2013) is the most extensive descriptive grammar of Uilta, supplemented by the article Ozoliņa (2018) specifically on attributive constructions, including attributive adjectives. The grammar sketches of Tsumagari (2009) and Tsumagari and Yamada (2024) are especially useful in providing more analysis, while Petrova (1967) remains an important resource. As dictionaries, I have consulted Ozoliņa (2001) and Ozoliņa and Fedjaeva (2003) (for the Northern dialect), and Ikegami (1997) and Magata (1981) (for the Southern dialect); the vocabulary in Ikegami et al. (2008: 89–107) also proved useful, especially in identifying Northern and Southern dialect variants.

Uilta examples are given in the transcription used by Ikegami (1997), which has two practical advantages. First, it corresponds to the transcription used in most of the source examples. Second, it corresponds closely to a broad IPA transcription, with the following exceptions: Palatal affricates are indicated as voiceless *č* and voiced *ǰ*; long segments (vowels and consonants) are indicated by doubling the letter; other combinations of a vowel followed by *i* or *u* indicate falling diphthongs. Examples from other sources are adapted to this system. Uilta has vowel harmony, and upper-case letters are used to indicate archiphonemes.

### 3 Sources

In assembling a corpus of Uilta textual material for the purposes of this investigation, I was guided by the following rather stringent selection principles:

- a) Only published texts were used. My material and analysis can thus readily be checked against the original.

- b) Only prose texts were considered. This was to avoid any possible bias from the influence of metrical or other constraints leading to atypical constituent orders.
- c) Only original texts were included. For the purposes of this investigation, involving as it does the study of a single language, this seems a reasonable constraint.<sup>2</sup> This involved excluding not only translations, but also one text, Pl\_19, that is presented as a kind of phrasebook without clarifying how the Uilta material was collected, i.e. it is possible that the Uilta versions were elicited as translations from some other language. Text Ym\_08 consists of a series of extensive responses by a native speaker to brief prompts in Uilta from the researcher; only the responses were included.
- d) Only transcriptions of oral texts were included. The recent development of Uilta written by native speakers is a major event in the social life of the language, but brings with it possible influences from the new media or from writing practices acquired in other languages. For consistency, therefore, only oral texts were considered.
- e) Duplication of texts was usually avoided. In particular, some of the texts in Ikegami (2002) part 3 are different editions of the same texts that appear in part 1; for details, see (iii) below. I used the texts in part 3, since they are provided with English translations in the publication and are thus more widely accessible. However, while Pl\_02 and Ik\_01\_20, and likewise Pt\_08 and Ik\_01\_17, are different versions of the same story, I included both since there are substantial differences between the two versions of each.

This left a substantial body of texts, as detailed below, sufficient, in my opinion, to investigate reliably the position of attributive adjective relative to head noun in Uilta.

One caveat should, however, be mentioned. Petrova (1967: 55) says:

A characteristic of adjectives in Uilta is their more frequent use in the role of predicate than in the role of attribute. As far as can be judged from the textual material, the Uilta, having once delineated an object in terms of a feature, in what follows do not consider it necessary to repeat these features as an attribute of the given object. An exception is made only for a few words, such as *daaji* 'big', *nuučī* 'small' and some others. [My translation]

I have not investigated whether Uilta does indeed have a low density of attributive adjectives in comparison with other languages, and even if this should turn out to

---

<sup>2</sup> Those interested in the use of parallel translations for purposes of cross-linguistic comparisons will be pleased to note that there is now an Uilta translation of *Le petit Prince* (Saint-Exupéry 2016).

be the case, the number of attributive adjectives identified in the textual material seems sufficient on which to base conclusions.

The following are the texts that have been used as a source of material:

i) Piłsudski (2011): These texts were collected in Sakhalin at the beginning of the 20th century. The published version reflects the efforts of the volume editors in interpreting Piłsudski's materials. The texts are numbered in the volume and I have used this numbering to identify texts using the formula Pl\_xy. In fact, nearly all of the (not particularly numerous) examples are from Pl\_02, by far the longest of Piłsudski's prose texts. In the 2011 volume, lines are numbered, and this numbering has been adopted here following the formula pq (i.e. a sequence of two digits); note that a numbered "line" may in fact contain more than one line of printed text. The general pattern of presentation of examples is: Piłsudski's original Uilta version – he used a Latin-script spelling that is at least in part impressionistic rather than phonemic; then Piłsudski's Polish word-by-word translation; then the editors' English word-by-word translation. This is followed by a Cyrillic version of the text corresponding to the editors' understanding of the original, a free Russian translation of the editors' version, and a free English translation of the same. There are sometimes notes by Piłsudski or by the editors. It should be borne in mind that this invaluable early collection of Uilta material does require philological interpretation. The 2011 volume editors in general follow a conservative policy, adhering as closely as possible to what Piłsudski wrote. For the long text Pl\_02, Tsumagari (2014a) offers an alternative analysis, as Piłsudski's text might be edited by a linguist working with native speakers a century later; this includes Tsumagari's amended version of Piłsudski's original in Latin script, interlinear morphological analysis, glossing using English as the metalanguage, and a free Japanese translation. There are occasional stark discrepancies between the two versions, including some relevant to attributive adjective order. The texts from Piłsudski (2011) included in the corpus are Pl\_01, Pl\_02, Pl\_07, and Pl\_09; some of the other short texts might meet criteria (a)–(e) but do not contain any relevant material.

ii) Petrova (1967): These texts were collected in Leningrad (now: St. Petersburg) in 1936 and 1949 from Uilta students from Sakhalin, speakers of the Northern dialect. The texts are numbered 1–10 in the original. Most of the texts are traditional stories, though some are descriptive of traditional life, and Pt\_10 is a collection of riddles. The texts are identified using the formula Pt\_xy, where "xy" indicates Petrova's text number. The starting line of an example is identified as mn\_pq, where "mn" indicates the number of the paragraph within that text (and "00" indicates the title), "pq" the line number within that paragraph. Each text is presented in the original in Uilta in Cyrillic script without interruption, followed by a free Russian trans-

lation without interruption or (with rare exceptions) annotations; for text Pt\_10, however, each individual riddle is presented separately and immediately followed by its Russian translation.

iii) Ikegami (2002): These texts were audio-recorded from speakers of the Southern dialect in Hokkaido in 1955–1958 and 1966 (parts 1 and 3, with the texts in part 3 supplemented from dictated versions from 1949–1951), and from speakers of both Southern and Northern dialects in Sakhalin in 1990–1992 (part 2). They provide by far the most extensive published original text material in Uilta, primarily traditional stories. The 2002 publication gathers together, with revisions, texts that had previously been published in various outlets. The texts are divided into three parts, with the texts numbered separately within each part. Texts are identified according to the formula Ik\_vw\_xy, where “vw” indicates the part and “xy” the number of the text within that part. For texts in parts 1 and 2, the line at which an example starts is identified by page number and line on that page, using the formula (1)mn\_pq, where “(1)mn” indicates the page and “pq” the line. A “line” consists of the original Uilta text in Latin script followed by a word-by-word translation into Japanese. The texts in part 3 are presented first as continuous text in Uilta in Latin script, with each sentence numbered, then as a free English translation with sentences numbered in the same way. For texts in part 3, the sentence in which an example occurs is identified using the formula mn for the sentence as numbered by Ikegami. Text Ik\_03\_04, a set of riddles, departs somewhat from this convention, and here the formula mn corresponds to the Roman numeral in Ikegami which identifies an individual riddle – all examples are short, so I did not feel it necessary to number individual sentences separately. In all parts, texts are often accompanied by notes. Incidentally, the four texts in part 3 correspond closely to four texts in part 1 (Ik\_01\_12, Ik\_01\_04, Ik\_01\_11, Ik\_01\_19, following the order of texts in part 3), so these texts from part 1 are not included in the corpus. The texts that are included in the corpus are Ik\_01\_01 through Ik\_01\_17, Ik\_01\_20, and Ik\_01\_28; Ik\_02\_01 through Ik\_02\_04 and Ik\_02\_16 (all from speakers of the Northern dialect); and Ik\_03\_01 through Ik\_03\_04. There is also a Russian-language edition of this collection (Ikegami 2007), with the Uilta text in Cyrillic script and a word-by-word (occasionally phrase-by-phrase) translation into Russian. Tsumagari (2009: 18–19) presents an analysis of Ik\_01\_01 with morpheme divisions, interlinear glosses using English as the metalanguage, and a free English translation.

iv) Pevnov (2014): This single text was audio-recorded in 2012 in Sakhalin from a speaker of the Northern dialect. It is a renarration of a dream originally told to the speaker by her mother. The text is identified as Pv. The starting line of an example is identified using the formula mn, where for this purpose a “line” is a block of

Uilta text in Cyrillic script (with indication of morpheme boundaries) accompanied by interlinear glosses using Russian as the metalanguage, and a single continuous Russian translation; it thus occupies several physical lines of actual text. Note that line numbers are not indicated in the original.

v) Yamada (2011a, b, c, 2012, 2014, 2015, 2016, 2023): These texts were audio-recorded in Sakhalin in the early 21st century, from speakers of the Northern dialect. They include traditional stories and descriptions of traditional life, but also discussions of the present day, and some conversational material. Each text was published separately; Table 1 shows correspondences between my text-identifying formula Ym\_xy and the publication. Some of Yamada’s text publications (Ym\_01, Ym\_02, Ym\_04, Ym\_07) contain more than one text, in which case the formula Ym\_xy\_vw is used to identify individual texts. Within each text, the formula mn indicates the number of the line using the numbering of the original, where “line” is a block consisting of original Uilta text in Latin script, interlinear morphological analysis, glossing using English as the metalanguage, and free Japanese and Russian translations.

**Table 1:** Correspondences between text codes and Yamada publications.

Text code	Publication	Text code	Publication
Ym_01	Yamada 2011a	Ym_05	Yamada 2014
Ym_02	Yamada 2011b	Ym_06	Yamada 2015
Ym_03	Yamada 2011c	Ym_07	Yamada 2016
Ym_04	Yamada 2012	Ym_08	Yamada 2023

## 4 Methodology

The present article adopts the same general typological approach to the notion “adjective” as does Dryer (2013b), so it will be useful to quote his characterization of the notion; see (6).

- (6) For the purposes of this map, the term *adjective* should be interpreted in a semantic sense, as a word denoting a descriptive property, with meanings such as ‘big’, ‘good’, or ‘red’. It does not include nondescriptive words that commonly modify nouns, such as demonstratives [ . . . ], numerals [ . . . ], or words meaning ‘other’ [ . . . ]. In some languages, like English, adjectives form a distinct word class. In other languages, however, adjectives do not form a distinct word class and are verbs or nouns [ . . . ].

It should be emphasized that this characterization serves to define the task at hand. It should not be interpreted as a “definition” of the word class “adjective”, indeed the last sentence of (6) explicitly excludes this as a definition of a language-specific word class.

I go, however, somewhat beyond (6) in the direction of using Uilta-specific word classes as follows. If an item is identified as an “adjective” on the basis of (6) and shares properties with a wide range of other such “adjectives”, then I consider this to define a language-specific class of adjectives. If an item does not share these properties, then I exclude it; relevant cases are discussed in Section 5.5. Conversely, if an item does not satisfy (6), or the issue is questionable, then I nonetheless assign it to the language-specific class of adjectives if it shares properties with “adjectives” as defined by (6) and the sentence preceding this one. I thus use a process of bootstrapping from (6) to a language-specific word class characterization.

The first task was thus to establish a subcorpus of examples of combinations of attributive adjective and head noun following (6). This was done by reading through the texts in the corpus, ensuring that I understood them, i.e. not only the meaning but also how the grammatical structures and lexical items combine to convey this meaning, at least with respect to data relevant to the enterprise. I was then able to extract clear instances of combinations of attributive adjective and head noun, plus less clear cases that might require further investigation. Each individual combination of attributive adjective and head noun identified in the texts constitutes one token. In cases where two (or more) adjectives are attributes to a single head noun, this constitutes two (or more) tokens, i.e.  $[[A_1 \text{ and } A_2] N]$  is considered to be two tokens,  $[A_1 N]$  and  $[A_2 N]$  – note that the formulae given here do not presuppose any particular constituent order. Likewise, if two (or more) nouns are qualified by the same adjective, this constitutes two (or more) tokens, i.e.  $[[A [N_1 \text{ and } N_2]]]$  is considered to be two tokens,  $[A N_1]$  and  $[A N_2]$ . This establishes a list of tokens. All tokens consisting of the same combination of attributive adjective and head noun, and having the same syntactic construction, including constituent order (see below), are considered a single type. Needless to say, the actual process of working through the data involved numerous attempts at shortcuts (some successful, some less so), much backtracking, and useful hints from the existing literature on Uilta, both grammatical and lexical. Since the extraction of examples was done manually, some omissions surely occurred inadvertently, but this should not affect the overall picture.

On the basis of this methodology, I established four candidate constructions for combinations of attributive adjective and head noun, as follows. The formula for each construction is closed with an asterisk (\*), to distinguish these four specific constructions from the more general distinction between whether the adject-

tive precedes or follows the noun. In these formulae, and more generally in what follows, the order of constituents is important.

(i) AN\*: This is simply an attributive adjective followed by a head noun, as in (7).

- (7) (Ik\_01\_10 42\_06)  
*daaji bəjə*  
 big bear  
 'big bear'

While most such examples of this construction consist of just an adjective and a noun, it is also possible for the overall noun phrase to contain other constituents, e.g. demonstratives, numerals, or possessors. The adjective is usually a single word, although there are numerous examples of adjectives accompanied by intensifiers, the most frequent being *jij* 'very'. Other instances of branching attributive adjective phrases are rare, except in the case of proprietive adjectives – see Section 5.5. Complex heads are also occasionally found, as in (17), where the attribute *sagji* 'old' qualifies the complex head *uilta əkkə-səl-ni* 'Uilta women', lit. 'Uilta's women'. Construction (i) is by far the most frequent type. Constructions (ii)–(iv) are not found in the subcorpus with as much variation in terms of optional additions as is construction (i), but this may of course be a feature of the limited corpus. Indeed, several questions that can be answered from the subcorpus or the literature for construction (i) remain unanswered for the other constructions.

The noun phrase, and thus the head noun, can be marked for plurality – this is more frequent for nouns higher in animacy – and will be marked for case according to its syntactic-semantic role in the clause. Where appropriate, the head noun will also be marked for possessor, as Uilta has a head-marking possessive construction as illustrated in (8); the possessive suffix follows the case suffix (if any).<sup>3</sup> Note that the possessive suffix can be 3SG even if the possessor is plural (Tsumagari 2009: 6).

- (8) (Ik\_01\_12 49\_05)  
*namu kira-tai-ni*  
 sea shore-ALL-3SG  
 'to the shore of the sea'

<sup>3</sup> This is actually the inalienable possessive construction. Uilta also has an alienable possessive construction, with the suffix *-ju* before any case suffix and the possessor suffix. For further details on Uilta possessive constructions, see Ozoliņa (2013: 118–133).

Agreement of attributive adjectives with their head noun is complex in Uilta. Ozoliņa (2013: 103, 111, 247–248), and earlier Petrova (1967: 55), discuss the situation in the Northern dialect, whereby an attributive adjective may be marked for plural if its head noun is marked for plural, and an attributive adjective may be marked for accusative case if its head noun stands in the accusative. There is no agreement for other cases or for possessor. Tsumagari (2009: 10) says that in the Southern dialect there is no agreement of attributive adjectives.

(ii) NA\*: This is simply a head noun followed by an attributive adjective, as in (9).

- (9) (Pt\_05 01\_08)  
*nooni nəktə nari, moomi, masi*  
 3SG low person thick strong  
 ‘He was a short person, thickset and strong’

The noun phrase resolves to three tokens, *nəktə nari*, *nari moomi*, and *nari masi*, of which the first shows the construction AN\*, the other two NA\*. Clear instances of NA\* are rare, and a relevant factor in (9) seems to be that the adjectives form a list, apparently leading to greater likelihood of postposing some of them.

(iii) NA-ni\*: In this construction, the attributive adjective follows the head noun and takes the 3SG possessor suffix *-ni*, as in (10).

- (10) (Pt\_09 00\_01)  
*ulaa masi-ni*  
 reindeer strong-3SG  
 ‘strong reindeer’

This construction is mentioned in the literature, e.g. Petrova (1967: 151), Tsumagari (2009: 10), Tsumagari and Yamada (2024: 451), but is usually excluded from consideration on the grounds that it is a nominalization. Petrova suggests a literal translation of the type ‘the reindeer’s strength’, but the resulting noun phrase does not denote an abstract quality, so it seems better to go with Tsumagari and a literal translation of the type ‘the reindeer’s strong one’. In Uilta adjectives can be used in headless noun phrases, so that an adjective like *masi* ‘strong’ can participate in a noun phrase *masi* meaning ‘strong one’. Here, there is no explicit nominalization, and no need to assume covert nominalization once one recognizes the existence of headless noun phrases. The precise function of construction (iii) is unclear from the limited subcorpus examples, but it seems to have a delimiting function of the type ‘the/a strong one from among the reindeer’.

If we follow (6), then there seems to be no reason not to consider this construction a combination of attributive adjective and head noun. The A constituent comes from the same set of lexical items as can be A in construction (i). The meaning is to attribute the property A to the referent N, and while literal translations of the kind given in the previous paragraph can be constructed, the most natural translation is usually simply ‘strong reindeer’. It is even possible that the analysis as a nominalization is only historically valid. I therefore include this construction.

I exclude, however, examples where the second component is overtly marked by derivational morphology as a noun, as in (11), since while *ulinga* ‘good’ is an adjective, *ulinga-hta* ‘good thing’ is not – see the entries *ulinga* and *ulingakta* in Ikegami (1997: 219).

(11) (Pt\_06\_07\_06)

*pattə-l ulinga-hta-č'i*  
 seal-PL good-NMLZ-3PL  
 ‘best seals’

(iv) NA-niN\*: This construction is like construction (iii) but with repetition of the head noun at the end, as in (12).

(12) (Ik\_01\_17\_66\_07)

*gasa daaji-ni gasa*  
 bird big-3SG bird  
 ‘big bird’

The structure may be an apposition involving type (iii) followed by a copy of the head noun, i.e. ‘the big bird, the bird’; the adjective thus goes with the preceding N, not with the following one. This construction is only found in Ikegami’s material, and the small number of attestations makes it difficult to discern a specific function.

In Section 5, then, I will work with these four types, one of which is straightforwardly AN, the other three all instances of NA, whether just this bare combination, or together with a possessive suffix, or together with a possessive suffix and repetition of the head noun.

Once the subcorpus of examples of attributive adjective plus head noun has been constructed, the next step is to count how many instances there are of each of the four constructions listed above, although for the purposes of establishing dominant order only the binary distinction between construction (i) on the one hand (AN) and (ii)–(iv) on the other (NA) is relevant. In then establishing whether Uilta is AN, NA, or NDO, I follow the heuristic rule in (13) proposed by Dryer (2013a).

- (13) The rule of thumb employed is that if text counts reveal one order of a pair of elements to be more than twice as common as the other order, then that order is considered dominant, while if the frequency of the two orders is such that the more frequent order is less than twice as common as the other, the language is treated as lacking a dominant order for that pair of elements. For sets of three elements, one order is considered dominant if text counts reveal it to be more than twice as common as the next most frequent order; if no order has this property, then the language is treated as lacking a dominant order for that set of elements.

Since we are dealing with a pair of elements (attributive adjective and head noun), only the first sentence of (13) is relevant and will be applied in Section 5.

## 5 Data and analysis

This section presents and, where necessary, comments on the data in the subcorpus of combinations of attributive adjective and head noun; the material is presented in table form. The tables are ordered following the order of the four constructions in Section 4, with some further subdivision in the case of construction (i) (AN\*). In each table, the first column gives the adjective, the second column the nouns with which this adjective combines. Uilta lexical items are given in the citation form taken from Ikegami (1997), or using the same transcription system for items not included in Ikegami. Where there is a clear difference between Southern and Northern dialectal forms, this is indicated using the formula “Southern/Northern”; however, variation between the vowel qualities *ø* and *o*, *u* is not indicated, as this represents an ongoing loss of the distinct phoneme *ø* during the documentation of Uilta (Tsumagari and Yamada 2024: 439). There is often variation in indicated vowel length, even within the same source. English translations attempt to capture the general or basic meaning of the adjective, though sometimes a contextually appropriate form is given. Sometimes more than one English translation is given, though this is probably as much due to idiosyncrasies of the English lexicon as to those of Uilta. To avoid undue repetition, *ulaa* is translated in the tables as ‘reindeer’, although more accurately it means ‘domestic reindeer’; *nari* sometimes translates more naturally as ‘man’ than as ‘person’, but in the table I have given only the translation ‘person’. It should be borne in mind that all translations will have passed through Japanese or Russian, although Ikegami (1997) does give English translations for some lexical items. The next column lists the textual occurrences (i.e. of tokens) of each type (i.e. combination of attributive adjective and head noun); different texts are separated

by semicolons, while different occurrences within a text are separated by a comma without repetition of the text identifier. The last two columns give a total of tokens (penultimate column) or types (last column) for the given part of the table. Adjectives are ordered in terms of semantic similarity.

## 5.1 Construction (i) (AN\*)

The adjectives found in this construction are divided into two, with Table 2 showing adjectives with core adjectival semantics, Table 3 those with peripheral adjectival semantics. The dividing line proved to be somewhat subjective, and in case of doubt I followed the devil's advocate principle and assigned an item or group of items to the peripheral category. Adjectives in the core category are usually monomorphemic, at least synchronically, although some may contain the productive suffix *-uli* deriving adjectives related to perception (Ozoliņa 2013: 53). The adjective *irgala* 'patterned' (though English prefers 'spotted' in reference to a seal or its hide, identifying the pattern in question) is derived from the noun *irga* 'pattern', but this is not the usual proprietive suffix *-lu* discussed in Section 5.5.

**Table 2:** Construction (i) (AN\*) with core-semantic adjectives.

Adjective	Noun	Textual occurrences	Frequency	
			Token	Type
<i>daaji</i> 'big, great'	<i>nari</i> 'person'	Ik_01_08 28_17; Pt_04 05_03	2	
	<i>jangee</i> 'official'	Ik_01_17 69_05	1	
	<i>beje</i> 'bear, beast'	Ik_01_10 42_06	1	
	<i>moo</i> 'tree, wood'	Ik_01_06 16_11	1	
	<i>gasa</i> 'village'	Ym_08 51, 52	2	
	<i>duxu/duku</i> 'house'	Ik_01_16 62_06, 63_11; Ik_02_01 108_03	3	
	<i>møluutu</i> 'well'	Ik_01_16 63_12	1	
	<i>tugdala</i> 'bridge'	Ym_05 20	1	
	<i>bajausa</i> PN	Ym_07_01 11	1	
	<i>torisa</i> PN	Ym_07_01 21	1	
<i>kadara</i> 'big'	<i>beje</i> 'bear, beast'	Ik_01_10 42_06	1	
	<i>duxu/duku</i> 'house'	Ik_02_01 108_02	1	
<i>nučiikā</i> 'small'	<i>anduma</i> 'box'	Ik_01_15 60_06, 60_10	2	
	<i>xuldaa</i> 'box'	Ik_01_17 79_11	1	
	<i>abdu</i> 'goods'	Ym_03 03	1	

Table 2 (continued)

Adjective	Noun	Textual occurrences	Frequency	
			Token	Type
<i>jiktu</i> 'thick'	<i>gara</i> 'branch'	Ik_01_06 16_12	1	
<i>moomi</i> 'thick, thickset'	<i>moo</i> 'tree, wood'	Ym_06_21	1	
<i>nəmdə</i> 'thin'	<i>moo</i> 'tree, wood'	Pt_02 03_05	1	
<i>gugda</i> 'high, tall'	<i>nari</i> 'person'	Ik_01_07 20_09	1	
<i>nəktə</i> 'low, short'	<i>nari</i> 'person'	Pt_05 01_08	1	
<i>sagji</i> 'old (aged)'	<i>ulaa</i> 'reindeer'	Ik_01_05 11_08	1	
	<i>sama</i> 'shaman'	Ik_01_09 37_05	1	
	<i>mama</i> 'old woman'	Ik_02_04 116_06, 116_06; Pt_05 01_06	3	
	<i>əəktə</i> 'woman'	Ym_04_01 09	1	
<i>purigə/purəə</i> 'young'	<i>nari</i> 'person'	Ik_01_06 16_01; Ik_01_10 40_06, 41_11, 42_11	4	
	<i>əəktə</i> 'woman'	Ym_07_02 25	1	
<i>aja</i> 'good, beautiful'	<i>nari</i> 'person'	Ym_04_02 25; Ym_07_03 13	2	
	<i>əəktə</i> 'woman'	Ym_07_01 07	1	
	<i>inəŋi</i> 'day'	Ik_01_06 16_05, 16_05, 16_06	3	
	<i>baawui</i> 'k.o. seal'	Ik_01_08 30_04	1	
	<i>namba</i> 'load'	Ym_07_03 08	1	
<i>ulinga</i> 'good'	<i>nari</i> 'person'	Ym_04_01 19	1	
	<i>jakka</i> 'treasure'	Ik_01_16 63_07	1	
	<i>tətuwə</i> 'clothes'	Ik_01_16 63_07, 63_08	2	
	<i>billaatu</i> 'kerchief'	Ik_01_16 63_09	1	
	<i>xuisa</i> 'travel food'	Ym_02_02 13	1	
	<i>orki</i> 'bad, weak'	<i>gewxaatu</i> PN	Ik_01_17 79_07	1
<i>nari</i> 'person'		Pt_03 03_01	1	
<i>dəppi</i> 'food'		Pt_03 02_02	1	
<i>optauli</i> 'delicious'	<i>dəppi</i> 'food'	Ym_02_01 12	1	
<i>saari/sagari</i> 'black'	<i>tətuwə</i> 'clothes'	Ik_01_03 07_04; Ym_01_01 03	2	
	<i>suli</i> 'fox'	Ik_01_03 09_04, 09_08	2	
	<i>xileepu</i> 'bread'	Ym_05 39	1	
<i>pakal</i> 'dark'	<i>ojo</i> 'skin'	Pt_09 09_02	1	
<i>taagda</i> 'white'	<i>ulaa</i> 'reindeer'	Ik_02_03 113_14	1	
	<i>soondoo</i> 'fawn'	Ik_03_03 31	1	
	<i>məə</i> 'water'	Ik_01_17 73_11, 73_12, 73_13, 74_06	4	

Table 2 (continued)

Adjective	Noun	Textual occurrences	Frequency	
			Token	Type
<i>sægda</i> 'red'	<i>sinakta</i> 'hair, fur'	Ik_01_17 71_02, 72_06, 72_14	3	
	<i>nirukta</i> 'hair'	Pt_08 06_11, 06_13	2	
	<i>urakta</i> 'willow branch'	Pl_01 01	1	
	<i>møø</i> 'water'	Ik_01_17 73_10, 73_11	2	
<i>irgala</i> 'patterned'	<i>pæta</i> 'seal'	Pt_06 04_02, 04_05	2	
	<i>natta</i> 'hide'	Pt_06 02_04	1	
<i>kæakku</i> 'empty'	<i>ugda</i> 'boat'	Ik_01_06 16_15	1	
<i>bødø</i> 'fat'	<i>ulaa</i> 'reindeer'	Pv 06	1	
<i>pilji</i> 'sturdy'	<i>nari</i> 'person'	Ik_01_09 32_03	1	
<i>manga</i> 'strong'	<i>nari</i> 'person'	Pt_05 01_05	1	
	<i>sama</i> 'shaman'	Pl_02 67, 75	2	
	<i>æakta</i> 'woman'	Ym_08 60	1	
<i>tanadala</i> 'skillful'	<i>nari</i> 'person'	Pt_05 13_09	1	
<i>iimau</i> 'fresh'	<i>bøjø</i> 'bear, beast'	Ik_01_17 72_08	1	
	<i>sundatta</i> 'fish'	Ym_02_01 02	1	
<i>gægda</i> 'clean'	<i>busu</i> 'cloth'	Ym_02_01 07	1	
<i>nunjuuli</i> 'cold'	<i>xædu</i> 'wind'	Ym_05 08	1	
<i>nama</i> 'warm'	<i>inæji</i> 'day'	Ik_01_06 16_05	1	
<i>namauli</i> 'warm'	<i>inæji</i> 'day'	Ik_01_06 16_06	1	
<i>ajakta</i> 'fierce'	<i>køerbø</i> 'male deer'	Pt_05 09_04	1	
	<i>bøjø</i> 'bear, beast'	Pt_09 10_03	1	
<i>bæræmi</i> 'modest'	<i>nari</i> 'person'	Ym_04_02 10	1	
<i>kæalæ</i> 'speedy'	<i>ulaa</i> 'reindeer'	Ik_03_03 48	1	
<i>nilau</i> 'naked'	<i>nari</i> 'person'	Ik_01_02 05_14, 06_04, 06_05, 06_07	4	
<i>baja</i> 'rich'	<i>nari</i> 'person'	Ik_01_03 07_07	1	
<i>jobbee</i> 'poor'	<i>mapa</i> 'old man'	Ik_03_01 55, 56, 65	3	
<i>paskapsuuli</i> 'unusual'	<i>ulaa</i> 'reindeer'	Ik_01_05 11_07	1	
<i>æsærai</i> 'pregnant'	<i>æakta</i> 'woman'	Ik_01_08 27_01	1	
<i>tæda</i> 'true'	<i>tæaluŋu</i> 'legend'	Ik_01_08 27_10, 27_11	2	
Total			107	76

The peripheral adjectives in Table 3 are all derived. The first block contains the suffix *-pčĭ*, which derives adjectives from temporal adverbs (Ozoliņa 2013: 53). The second block contains the suffix *-mA*, which derives adjectives from nouns indicating the material from which something is made (as with 'silk ribbon'), or at least one

of its major constituents (as with ‘soapy water’) (Ozoliņa 2013: 51); for simplicity, I have used the translation ‘of X’. The last example illustrates the simulative suffix *-ŋAčĭ* ‘similar to, -like’ (Ozoliņa 2013: 58) – while there are numerous instances of this formation in the corpus, most are adverbial or predicative.

**Table 3:** Construction (i) (AN\*) with peripheral-semantic adjectives.

Adjective	Noun	Textual occurrences	Frequency	
			Token	Type
<i>goro-pčĭ</i> ‘old (traditional)’	<i>saxuri</i> ‘tale’	Ik_01_13 54_11; Ik_01_16 65_14; Ik_03_01 68	3	
	<i>dooro</i> ‘path’	Ik_01_06 16_15	1	
	<i>sama</i> ‘shaman’	Ik_01_09 37_05	1	
<i>ooro-pčĭ</i> ‘aforementioned’	<i>manga</i> ‘hero’	Ik_01_10 43_02	1	
	<i>mөө</i> ‘water’	Pt_08 10_04	1	
<i>čĭisä-pčĭ</i> ‘former’	<i>saŋa</i> ‘hole’	Pt_08 07_01, 12_01	2	
	<i>duxu/duku</i> ‘house’	Pt_08 08_02	1	
<i>xuja-mä</i> ‘of horn’	<i>nari</i> ‘person’	Pt_10 02_01	1	
<i>talū-ma</i> ‘of birch bark’	<i>ugda</i> ‘boat’	Ik_01_06 16_07, 16_07	2	
<i>orokto-mo</i> ‘of grass’	<i>duxu/duku</i> ‘house’	Ik_01_17 66_02, 66_05, 77_15, 79_05	4	
<i>čaawakta-ma</i> ‘of claw’	<i>pokto</i> ‘track’	Ym_01_02 14	1	
<i>saura-mä</i> ‘of silk’	<i>leentä</i> ‘ribbon’	Ym_03 22	1	
<i>sala-mä</i> ‘of iron’	<i>sabuu</i> ‘chopstick’	Ik_01_07 24_11	1	
	<i>masaari</i> ‘ax’	Ik_01_07 24_14	1	
	<i>nari</i> ‘person’	Ik_01_09 32_03	1	
	<i>koori</i> ‘mythical bird’	Ik_01_17 66_08, 68_11	2	
	<i>uraktä</i> ‘willow branch’	Ik_01_20 89_17, 90_03	2	
	<i>oljĭga</i> ‘hook’	Pl_02 54, 87	2	
	<i>poroxootu</i> ‘steamer’	Ik_01_17 69_08	1	
<i>aĭsi-ma</i> ‘of gold’	<i>xunaaptu</i> ‘ring’	Ik_01_17 73_05, 78_04, 78_15	3	
	<i>xuldaa</i> ‘box’	Ik_01_17 79_11	1	
<i>mäŋu-mä</i> ‘of silver’	<i>xunaaptu</i> ‘ring’	Pt_08 01_11	1	
<i>xulmäu-mä</i> ‘of container’	<i>namba</i> ‘load’	Ym_07_03 08	1	
<i>putta-ŋäčĭ</i> ‘childlike’	<i>nari</i> ‘person’	Ym_06 25	1	
Total			36	24

## 5.2 Construction (ii) (NA\*)

None of the examples of this construction is straightforward, as indicated in the discussion following Table 4, and even so, only core-semantic adjectives are attested, though this could be an artefact of the small number of examples.

**Table 4:** Construction (ii) (NA\*).

Adjective	Noun	Textual occurrences	Frequency	
			Token	Type
<i>daaji</i> ‘big, great’	<i>bəjə</i> ‘bear, beast’	Pl_02 34	1	
<i>nučīikə</i> ‘small’	<i>poo</i> ‘relative’	Ym_07_01 10	1	
<i>moomi</i> ‘thick, thickset’	<i>nari</i> ‘person’	Pt_05 01_08	1	
<i>masi</i> ‘strong’	<i>nari</i> ‘person’	Pt_05 01_08	1	
Total			4	4

The first example, *bəjə daaji*, is the only one consisting solely of a noun followed by an attributive adjective. Tsumagari (2014a: 221) amends this to *bəjə daaji-ni*, but while this may well be correct, accepting it for present purposes would be circular, so we retain the version given by Piłsudski. The second one is more fully as in (14).

(14) (Ym\_07\_01 10)

*poo-l-bi*                      *nučīikə-səl-bə*  
 relative-PL-REFL    small-PL-ACC  
 ‘her own small relatives (i.e. junior kin)’

The plural suffix *-səl* normally goes on nouns, and (Ozoliņa 2013: 32) notes that this plural suffix added to some adjectives also has a nominalizing function. So it is possible that (14) actually consists of two noun phrases in apposition, i.e. ‘relatives, small ones’. The two examples at Pt\_05 01\_08 are those already discussed as example (9). So extracting even 4 examples from the corpus involves a fair amount of application of the devil’s advocate principle.

The text Pl\_02 contains a number of instances of an item, in Piłsudski’s spelling, *nuči* or *nūci*, sometimes written hyphenated to the preceding noun and sometimes written separately, and which he glosses with the Polish lexical item *mały* ‘small’, i.e. identifying the Uilta item as *nuučī* ‘small’. The editorial team of Piłsudski (2011), following their general principle of adhering closely to Piłsudski’s presentation and

analysis, translate this item into English as ‘small’ or ‘little’. However, in his Uilta dictionary, Piłsudski (2011: 386) has an entry *núci*, with a variant *múci*, distinct from the entry *núci* (i.e. *nuučĩ*) ‘small’ (Piłsudski 2011: 385), and with the definition “is added to nouns about which is narrated in tales” [my literal translation], i.e. with no reference to smallness or other descriptive properties. Tsumagari (2014b: 86–90) argues cogently against the analysis of this item as an adjective meaning ‘small’, even leaving aside general doubts concerning the viability of the construction NA\* in Uilta. The following is a summary of his argument (except for my addition of the parallel to *mərgə* ‘wise’). First, the interpretation ‘small’ often does not make sense. Second, the adjective could have been conventionalized as a postposed bound diminutive morpheme, though once again this is often semantically problematic, and this conventionalization as a suffix is not attested in other sources or by contemporary speakers. Note, however, that Ikegami (2002) does show one instance of conventionalization of an adjective when postposed, namely of *mərgə* ‘wise’ as an honorific marker in (15).

- (15) (Ik\_01\_16 64\_03)  
*andaxa mərgə* (or perhaps: *andaxa-mərgə*)  
 guest wise (or perhaps: guest-HON)  
 ‘honored guest’

Third, it is conceivable that this is the third person plural pronoun *noočĩ* ‘they’ in apposition to the preceding noun, though this hypothesis runs into two problems: The difference in vowel quality is unexpected even given Piłsudski’s impressionistic spelling; the plural pronoun would be inconsistent with instances of a preceding noun singular in both form and interpretation. Finally, Tsumagari proposes that *nuci*, as well as its variant *muci* (found after *o* by assimilation), might reflect a basic representation *-ŋu-čĩ* – Piłsudski sometimes fails to distinguish *n* and *ŋ* elsewhere too – ‘AL-3PL’, i.e. ‘their’, but reinterpreted as a clitic topic marker; Tsumagari cites close parallels from work by Ikegami and Magata. Given this, I have not included examples of postnominal *nuci* as instances of the NA\* construction. Even if they were included, this would reflect idiosyncratic ordering of a single adjectival lexical item.

Pevnov’s text provides another example that might be parallel to (9) in terms of providing a list of properties expressed by adjectives, namely (16).

(16) (Pv 05)<sup>4</sup>

*j̄iŋ bara ulaa-l: peemura=ddaa, gəlči=ddəə, karau=ddaa,*  
 very many reindeer-PL dappled=CONC gray=CONC brown=CONC  
*taagda=ddaa*  
 white=CONC  
 ‘very many reindeer: both dappled and gray and brown and white’

However, there are also reasons for thinking that this is a different structure, as is suggested by Pevnov’s translation, with a colon between the noun *ulaa-l* and the string of adjectives, implying at least a different prosody from simple postposition of a string of adjectives.

### 5.3 Construction (iii) (NA-ni\*)

With this construction again, only core-semantic adjectives are attested (Table 5), though again this could reflect the small number of instances.

**Table 5:** Construction (ii) (NA-ni\*).

Adjective	Noun	Textual occurrences	Frequency	
			Token	Type
<i>gugda</i> ‘high, tall’	<i>nari</i> ‘person’	Ik_01_07 20_09, 20_09	2	
<i>aja</i> ‘good, beautiful’	<i>nari</i> ‘person’	Ik_01_03 07_05	1	
<i>manga</i> ‘strong’	<i>nari</i> ‘person’	Ik_01_09 32_02; Pt_07 06_04, 06_07	3	
<i>masi</i> ‘strong’	<i>ulaa</i> ‘reindeer’	Pt_09 00_01, 01_02, 04_01, 05_01, 06_01, 07_01, 09_02, 09_03, 11_02, 11_03, 11_07	11	
<i>ajakta</i> ‘fierce’	<i>bəjə</i> ‘bear, beast’	Pt_09 00_01, 01_03, 02_01, 02_08, 04_01, 05_01, 06_01, 07_01, 07_07, 09_03	10	
Total			27	5

It is noticeable that the vast majority of tokens – 21/27 or 77.8% – are provided by two types found exclusively in one text, where they refer to the two main participants: *ulaa masi-ni* ‘the strong reindeer’ and *bəjə ajakta-ni* ‘the fierce bear’. This is the clearest instance of a substantial discrepancy between token and type frequency.

<sup>4</sup> The concessive particle is frequently used as a conjunctive coordinator ‘and’ (Ozoliņa 2013: 362–363).

## 5.4 Construction (iv) (NA-niN\*)

Only three adjectives are attested in this construction (Table 6), and all belong to the semantic core.

**Table 6:** Construction (ii) (NA-niN\*).

Adjective	Noun	Textual occurrences	Frequency	
			Token	Type
<i>aĵa</i> ‘good, beautiful’	<i>patala</i> ‘girl’	Ik_01_16 63_13	1	
	<i>ǎǎktǎ</i> ‘woman’	Ik_01_17 67_02, 68_12	2	
	<i>inǎŋi</i> ‘day’	Ik_01_17 66_06	1	
<i>daaĵi</i> ‘big, great’	<i>gasa</i> ‘bird’	Ik_01_17 66_07, 68_09	2	
<i>gugda</i> ‘high, tall’	<i>nari</i> ‘person’	Ik_01_06 16_10	1	
Total			7	5

## 5.5 Questionable and excluded examples

A number of items that might at first sight appear to be adjectives, at least following the characterization in (6), should probably not be considered within the remit of the present investigation. I discuss them in this subsection, starting with those that I would most surely exclude and working toward cases that are less clear. All of the excluded and questionable classes have the attributive element before the head noun, either as the only possibility or as the clearly more frequent possibility, so that including these items would only increase the incidence of AN relative to NA. Once again, the devil’s advocate principle is also at work.

Nationality attributive expressions of the type ‘Uilta woman’ are expressed in Uilta as ‘woman of the Uilta’, i.e. using a possessive construction, as in (17).

- (17) (Ym04\_01 09)  
*sagĵi uilta ǎkkǎ-sǎl-ĵi-ni*  
 old Uilta woman-PL-INS-3SG  
 ‘with old Uilta women’

I consider these to be possessive constructions, not attributive adjective constructions. This may seem inconsistent with the treatment of the construction NA-ni\* as an attributive adjective construction, but there is a crucial difference. In an example like (10), *masi* is qua lexical item an adjective. By contrast, in (17) *uilta* is

qua lexical item a noun and cannot be used as an adjective. There is one possible counterexample to this in the corpus, alongside several examples of the construction in (17), namely (18).

- (18) (Ym\_04\_01 12)  
*uilta nari-sal*  
 Uilta person-PL  
 ‘Uilta people’

I assume that this is an instance of apposition between two nouns.

Sex-expressing attributes, comparable to English ‘male’ and ‘female’, precede the noun that they qualify. The most general are *xusə* ‘male’ and *əəktə* ‘female’, though there are also more specific terms used with some animals. However, all of these words also function as nouns, e.g. *xusə* ‘man’, *əəktə* ‘woman’, so collocations like (19) are best analyzed as nouns in apposition.

- (19) (Ik\_01\_10 42\_02)  
*əəktə bəjə*  
 woman bear  
 ‘female bear’

Quantifiers do not seem to form a single formal category, or a set within a single formal category, in Uilta. Numerals precede their head noun, but are excluded from consideration by Dryer’s characterization (6). The quantifier *čipal* ‘all’ never appears with nominal morphology, and is probably an adverb, following Ikegami (1997: 34). The quantifier *bara* ‘much, many’ is more complex. It appears both before and after the noun to which it refers, and in the latter case when relating to a direct object it takes accusative case marking, as does the head noun, as in (20).

- (20) (Pt\_06 07\_08)  
*pəttə-l-bə baram-ba*  
 seal-PL-ACC many-ACC  
 ‘many seals’

I have not found this pattern with any candidate for adjective status in the corpus other than the quantifiers *bara* ‘much, many’ (numerous examples) and *ojuuka* ‘a little, a few’ (one example at Ym\_05 01). I conclude that *bara* and *ojuuka* are either not adjectives, or constitute an irregular subclass of adjectives whose behavior cannot be taken to represent dominant order. Ikegami (1997: 16), however, does consider *bara* and *ojuuka* to be adjectives.

Uilta has a very productive suffix *-lu* deriving proprietive adjectives, with the meaning ‘having X, with X’. Indeed, in predicative position this is the usual Uilta translation equivalent of English ‘to have’, and presumably any noun phrase, however complex its internal structure, can occupy the position X. Thus, although *-lu* is attached to the last word in the base noun phrase, it has scope over the whole of that noun phrase. Note that attributes dependent on the proprietive adjective stand in the instrumental case, as in (22).

(21) (Ik\_01\_17 67\_17)

*lauta-lu jängee*  
 sword-PROP official  
 ‘an official with a sword’

(22) (Pt\_08 11\_02)

*du-ji jili-lu səwəm-bə*  
 two-INS head-PROP idol-ACC  
 ‘a two-headed idol’

The order in clear instances of this construction in the subcorpus is exclusively with the attribute before the head; see Table 7.

**Table 7:** Construction (i) (AN\*) with proprietive adjectives.

Adjective	Noun	Textual occurrences	Frequency	
			Token	Type
<i>bokko-lu</i> ‘having belly’	<i>mama</i> ‘old woman’	Ik_01_15 56_03, 56_06, 57_08, 57_14, 59_06, 61_03	6	
<i>lauta-lu</i> ‘having sword’	<i>jängee</i> ‘official’	Ik_01_17 67_17	1	
<i>jili-lu</i> ‘having head’	<i>kuukku</i> ‘swan’	Pt_07 04_06	1	
	<i>səwə</i> ‘idol’	Pt_08 11_02, 11_08	2	
<i>ananiju-lu</i> ‘having year’	<i>kəərbə</i> ‘male deer’	Pt_09 02_05	1	
	<i>xusə</i> ‘boy’	Ym_07_02 24	1	
<i>meela-lu</i> ‘having soap’	<i>məə</i> ‘water’	Ym_03 10	1	
<i>dausu-lu</i> ‘having salt’	<i>məə</i> ‘water’	Ym_03 11	1	
Total			14	8

This is perhaps surprising, since Ozoliņa (2018: 21) says that proprietive adjectives usually follow their head noun. However, she qualifies this claim (Ozoliņa 2018: 22) by noting that when proprietive adjectives themselves have attributes (i.e. are

branching), then the usual order is for them to precede, while otherwise they usually follow. Of the examples in our subcorpus, only 3 tokens representing 3 types are non-branching (*lauta-lu jangee*, *meela-lu mөө*, and *dausu-lu mөө*), while the other 11 tokens representing 5 types are branching (with the type *bokko-lu mama* accounting for 6 of these tokens). The prevalence of preposed nonbranching proprietive adjectives may therefore be an artefact of the small corpus. Overall, proprietive expressions present a mixed picture. On the one hand, they have a derivational suffix that affects the noun phrase status of their base, so treating them as derived adjectives would be plausible, even if they do not belong to the semantic core. On the other hand, their frequent complex internal structure with branching is atypical for Uilta adjectives. I have not included them in the statistics for constructions (i)–(iv).

Attributes of spatial orientation denote the location of the referent, as illustrated by the English translations of the examples from the Uilta subcorpus in Table 8. In Uilta, those without the suffix *-duma* can certainly be used as nouns, and are usually classified only as nouns by Ikegami (1997), while Ozoliņa (2001) classifies them as both nouns and adjectives. Those with the suffix *-duma* seem to be more clearly adjectives in Uilta. If the attribute of spatial orientation is an adjective, then all these Uilta examples illustrate construction (i) AN\*. If it is a noun, then the exact nature of the construction is unclear; it does not appear to be simple apposition, as this does not make sense semantically, in contrast to examples like (19). Of the types discussed in this subsection, attributes of spatial adjectives are the ones I would be most inclined to classify as adjectives. However, I have not included them in the statistics for constructions (i)–(iv).

**Table 8:** Construction (i) (AN\*) with adjectives denoting spatial orientation.<sup>5</sup>

Adjective	Noun	Textual occurrences	Frequency	
			Token	Type
<i>nauramji</i> ‘front’	<i>ŋinda</i> ‘dog’	Ik_01_03 07_08	1	
<i>xamarree</i> ‘back’	<i>bagji</i> ‘leg’	Pt_01 02_09	1	
<i>uuwu</i> ‘top’	<i>kumulta</i> ‘fur-lined ski’	Ik_03_04 12	1	
	<i>kipərə</i> ‘fish-frying rack’	Ik_03_04 12	1	

<sup>5</sup> A couple of comments are in order concerning the English translations of some of the items in Table 8. Petrova consistently translates *aanjēe* as ‘left’, but this seems to be an error, as only the translation ‘right’ is given by Ozoliņa (2001: 25–26), Ikegami (1997: 1), and the translations in Ikegami (2002). The translation of *gajau* as ‘little finger’ does, however, seem to be correct for the Northern dialect, where it contrasts with *gajau aan-ni* ‘ring finger’, lit. ‘little finger’s older brother’ (Ozoliņa 2001: 55). In the Southern dialect, *gajau* means ‘ring finger’, and ‘little finger’ is a completely different word, *čimutəə* (Ikegami 1997: 33, 63; Magata 1981: 27, 62).

Table 8 (continued)

Adjective	Noun	Textual occurrences	Frequency	
			Token	Type
<i>uuwu-dumə</i> 'upper'	<i>xanʃki</i> 'room'	Pt_08_07_02	1	
<i>pəduu</i> 'bottom'	<i>kumultə</i> 'fur-lined ski'	Ik_03_04_12	1	
	<i>kipəre</i> 'fish-frying rack'	Ik_03_04_12	1	
<i>bajʃee</i> 'opposite (side of)'	<i>namu</i> 'sea'	Ik_01_08_29_06, 09	2	
<i>dəunʃee</i> 'left'	<i>xoldo</i> 'side'	Ik_01_10_43_08	1	
	<i>isal</i> 'eye'	Ik_01_17_67_05, 68_16	2	
	<i>xawani</i> 'armpit'	Ik_01_17_70_08, 71_10	2	
<i>aanʃee</i> 'right'	<i>xoldo</i> 'side'	Ik_01_10_43_07	1	
	<i>gajau</i> 'little finger'	Pt_08_01_11	1	
	<i>ŋaala</i> 'hand'	Pt_08_03_05	1	
	<i>xawani</i> 'armpit'	Ik_01_17_70_07, 71_09	2	
	<i>isal</i> 'eye'	Ik_01_17_67_06, 68_15	2	
<i>ojo-duma</i> 'outer'	<i>duxu/duku</i> 'house'	Pt_08_10_01	2	
Total			23	17

## 5.6 Frequency summary

Table 9 sets out the relative frequencies of the four different constructions, combining Tables 2–6, for both tokens and types, and both excluding and including peripheral-semantic adjectives (though this turns out to be relevant only for the AN\* construction). For ease of comparison, both raw figures and percentages of the relevant total are given.

Table 9: Frequency of different constructions.

	Token				Type			
	Core		Core + Peripheral		Core		Core + Peripheral	
AN*	107	73.8%	143	79.0%	76	84.4%	100	87.7%
NA*	4	2.8%	4	2.2%	4	4.4%	4	3.5%
NA-ni*	27	18.6%	27	14.9%	5	5.6%	5	4.4%
NA-niN*	7	4.8%	7	3.9%	5	5.6%	5	4.4%
Total NA	38	26.2%	38	21.0%	14	15.6%	14	12.3%
Total	145	100.0%	181	100.0%	90	100.0%	114	100.0%

Dryer's rule of thumb given in (13) refers to tokens, so the ratio of AN: NA in Uilta, on the basis of my subcorpus of core-semantic attributive adjectives, is 73.8: 26.2, clearly above the minimum needed to identify AN as the dominant order (66.7: 33.3). Including peripheral-semantic adjectives, or shifting from tokens to types, serves only to increase the ratio in favor of AN. I conclude that the dominant order of attributive adjective and head noun in Uilta is AN.

## 6 Conclusions and prospect

I hope to have shown that Uilta should be classified as an AN language in the terminology of WALS. This has been quite a time-consuming task, and one might well wonder whether it was worth the effort. It may result in a change to the classification of Uilta for feature 87 in WALS, in which case the present article might serve as a footnote to history. But now that Grambank (Skirgård et al. 2023 a, b) is available, it is quite possible that in the future linguists will consult Grambank rather than WALS, and Grambank (feature GB193, with Jay Latache and Jeremy Collins as patrons) classifies Uilta as AN (using WALS terminology), citing Tsumagari (2009) as its source; in a world where Grambank gave this information and WALS gave either the same or no information, I would not have been impelled to write this article.

But there are nonetheless indications that not all is well, or at least clear, at the edges of the vast AN area in northern Eurasia, even if elsewhere within this area, with the exception of Uilta, WALS and Grambank are in agreement. At the eastern edge stands the Chukotko-Kamchatkan family. WALS gives information for three languages of the family, with Itelmen classified as NDO, Chukchi and the closely related Koryak as AN. Grambank includes only Chukchi, but classifies it as “Both orders possible” (BOP; see further below). Both cite Dunn (1999), more specifically p. 161ff (WALS) or p. 292 (Grambank). Reading these pages confirms that both AN and NA are possible orders in Chukchi, but does not include any clear statement that one or the other is dominant. On the basis of the sources cited, at this eastern edge Grambank seems to do better than WALS.

What about the Western edge? Here we find discrepancies within the Uralic family and within the Slavic branch of Indo-European. WALS includes 16 Uralic languages for this feature, all with the value AN. Grambank includes 30 Uralic languages, all AN with two exceptions, since both Eastern Khanty (more specifically: Surgut Khanty) and Livvi are classified as BOP. Eastern Khanty is included in both WALS and Grambank, and both refer to the same source, Filchenko (2007). Filchenko's account is unusually detailed in that it gives percentages for the relative frequency of AN and NA in his corpus: AN 93.5%, NA 6.5% (Filchenko 2007: 132). Faced

with these figures, I would have had no hesitation in calling AN as the dominant order in Eastern Khanty, and the relevant Grambank entry does include a note that AN is more frequent than NA. In order to see why Grambank might have classified Eastern Khanty as BOP, it is necessary to examine the coding definitions for this feature in Grambank, the relevant parts of which are given in (23).

(23) **Summary**

Adnominal property words are also known as “adjectives”, in particular in those languages where they make up a separate word class. We want to include elements that mark ‘adjectival’ function but that might not be described as ‘adjectives’ [. . .] The question concerns the pragmatically neutral order(s).

**Procedure**

Code 1 if most adnominal property words are placed before nouns.

Code 2 if most adnominal property words are placed after nouns.

Code 3 if adnominal property words can be placed before or after nouns, either because both orders are possible or because some adjectives precede and some follow the noun.

Grambank’s term “adnominal property word” is equivalent to WALs’s “adjective”, and for consistency I will continue to use the abbreviations AN (Grambank: 1) and NA (Grambank: 2) in relation to Grambank. It is less clear that Grambank’s 3 can be identified with WALs’s NDO, so I will use a distinct abbreviation BOP. The problem that arises in practice with (23) is determining the dividing line between AN and BOP or between NA and BOP. For each of these pairs, many languages are in principle consistent with either coding; in Eastern Khanty, for instance, it is true that most attributive adjectives are placed before nouns, but it is also true that adjectives can be placed before or after nouns, and this will hold for any language where the ratio of the two orders is neither 100: 0 nor 50: 50, which probably means most languages of the world. Unlike WALs, where Dryer suggests a cut-off point where one possibility occurs two-thirds of the time – see (13) – Grambank provides no guidance, and it remains unclear if different authors have used the same criteria. For Eastern Khanty, WALs seems to do better than Grambank. For the other Uralic language where Grambank has BOP, Livvi, I lack evidence comparable to the statistical evidence provided by Filchenko (2007) for Eastern Khanty, so the issue must for now remain open.

Turning now to Slavic, WALs includes 10 languages, all classified as AN, while Grambank includes 8 (a subset of the WALs sample), of which 5 are classified as AN, 1 as BOP (Ukrainian), 2 as NA (Czech and Polish). The classification of Czech as NA seems to be a coding error, since the source (Naughton 2005: 49) says that attributive adjectives are “normally” placed before their nouns, only “occasionally”

after, which I would interpret as AN, though with recurrence of the already noted problem in dividing AN from BOP in Grambank. For Ukrainian, the source (Pugh and Press 1999: 166) says: “the attributive adjective as a rule precedes, although it may follow when emphatic or for stylistic reasons”. Since emphasis is pragmatically marked, the order used for emphasis should be disregarded according to (23); it is unclear whether the same applies to “stylistic reasons”. My call here would again be AN, though again noting the problem with the dividing line between AN and BPO. Polish is a more complicated case, and certainly has a much higher incidence of NA than does the general run of Slavic languages; further discussion would exceed the bounds of this section and must for now remain open. But in general, for Slavic, and thus for the western edge in general, WALs seems to do better than Grambank.

The moral is clear: The price of typology is eternal vigilance.

## Abbreviations

3	third person
A	Adjective
ACC	accusative
AL	alienable
ALL	allative
AN	Adjective–Noun
BOP	Both orders possible, or some adjectives precede and some follow the noun
CONC	concessive
HON	honorific
Ik	Ikegami (2002)
INS	instrumental
k.o.	kind of
N	Noun
NA	Noun–Adjective
NDO	No dominant order
NMLZ	nominalizer
PI	Piśsudski (2011)
PL	plural
PN	proper name
PROP	proprietary
Pt	Petrova (1967)
Pv	Pevnov (2014)
REFL	reflexive
SG	singular
Ym	Yamada [See Table 1]

## References

Citing sources in different languages and scripts and from different publishing traditions (print, digital, and database) has been challenging. For authors whose last name is cited in the main text, I have adopted a single version of the last name, with the author's name as it appears in the original in square brackets. For titles of books, articles, and journals in languages other than English or German, an English translation is added in square brackets, using the translation given in the original publication if there is one, even if this is not a literal translation of the original; for Uilta titles in Cyrillic script, a version using Ikegami's transcription is added in parentheses. Names of publishers are given as in the original, with an English translation only when provided in the original publication and following the version in that publication.

- Dryer, Matthew S. 2005. Order of adjective and noun. In Matthew S. Dryer, Martin Haspelmath, David Gil & Bernard Comrie (eds.), *The world atlas of language structures*, 354–357. Oxford: Oxford University Press.
- Dryer, Matthew S. 2013a. Determining dominant word order. In Matthew S. Dryer & Martin Haspelmath (eds.), WALS Online (v2020.4) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13950591> (Available online at <http://wals.info/chapter/s6>, Accessed on 2024-11-15.)
- Dryer, Matthew S. 2013b. Order of adjective and noun. In Matthew S. Dryer & Martin Haspelmath (eds.), WALS Online (v2020.4) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13950591> (Available online at <http://wals.info/chapter/87>, Accessed on 2024-11-15.)
- Dunn, Michael John. 1999. *A grammar of Chukchi*. Australian National University PhD thesis.
- Filchenko, Andrey Yuri. 2007. *A grammar of Eastern Khanty*. Rice University PhD dissertation.
- Ikegami, Jirô [池上二良]. 1997. *ウイльта語辞典 [A dictionary of the Uilta language spoken on Sakhalin]*. Sapporo: 北海道大学図書刊行会 [Hokkaido University Press].
- Ikegami, Jirô [池上二良]. 2002. 増訂 ウイльта口頭文芸原文集 [*Uilta oral literature: A collection of texts. Revised and enlarged edition.*] Osaka: 環太平洋の言語 [ELPR Endangered Languages of the Pacific Rim].
- Ikegami, Jirô [Икегами, Дзиро]. 2007. *Сказания и легенды народа уилта [Uilta oral literature: A collection of texts]*. Sapporo: 北海道大学大学院文学研究科 [Graduate School of Letters, Hokkaido University].
- Ikegami, Jirô [Икэгами, Дзиро], E[lena] A. Bibikova [Е. А. Бибикова], L[jubov] R. Katazima [Л. Р. Катазима], S[ir]juko Minato [С. Минато], T[at]jana P. Roon [Т. П. Роон] & I[rina] Ja. Fedjaeva [И. Я. Федяева]. 2008. *Уилта даирису (Uilta dairisu) / Говорим по-уилтински [Let's speak Uilta]*. Yuzhno-Sakhalinsk: Сахалинское книжное издательство [Sakhalin Book Publisher].
- Janhunen, Juha. 2024. Tungusic as a language family. In Alexander Vovin, José Andrés Alonso de la Fuente & Juha Janhunen (eds.), *The Tungusic languages*, 1–18. London: Routledge.
- Magata, Hisaharu [澗瀧久治]. 1981. *ウイльта語辞典 [A dictionary of the Uilta language]*. Abashiri: 北方民俗文化保存協会 [Society for the Preservation of Northern Region Culture and Folklore].
- Nakanome, Akira. 1928. *Grammatik der Orokko-Sprache*. Osaka: Osaka Asiatic Society. [Translated from the Japanese original of 1917.]
- Naughton, James. 2005. *Czech: An essential grammar*. London: Routledge.
- Ozoliņa, Larisa [Озолиня, Л. В.]. 2001. *Орокско-русский словарь [Orok-Russian dictionary]*. Novosibirsk: Издательство СО РАН [Publisher of the Siberian Branch of the Russian Academy of Sciences].
- Ozoliņa, Larisa [Озолиня, Л. В.]. 2013. *Грамматика орокского языка [A grammar of Orok (Uilta)]*. Novosibirsk: Издательство Гео [Geo Publisher].

- Ozoliņa, Larisa [Озолиня, Л. В.]. 2018. О непредикативных синтаксических структурах в тунгусо-маньчжурских языках: атрибутивная конструкция в орокском языке [On non-predicative syntactic structures in the Orok language]. *Языки и фольклор коренных народов Сибири [Languages and Folklore of Indigenous Peoples of Siberia]* 36. 20–32.
- Ozoliņa, Larisa [Озолиня, Л. В.] & I[rina] Ja. Fedjaeva [И. Я. Федяева]. 2003. *Орокско-русский и русско-орокский словарь [Orok-Russian and Russian-Orok dictionary]*. Yuzhno-Sakhalinsk: Сахалинское книжное издательство [Sakhalin Book Publisher].
- Petrova, T[aisija] I. [Петрова, Т. И.]. 1967. *Язык ороков (ульта) [The language of the Oroks (Uilta)]*. Leningrad: Наука [Nauka].
- Pevnov, A[leksandr] M. [Певнов, А. М.]. 2014. Текст на орокском (ульта) языке с комментариями [A text in the Orok (Uilta) language with notes]. In Valentin F. Vydrin [В. Ф. Выдрин] & Natalia V. Kuznetsova [Н. В. Кузнецова] (eds.), *От Бикина до Бамбалумы, из варяг в греки. Экспедиционные этюды в честь Елены Всеволодовны Перехвальской [From Bikin to Vanbaluma, from Varangians to the Greeks. Field-inspired essays in honour of Elena V. Perehval'skaya]*, 55–66. St. Petersburg: Издательство Нестор-История [Nestor-Istoria].
- Piłsudski, Bronisław. 2011. Materials for the study of the Orok / Uilta language and folklore. In Alfred Majewicz (ed.) with the assistance of Larisa V. Ozoliņa, Mikhail D. Simonov, Tatyana Bulgakova, Elżbieta Majewicz, Tatyana P. Roon, Tomasz Wicherkiewicz & Werner Winter, *The collected works of Bronisław Piłsudski, Volume 4: Materials for the study of Tungusic languages and folklore*, 113–745. Berlin: Mouton de Gruyter.
- Pugh, Stefan M. & Ian Press. 1999. *Ukrainian: A comprehensive grammar*. London: Routledge.
- Saint-Exupéry, Antoine de [Антуану зе Сенту Экзупери (Antuaanu je Seentu Ekjuperi)]. 2016. *Нучйкэ принсу (Nučička prinsu) [The little prince]*. Sapporo: 北海道大学大学院文学研究科 [Graduate School of Letters, Hokkaido University]. [Translator: Elena A. Bibikova; Editor: Yoshiko Yamada; Supervisor: Toshiro Tsumagari.]
- Skirgård, Hedvig, Hannah J. Haynie, Harald Hammarström, Damián E. Blasi, Jeremy Collins, Jay Latache, Jakob Lesage, Tobias Weber, Alena Witzlack-Makarevich, Michael Dunn et al. 2023a. Grambank v1.0 (v1.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.7740140> (Available online at <https://grambank.clld.org/>, Accessed on 2024-11-15.)
- Skirgård, Hedvig, Hannah J. Haynie, Damián E. Blasi, Harald Hammarström, Jeremy Collins, Jay J. Latache, Jakob Lesage, Tobias Weber, Alena Witzlack-Makarevich, Sam Passmore et al. 2023b. Grambank reveals global patterns in the structural diversity of the world's languages. *Science Advances* 9. DOI:10.1126/sciadv.adg6175
- Tsumagari, Toshiro. 2009. Grammatical outline of Uilta (revised). *Journal of the Graduate School of Letters, Hokkaido University* 4. 1–21.
- Tsumagari, Toshiro [津曲敏郎]. 2014a. ピウスツキ採集のウイльта語民話テキスト「ネズミの母親とカエルの母親」 [An Uilta folktale text by B. Piłsudski: “A rat mother and a frog mother” with grammatical analysis and Japanese translation]. *北方言語研究 [Northern Language Studies]* 4. 213–231.
- Tsumagari, Toshiro. 2014b. Remarks on the Uilta folktale text collected by B. Piłsudski. *北方人文研究 [Journal of the Center for Northern Humanities]* 7. 83–94. [Revised English version of Japanese original of 2013.]
- Tsumagari, Toshiro & Yoshiko Yamada. 2024. Uilta. In Alexander Vovin, José Andrés Alonso de la Fuente and Juha Janhunen (eds.), *The Tungusic languages*, 436–462. London: Routledge.
- Yamada, Yoshiko [山田祥子]. 2011a. ウイльта語北方言テキスト: 思い出話 2 編 [Uilta texts in Northern dialect: Two recollections]. *北海道民族学 [Hokkaido Journal of Ethnology]* 7. 48–59.

- Yamada, Yoshiko [山田祥子]. 2011b. ウイルタ語北方言テキスト: スルクタの作り方 [Uilta texts in Northern dialect: How to cook *sulukta*]. 北方言語研究 [*Northern Language Studies*] 1. 217–228.
- Yamada, Yoshiko [山田祥子]. 2011c. ウイルタの振楽器「ヨードプ」: 復元工程についてのウイルタ語北方言テキストを中心に [A type of shaken musical instruments *joodopu* of the Uilta: with the text in Uilta Northern dialect representing process of restoration]. 北方人文研究 [*Journal of the Center for Northern Humanities*] 4. 25–49.
- Yamada, Yoshiko [山田祥子]. 2012. ウイルタ語北方言テキスト: ありがとう、池上先生 [Uilta texts in Northern dialect: Thank you, Ikegami-sensei]. 北方人文研究 [*Journal of the Center for Northern Humanities*] 5. 159–172.
- Yamada, Yoshiko [山田祥子]. 2014. ウイルタ語北方言テキスト: アザラシ肉に関する体験談 [An Uilta text in Northern dialect: A story about seal meat]. 北海道民族学 [*Hokkaido Journal of Ethnology*] 10. 104–113.
- Yamada, Yoshiko [山田祥子]. 2015. ウイルタ語北方言テキスト: 人喰いお化けの話 [An Uilta text in Northern dialect: A story about ogre]. 北方言語研究 [*Northern Language Studies*] 5. 261–280.
- Yamada, Yoshiko [山田祥子]. 2016. ギシクタウダ (マリーヤ・ミヘエワ) の生涯: ウイルタ語北方言テキスト [The life of Gisiktauda (Marija Mikheeva): An Uilta text in Northern dialect]. 北方言語研究 [*Northern Language Studies*] 6. 179–201.
- Yamada, Yoshiko [山田祥子]. 2023. ウイルタ語北方言テキスト: お母さんのこと [A text in the Northern dialect of Uilta: Mother]. 北海道言語文化研究 [*Journal of Language and Culture of Hokkaido*] 21. 1–16.



Martin Haspelmath

# Roots and root classes in comparative grammar

**Abstract:** This paper proposes that a root should be defined as a contentful morph that can occur as part of a free form without another contentful morph. This definition can be applied to all languages using the same criteria and is very largely in line with existing usage. Roots are concrete forms with a shape consisting of a contiguous string of segments, so that consonantal skeletons of Semitic type do not fall under the definition. They differ from affixes in that they may occur freely and have contentful meaning, i.e. denote an action, an object or a property. These three root classes are closely related to the word class notions verb, noun and adjective. The definition of *root* proposed here is largely intuitive, but it must be noted that heterosemous root pairs (such as English *hammer* (noun) and *hammer* (verb)) cannot be seen as having “the same root”, but must be treated as sister roots.

**Keywords:** morphology, terminology, root, contentful morph

## 1 Introduction: Roots in general grammar

The term *root* has long been an important technical term in research on grammar, more technical than “word”, but what exactly is a root? This has rarely been discussed, and this short paper is devoted to this topic. It offers a definition of the term for textbook purposes and comparative purposes and discusses some of the conceptual issues that arise. The definition is meant as a contribution to the methodology of comparative grammar, not as an empirical claim. Linguists often use *root* as a comparative concept, and as such the term should have a definition that can be applied uniformly to all languages.

In addition, I briefly discuss the role of the three root classes (object roots, action roots, property roots) for comparative grammar, following my earlier treatment in Haspelmath (2023b). Because of typical associations between semantic classes and discourse functions, the general word class concepts noun, verb and adjective are closely related to the three semantic root classes.

---

**Martin Haspelmath**, Max Planck Institute for Evolutionary Anthropology,  
e-mail: martin\_haspelmath@eva.mpg.de

Finally, I discuss the phenomenon of heterosemous root sets, like *hammer* (noun) and *hammer* (verb), or *dance* (noun) and *dance* (verb) in English. Such root pairs (or larger root sets) have often been described in derivational terms (e.g. by zero-derivation or conversion), but I propose that they should be treated as sister roots that are formally identical (having the same shape) and semantically related.

Some aspects of the current definition may seem unintuitive, but the purpose is to provide a rigorous definition of *root* that is consistent with other concepts and can be applied to all languages using the same criteria.

## 2 Defining the root

The definition of the term *root* that I propose here is given in (1).

(1) **root** (Haspelmath 2023a: 287)

A root is a contentful morph (i.e. a morph denoting an action, an object or a property) that can occur as part of a free form without another contentful morph.

This definition presupposes the terms *morph* and *free form*, which I discussed at greater length in earlier work. Briefly, a morph is a minimal form, i.e. a form that does not consist of other forms (Haspelmath 2020). Colloquially, linguists often use the term *morpheme* in this way, but morphemes are often more abstract, so that they may have different “realizations”, or they may be zero or discontinuous. By contrast, morphs are by definition pairings of a shape (i.e. a sequence of segments) and a meaning. A free form is a form that may occur on its own, isolated from other forms (Bloomfield 1933: 160; Haspelmath 2021: §4).

Morphs are often divided into “lexical” and “grammatical” morphs (i.e. roots vs. markers), but these terms are not very clear. The term “lexical” has four different meanings (Haspelmath 2024b), and the term “grammatical meaning” is not immediately clear either. It thus seems best to define *root* with respect to concrete types of meanings, namely actions, objects and properties. The term *contentful* was adopted from the definition offered by Bauer et al. (2013: 17):

A root is the centre of a word, a lexically contentful morph, either free or bound, which is not further analysable.

Linguists sometimes contrast “function words” with “content words”, and while these terms are not much less vague than “grammatical items” and “lexical items”, I use “contentful” here in the more precise sense “denoting an action, an object or a property”.

The definition in (1) contains a qualifying clause (“that can occur as part of a free form without another contentful morph”) because there are some kinds of morphs with contentful meaning that we would not call roots. In particular, many languages have causative affixes, such as Japanese *-ase* (as in *yom-ase(ru)* ‘make (someone) read’). These can be said to denote an action, but they are not roots. In addition, some languages have affixes with “root-like” meanings such as Bella Coola *-ak* ‘hand’ and *-uc* ‘mouth’ which only occur together with a root, with an instrumental sense (Mithun 1997). Such elements are often called “lexical affixes”, as opposed to affixes with grammatical meanings or functional roles, but they are “lexical” only in the sense that they are contentful.<sup>1</sup> Contentful affixes are thus excluded from root status by the definition in (1) on the grounds that they do not occur on their own without another contentful morph.

Some languages also have compound-like forms with contentful morphs that do not occur in free forms without another contentful morph. For example, elements such as English *geo-* and *socio-* in “neoclassical compounds” are contentful but always occur together with other contentful morphs, so they do not count as roots either.<sup>2</sup>

### 3 How roots have been defined elsewhere

The term *root* has been used widely in linguistics since the 19th century, especially by Indo-Europeanists, following its use in Semitic linguistics.<sup>3</sup> For example, Marouzeau (1961: 194) defined *root* (French *racine*, German *Wurzel*, Italian *radice*) as in (1).

---

1 In Haspelmath (2024b), I distinguish four senses of “lexical”: relating to word-forms, relating to lexemes, relating to the inventorium, and relating to the mental lexicon. A lexeme must be based on a root, i.e. on a contentful morph (Haspelmath 2024a: 68–69), so it is the second sense of “lexical” that would be relevant here. But it seems best to avoid the use of “lexical” in the sense of “contentful” (or one may replace it by “lexemic”, as suggested in Section 5 below).

2 Such elements are often called “(obligatorily) bound roots” in English-language linguistics (e.g. Bauer et al. 2013: 18), but in many languages, all roots must occur with some affixes (e.g. most Russian noun roots, and most Italian verb roots). Thus, the fact that these elements are bound forms is not a remarkable feature from a cross-linguistic point of view. Their crucial property is that they need to combine with another contentful morph.

3 In Indo-European linguistics, the term *root* was originally used in a diachronic sense, referring to elements that are hypothesized to have existed in the deep past. This usage was apparently strongly influenced by Sanskrit, where “roots” play an important role in Pāṇini’s works (e.g. Whitney 1885).

(2) **racine** (*Wurzel / Root / Radice*)

Élément irréductible du mot, obtenu par élimination de tous les éléments de formation discernables, concevable comme caractéristique d'un concept donné et susceptible de figurer, intact ou modifié, dans les diverses formations qui constituent une famille de mots; ainsi l'élément qui apparaît sous la forme \**am-* dans fr. *amour, amitié, amateur*, sous la forme \**aim-* dans *aimer, aimable, aimant* . . .

The term *root* also appears in Nida's (1949: 82) classical introduction to descriptive morphology:

(3) **roots vs. nonroots**

Roots constitute the nuclei (or cores) of all words. There may be more than one root in a single word, e.g. *blackbird, catfish*, and *he-goat*, and some roots may have unique occurrences. For example, the unique element *cran-* in *cranberry* does not constitute the nucleus of any other words, but it occurs in the position occupied by roots; cf. *redberry, blueberry, blackberry*, and *strawberry*. All other distributional types of morphemes constitute nonroots.

These definitions are not very precise, as they make reference to vague notions such as “elements of formation” or “nucleus/core (of a word)”. Still, there is no doubt that these definitions aim at contentful morphs, i.e. minimal forms that denote actions (verb roots), objects (noun roots) and properties (adjective roots). Linguists' intuitions about the meaning of the term *root* are thus fairly clear, though it has proved difficult to provide a definition. (There is only one respect in which different linguists seem to have diverging intuitions: the question whether “roots” can show heterosemy, which will be discussed in Section 6.)

Let us consider a few more textbook definitions of the term *root* from more recent works:

## (4) a. Harley (2006: 288)

**root:** the morpheme conveying the main meaning in a word. In *cats*, *cat* is the root. In *teacher*, *teach* is the root. In *economics* and *economy*, *econom-* is the root.

## b. Lieber (2009: 204)

**root:** the part of a word that is left after all affixes have been removed. Roots may be free bases, as is frequently the case in English, or bound morphemes, as is the case in Latin.

- c. Aronoff and Fudeman (2011: 2)  
A **root** is like a stem in constituting the core of the word to which other pieces attach, but the term refers only to morphologically simple units.
- d. Booij (2012: 29)  
Stems can be either simplex or complex. If they are simplex they are called **roots**. Roots may be turned into stems by the addition of a morpheme. . .

Harley's "main meaning" is of course very vague, and Lieber's "removal of affixes" presupposes that we know what an affix is.<sup>4</sup> But the notion of "affix" remains undefined.<sup>5</sup> In Aronoff and Fudeman's and in Booij's definitions, "root" is defined in terms of "stem", but it is not clear that the notion of "stem" should be more basic than "root".<sup>6</sup>

Linguists most often talk about roots in connection with affixed forms, but in a language where a noun has only a single form (as in Vietnamese, which lacks inflectional affixes), we would nevertheless say that it has noun roots. Thus, it is better not to define roots in terms of affixes, but vice versa (affixes and clitics are two types of bound forms that are not roots; see Haspelmath 2023c).

One of the most sophisticated treatments of the conceptual issues surrounding roots is offered by Mugdan (2015: §4.1), who notes that "minimal signs fall into two classes", roots and affixes, as in the form *book-s*, where everyone agrees that *book* is the root and *-s* is the affix. He first considers the possibility that roots are potentially free and affixes obligatorily bound, but then he notes that Slavic languages

---

4 Gebhardt (2023: xxiv) similarly defines a root as "an unanalyzable, monomorphemic form, free or bound, that's left when all affixes are removed".

5 According to Lieber (2009: 197), an affix is "a bound morpheme that consists of one or more segments that typically appear before, after, or within, a base morpheme", but there is no definition of "base", and as she notes in her definition of "root", roots may be bound forms just as affixes. Gebhardt (2023: xiii) says that an affix is "a bound morpheme, excluding bound roots, that is added to a stem", so he defines "root" in terms of "affix" and vice versa.

6 I propose the following definition, which is based on the "root" notion:

**stem**

A stem is a contiguous segment string that consists of at least one root and possibly some affixes and that can be combined with an affix.

This is very similar to the definition given by Aronoff and Fudeman ("a stem is a base unit to which another morphological piece is attached", 2011: 2), but more precise. Booij's (2012: 29) definition ("the stem of a word is the word form minus its inflectional affixes") is more narrow, as it excludes strings containing an inflectional affix, but we probably want to say, for example, that Latin *laud-ab-* is an "Imperfective Past stem" (consisting of a root and a tense suffix) that can be combined with person suffixes. (See also Mugdan 2015: §4.3 for some enlightening discussion).

such as Sorbian have many words (not only verbs but also nouns) whose root is not free but always occurs with an inflectional affix (e.g. *žon-* ‘wife’: nominative *žon-a*, genitive *žon-y*, accusative *žon-u*, etc.). Thus, Mugdan ends up bringing up the semantic criterion of “lexical” vs. “grammatical” meaning, even though he also notes that it is not always clear how to classify a morph’s meaning:

Another distinction that is frequently appealed to is that between *lexical meaning* and *grammatical meaning*. It is based on a variety of factors (cf. Croft 2000: 258–260); putting it simply, lexical meanings have something to do with entities, properties and states of affairs in the extra-linguistic world, and grammatical meanings are relational or structural. (Mugdan 2015: 257)<sup>7</sup>

We will come back to this criterion in Section 5 below.

## 4 Roots as concrete forms

In (1) above, a root was defined as a kind of morph, i.e. a kind of form (as discussed in Haspelmath 2020). A form is a pairing of a shape and a meaning (or function), and a shape is a contiguous string of segments. In this conception of morphs, there are no zero morphs, no discontinuous morphs, and no replacive morphs. Consider the German inflected verb forms in (5):

(5) some German inflected verb forms

<i>ich lauf-e</i>	‘I run’
<i>wir lauf-en</i>	‘we run’
<i>lauf-Ø!</i>	‘run!’
<i>ich bin ge-lauf-en</i>	‘I have run’
<i>sie lief</i>	‘she ran’

Many linguists have said that there is a “zero morph(eme)” in the imperative form *lauf*, and that there is a “circumfix” in the past participle form *ge-lauf-en*, but if a morph is a form, it cannot be zero or discontinuous. Zero elements are often useful to make a description more abstract and elegant, but they cannot be treated as forms. And “circumfixes” are best described in terms of constructions containing a prefix and a suffix.<sup>8</sup> Finally, cases where an inflected minimal form has a dif-

<sup>7</sup> The contrast between lexical and grammatical meaning is also the basis for Urbanczyk’s (2011) definitions of roots and affixes.

<sup>8</sup> Defining a form as a pairing of a shape and a meaning may not be controversial, but treating the notion of morph (a kind of form) as basic for other concepts is not the only possibility. More com-

ferent pronunciation, like the past tense form *lief* with vowel ablaut (meaningful sound change), are not treated as “replacive morphs” either. The forms *lauf* and *lief* are simply different roots, though they are related by ablaut (more generally, by endophony).

Since there are no discontinuous morphs, the “roots” of Semitic languages are not roots in the sense of the definition in this paper. Consider the four Standard Arabic forms in (6).

(6) Arabic

<b><i>katab-naa</i></b>	‘we wrote’
<b><i>na-ktub-u</i></b>	‘we write’
<b><i>kaatib</i></b>	‘writer’
<b><i>kitaab</i></b>	‘book’

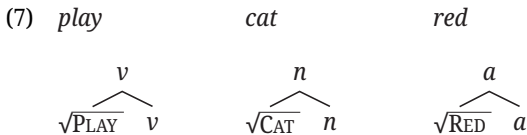
These forms contain four different roots (*katab*, *ktub*, *kaatib*, *kitaab*) that share the consonantal skeleton *k-t-b* and that have different vowel patterns. This kind of morphosyntactic organization could be called “skeleton-and-ablaut”, but a more common term is “root-and-pattern morphology” (e.g. Hudson 1986). It is perhaps somewhat unfortunate that the Semitic triconsonantal skeletons do not fall under the definition of “root” that is proposed here, because the term *root* ultimately derives from Hebrew (and Semitic) linguistics. However, Semitic-type consonantal skeletons are very rare in the world’s languages, and the term *root* is usually used by linguists in the sense given in (1).

Before moving on to root classes in the next section, I should briefly note another use of the term “root” that has become prominent in generative approaches, especially in Distributed Morphology (e.g. Bobaljik 2017; Embick 2021) and the Exoskeletal approach (e.g. Borer 2014; Lohndal 2020). In these approaches, a root is not a kind of minimal form, but an abstract element of the grammar, similar to an inflectional feature that is realized by an affix. Just as “abstract Case” is often written with a capital C in generative grammar (e.g. Chomsky 1981), the term “Root” is sometimes capitalized when referring to an abstract element (e.g. Embick 2021). There is much debate about the nature of abstract Roots (e.g. Harley 2014), but their most salient property is that they do not belong to one of the major word classes (“lexical category”) but are “acategorical”. In order to appear on the surface, they

---

mon is the term “morpheme”, which is often defined abstractly as a kind of “unit” (e.g. Gebhardt (2023: xx): “a morpheme is the smallest unit of meaning and function in a language”). However, the problems that arise in identifying abstract elements such as zeroes (and abstract operations) are well-known, and the concrete definition adopted here is simpler and more appropriate for general and comparative purposes.

must be categorized by a “category-defining head” (*v*, *n* or *a*), as illustrated in (7) (from Embick 2015: 45). This is called the “categorization assumption” (Embick and Marantz 2008: 6).



In the abstract Root view, it is possible to say things such as “some affixes are roots” (Creemers et al. 2018), which would be contradictory for the traditional use of the term *root*. In this paper, I will not consider these approaches further, because they are based on highly specific assumptions about the nature of innate grammatical knowledge that cannot be discussed here.

## 5 Root classes and word classes

Now that we have seen that roots are contentful morphs (Sections 2–3) and that they are concrete forms rather than abstract entities, let me say a bit more about the three classes of roots: action roots, object roots and property roots, which are exemplified in (8), using Maltese examples.

- (8) a. some action roots: *kanta* ‘sing’, *fetaħ* ‘open’, *ħataf* ‘snatch’  
 b. some object roots: *siġra* ‘tree’, *għasfur* ‘bird’, *borża* ‘bag’  
 c. some property roots: *tajjeb* ‘good’, *żgħir* ‘small’, *għani* ‘rich’

Why these three classes? The reason is that they have played an outsize role in many recent discussions of word classes and their diverse cross-linguistic behaviour (e.g. Croft 1991, 2000; Hengeveld and van Lier 2010; Bisang 2023), and there seems to be broad agreement that for the big picture, these are the crucial classes. As emphasized by Croft in various publications, action roots tend to occur frequently in predication function, object roots in reference function, and property roots in modification function (these are “prototypical combinations”; e.g. Croft 2003: §6.4; Baker & Croft 2017: 183), and as a result, they do not need function indicators in these discourse functions (see also Haspelmath 2021: §7.3):

- (9) semantic root classes and discourse functions
- a. action roots: no copula/verbalizer in predication function (e.g. *sing*)
  - b. object roots: no nominalizer/substantivizer in reference function (e.g. *tree*)
  - c. property roots: no relativizer/genitive in modification function (e.g. *good*)

These are very broad generalizations, but they seem to be robust across languages, regardless of the language type or language family. Further more fine-grained generalizations may be possible, e.g. about transitive vs. intransitive action roots, about inanimate vs. animate object roots, about permanent vs. temporary property roots, but these only lead to subclassifications, not to different major classes.<sup>9</sup>

Now how do these root classes relate to word classes in a comparative perspective? In Haspelmath (2023b: 34), I noted that when we consider the way in which the terms *verb*, *noun* and *adjective* are generally used in the literature, we can simply say the following:

- (10) word classes (as comparative concepts)
- a. a verb is an action-denoting root
  - b. a noun is an object-denoting root
  - c. an adjective is a property-denoting root

That is, word classes are not distinct from root classes in a comparative perspective, though the formal properties of these classes do of course differ strikingly across languages. However, regardless of what these formal properties are, action roots will almost always be called “verb”, object roots will almost be called “noun”, and property roots will typically be called “adjectives” (sometimes in seemingly contradictory expressions such as “in this language, adjectives are verbs”; cf. Croft 2022).

There is an enormous literature on word classes across languages (see van Lier (ed.) 2023), but the difficulties of word class assignment that have occupied linguists typically concern larger expressions, e.g. nominalized forms such as English *sing-ing* (which is sometimes called “noun”, even though the root *sing* denotes an action), or derived lexeme-stems such as English *sharp-en* (which is called “verb”, even though the root *sharp* denotes a property). There are many different views

---

<sup>9</sup> Of course, not all meanings expressed by languages fall into these three semantic classes. For meanings such as ‘wave’, ‘hate’, ‘die’, one would need still further categories. However, comparative concepts are typically restricted to a SHARED CORE, and there is little doubt that the meanings seen in (8)–(9) are the core of the meanings expressed by roots in the world’s languages. Thus, the fact that the classes are not exhaustive is not problematic in the present context.

about the word class assignment of such larger expressions, because there is no clear set of cross-linguistically applicable criteria, and it is easy to cherry-pick one's favorite properties for any particular language. Some authors have emphasized that “word class typology” can be done at multiple levels (root, stem, word-form, e.g. Lehmann 2008, 2013; Bisang 2023; or lexical vs. syntactic, e.g. Himmelmann 2008), but this has not led to the kinds of widely recognized results that Croft's approach has given us.

To conclude this section, let us ask how the definition proposed here relates to the intuition that roots denote “lexical meanings” and affixes (or function words) denote “grammatical meanings” (see the quotation from Mugdan in Section 3 above). There is a surprisingly simple answer, given in (11).

(11) **lexical meaning (or lexemic meaning)**

A lexical meaning (or lexemic meaning) is an action meaning, an object meaning, or a property meaning.

This definition may not seem very enlightening, but gaining deeper insight into the difference between lexemic and grammatical meanings is difficult. Perhaps the most sophisticated recent proposal has been summarized as follows:

lexical items (whether morphemes, words, or schematic constructions) are defined as potentially discursively primary; grammatical items, in contrast, are by convention discursively secondary. In other words, lexical items have the potential to express foreground meaning (whether they do so or not depends on context and speaker intentions), while grammatical items can only express their meaning as background information (outside metalinguistic and corrective contexts where conventions may be overridden). (Boye 2023: 81; see also Boye and Harder 2012)

In short, “lexical items” are characterized by “conventionalized discourse prominence”, and this is a feature of all roots in the sense of the definition in (1): All nouns, verbs and adjectives can be focused and occur on their own as answers, while the most typical “grammatical items” are bound forms (clitics or affixes).<sup>10</sup> As noted earlier, the term “contentful form” is better than “lexical form”, because “lexical” has a range of different meanings (Haspelmath 2024b). But while I would not recommend the characterization of roots in terms of “lexical (or lexemic)

---

<sup>10</sup> Note that there are some free forms such as *hello* or *ouch* which do not have lexemic meanings but would probably be said to exhibit conventional discourse prominence. These do not count as roots on the present proposal, which seems to be in line with normal usage.

meaning”, such a definition would actually be compatible with the definition given here.<sup>11</sup>

## 6 Heterosemy: Homonymous root sets with related meanings

In this section, I will address an issue that has loomed large in the literature on roots: the idea that roots can be “precategorial” or “acategorial”, which contradicts what I said above in (10). However, many languages have homonymous root pairs of the following type:

(12) English

- a. *hammer* ‘(the) hammer (i.e. an instrument)’
- b. *hammer* ‘(to) hammer (i.e. hit with a hammer)’

(13) Italian

- a. *strega* ‘(the) witch’
- b. *strega(-re)* ‘(to) bewitch’

(14) French

- a. *combat* ‘(the) fight’
- b. *combatt(-re)* ‘(to) fight’

(15) Russian

- a. *rabota* ‘(the) work’
- b. *rabota(-t’)* ‘(to) work’

Linguists have been uncertain how to deal with such cases, and a variety of approaches have been proposed and discussed. For cases like *hammer/hammer*, derivational analyses in terms of zero-derivation or conversion have been proposed (see Valera 2015), and for languages with required inflectional affixes (like the infinitival suffixes *-re* and *-t’* in Italian/French and Russian), these affixes have sometimes been treated as having a word-deriving function. For Romance lan-

---

<sup>11</sup> The reason why “lexemic meaning” is given as an alternative to “lexical meaning” is that “lexical” has four different meanings, which could be disambiguated by replacing them by “lectic”, “lexemic”, “inventorial”, and “mentalical”, respectively (see Haspelmath 2024b).

guages, there is an earlier tradition of treating nouns such as French *combat* as back-formed (as “postverbal nouns”, e.g. Malkiel 1977).

However, it does not seem to be advisable to adopt a derivational approach, because the direction of derivation is often unclear. Grestenberger and Kastner (2022: 1) formulate the issues that arise in a derivational approach as follows:

- (16) a. Is one of the forms derived from the other, or are both derived from one common ROOT?  
 b. If one is derived from the other, then which is the BASE and which is the DERIVATIVE?

The problem is that there are no cross-linguistically applicable criteria for answering either of these questions. It may sometimes seem evident that one of the two is derived from the other, e.g. the verb *to crown* is said to be derived from the noun *a crown* because it is impossible to say *to crown someone with a hat*. However, the semantic criteria are not robust,<sup>12</sup> and the alternative non-directional view is now perhaps more common.

However, there are (at least) two logically distinct non-directional views:

- (17) a. the precategorial view:  
 Forms such as *hammer* (verb) and *hammer* (noun) are both derived from an abstract unit.  
 b. the heterosemy view:  
 Forms such as *hammer* (verb) and *hammer* (noun) are different roots which are related via sister schemas (e.g. Audring 2019; Jackendoff and Audring 2020a, 2020b) and which can be said to be heterosemous root sets.<sup>13</sup>

The reason why this paper adopts the heterosemy view is that it is based on a conception of comparative concepts which requires that they are defined uniformly across languages, and this is possible only with concrete notions such as meaning, shape, form/expression, bound, as well as general concepts such as class,

---

<sup>12</sup> For example, while it has been argued that “true denominal” instrumental verbs in English are incompatible with instruments that are different from the base noun (*\*taping a picture to the wall with pushpins*, contrasting with *hammering the desk with a shoe*), this does not work consistently (for example, it is possible to talk about *taping a poster to a wall with band-aids*).

<sup>13</sup> The term *heterosemy* for the meaning relationship between words belonging to different word classes was introduced by Lichtenberk (1991).

construction, and inventorium.<sup>14</sup> Abstract “roots” can be posited for purposes of elegant description or analysis, but cannot be identified in a rigorous way in a general (or cross-linguistic) context.

Sister schemas can be illustrated by the sister words *ambition/ambitious* in (18) (from Jackendoff and Audring 2020b), which instantiate the more general sister schema pair in (19) that is also instantiated by word pairs such as *contagion/contagious*.

- (18) a. *ambition*  
 semantics: DESIRE<sub>1</sub>  
 morphosyntax: [N – aff<sub>2</sub>]<sub>1</sub>  
 phonology: /æmbɪʃ<sub>3</sub> ən<sub>2</sub>/<sub>1</sub>
- b. *ambitious*  
 semantics: [HAVING (DESIRE<sub>1</sub>)]<sub>5</sub>  
 morphosyntax: [A – aff<sub>4</sub>]<sub>5</sub>  
 phonology: /æmbɪʃ<sub>3</sub> əs<sub>4</sub>/<sub>5</sub>
- (19) *X-ion*  
 semantics: X<sub>1</sub>  
 morphosyntax: [N – aff<sub>2</sub>]<sub>1</sub>  
 phonology: /Y<sub>3</sub> ən<sub>2</sub>/<sub>1</sub>
- b. *X-ious*  
 semantics: [EXHIBITING (X<sub>1</sub>)]<sub>5</sub>  
 morphosyntax: [A – aff<sub>4</sub>]<sub>5</sub>  
 phonology: /Y<sub>3</sub> əs<sub>4</sub>/<sub>5</sub>

The two schemas are linked relationally because they share some of the same indices, specifically the index 1 for the meaning (‘desire’, or more generally ‘X’) and the index 2 for the shape (/æmbɪʃ/, or more generally /Y/). In much the same way, we can represent the relationship between *hammer* (noun) and *hammer* (verb) as in (20).

---

<sup>14</sup> This also means that concepts such as “word”, “lexeme”, “morphology”, “syntax” and “lexicon” are not needed and are best avoided, because they are associated with stereotypes and are rarely defined properly. (The concept “word” is defined in Haspelmath 2023a, but not in a natural way.) It should be noted that one could alternatively take comparative concepts to be defined differently in different languages, but this makes sense only if they are hypothesized to be innately given (as natural kinds, analogous to the elements in chemistry; see Haspelmath 2018).

- (20) a. *hammer* (noun)  
 semantics: HAMMER<sub>1</sub>  
 morphosyntax: [N -]<sub>1</sub>  
 phonology: /hæmər<sub>2/1</sub>
- b. *hammer* (verb)  
 semantics: [HIT (WITH SOMETHING LIKE HAMMER)]<sub>3</sub>  
 morphosyntax: [V -]<sub>3</sub>  
 phonology: /hæmər<sub>2/3</sub>

On this view, there is neither a derivational claim nor a claim of an abstract root. The identity of the shape is captured by the index 2, and the related meaning is captured by the presence of ‘hammer’ in the semantics in both words. More generally, we can say that English has the two sister schemas in (21).

- (21) a. *X* (noun)  
 semantics: OBJECT (RELATED TO X)<sub>1</sub>  
 morphosyntax: [N -]<sub>1</sub>  
 phonology: /Y<sub>2/1</sub>
- b. *X* (verb)  
 semantics: [DO (IN RELATION TO X)]<sub>3</sub>  
 morphosyntax: [V -]<sub>3</sub>  
 phonology: /Y<sub>2/3</sub>

The precise meaning relationship in these sister schemas is not always predictable, though there are some subschemas with more specific meaning relationships and with high productivity (as discussed in works such as Plag 1999).

Thus, there are good reasons to say that *hammer* (noun) and *hammer* (verb) are two different (homonymous and heterosemous) roots, that neither is derived from the other, and that they are not derived from anything else either (such as an “abstract root”). The question asked by Grestenberger and Kastner (2022) (see 16a above) seems to presuppose that one of these two options must be true (either one is derived from the other, or both are derived from something else), but as we have seen, a non-directional AND non-derivational view is possible and preferable.

It should be noted here that the “acategorical” or “precategorical” view has been seriously discussed or even adopted by some typologists, e.g. Wiltschko (2005) for Salishan languages, Evans and Osada (2005) for Mundari, and Himmelmann (2008) for Tagalog. And intuitively, many linguists (not only typologists) would probably say that English *hammer* (noun) and *hammer* (verb) “share the same root”. On the present proposal, they do not share the same root, but only the same shape. This

has the advantage of being conceptually consistent and of avoiding notions such as “derivation” and “abstract roots”.

Finally, I should mention that in addition to the derivational view and the heterosemy view, there is a third logical possibility, what can be called the “macro-class” view:

(22) the macro-class view:

Elements such as English *hammer* are neither verbs nor nouns but belong to a larger class (“nomiverbs”), and elements such as *clean* are neither verbs nor adjectives but belong to a larger class (“verbectives”), and so on.

For language-particular analyses, this is surely a viable approach in many cases, and especially for “verbectives” (lexemes that correspond to either adjectives or verbs in other languages), this approach has often been adopted. However, as the formal criteria that are used to distinguish between different word classes are typically language-particular, this does not generalize. By contrast, the three root classes discussed in Section 5 show uniform behaviour across languages (Haspelmath 2023b: §2.1), and they can thus be used for defining the notion of root.

## 7 Concluding remarks

In this paper I have provided a definition of the widely used term *root* that is very largely in line with the way it is actually used. The definition makes crucial use of the semantic classes of actions, objects and properties (first highlighted by Croft 1991), and of the notion of a free form (introduced by Bloomfield 1933). A root is defined as a special kind of morph (i.e. a minimal form), which is in line with most linguists’ intuitions, but because a form is required to have a continuous shape, triconsonantal skeletons such as the Arabic *k-t-b* for ‘write’ do not fall under the definition. It should perhaps be noted that roots have not only been discussed by morphologists, but also by phonologists who were interested in their properties compared to the properties of affixes (e.g. Bybee 2005; Urbanczyk 2011). For phonology, it is perhaps clearest that one needs a definition in terms of a minimal form that pairs a meaning with a shape (i.e. a contiguous segment string).

The term “root” became well-known in linguistics in the 19th century, during a period when historical-comparative linguistics made such rapid progress that it seemed possible to some linguists that we might identify the ultimate minimal constituents of which all languages were originally made up. It may be that the fact that minimal contentful morphs were called “roots” contributed to the erroneous

idea that finding the ultimate “roots of language” might be within our reach. But we now know that *root* is no more than a handy short term for discussions of general grammar or language-particular analyses.

Some readers may find the present definition of *root* (as a contentful morph that can occur as part of a free form without another contentful morph) unintuitive, because traditionally, the notion of a free form has not played a big role in basic morphosyntax. However, the purpose of the proposed definition is not to offer an *intuitive* characterization of roots. Instead, its purpose is to make linguists aware that until now, no really good definition has been proposed, and that it is possible to formulate a very good definition on the basis of widely agreed semantic criteria plus the free-form criterion (i.e. that roots can occur on their own, or at most in combination with some nonroots).

It is not clear whether roots might play a significant role in our ultimate understanding of the grammatical properties of human languages (for example, whether they are innately given as part of a “universal grammar” of some kind), but for the time being, it seems best if we have clear definitions of our traditional terms, including *root*.

## References

- Aronoff, Mark & Kirsten Anne Fudeman. 2011. *What is morphology?* 2nd edn. Chichester: Wiley-Blackwell.
- Audring, Jenny. 2019. Mothers or sisters? The encoding of morphological knowledge. *Word structure* 12(3). 274–296.
- Baker, Mark C. & William Croft. 2017. Lexical categories: Legacy, lacuna, and opportunity for functionalists and formalists. *Annual Review of Linguistics* 3. 179–197.
- Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.
- Bisang, Walter. 2023. Levels of analysis and word classes (root, stem, word). In Eva van Lier (ed.), *The Oxford handbook of word classes*, 41–71. Oxford: Oxford University Press.
- Bloomfield, Leonard. 1933. *Language*. New York: H. Holt and Company.
- Bobaljik, Jonathan David. 2017. Distributed Morphology. *Oxford Research Encyclopedia of Linguistics*. Available at [https://static1.squarespace.com/static/586960e5197aea52834230a2/t/5c07fca9898583587235007f/1544027307865/DM\\_ORE.pdf](https://static1.squarespace.com/static/586960e5197aea52834230a2/t/5c07fca9898583587235007f/1544027307865/DM_ORE.pdf)
- Booij, Geert E. 2012. *The grammar of words: An introduction to linguistic morphology*. 3rd edn. Oxford: Oxford University Press.
- Borer, Hagit. 2014. The category of roots. In Artemis Alexiadou, Hagit Borer & Florian Schäfer (eds.), *The syntax of roots and the roots of syntax*, 112–148. Oxford: Oxford University Press.
- Boye, Kasper. 2023. Lexical vs. grammatical words. In Eva van Lier (ed.), *The Oxford handbook of word classes*, 72–84. Oxford: Oxford University Press.

- Boye, Kasper & Peter Harder. 2012. A usage-based theory of grammatical status and grammaticalization. *Language* 88(1). 1–44.
- Bybee, Joan. 2005. Restrictions on phonemes in affixes: A crosslinguistic test of a popular hypothesis. *Linguistic Typology* 9(2). 165–222. (doi:10.1515/lity.2005.9.2.165)
- Chomsky, Noam A. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Creemers, Ava, Jan Don & Paula Fenger. 2018. Some affixes are roots, others are heads. *Natural Language & Linguistic Theory* 36(1). 45–84. (doi:10.1007/s11049-017-9372-1)
- Croft, William. 1991. *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press.
- Croft, William. 2000. Parts of speech as language universals and as language-particular categories. In Petra M. Vogel & Bernard Comrie (eds.), *Approaches to the typology of word classes*, 65–102. Berlin: Mouton de Gruyter.
- Croft, William. 2003. *Typology and universals*. 2nd edition. Cambridge: Cambridge University Press.
- Croft, William. 2022. Word classes in radical construction grammar. In Eva van Lier (ed.), *Oxford handbook of word classes*, 213–230. Oxford: Oxford University Press.
- Embick, David. 2015. *The morpheme: A theoretical introduction*. Berlin: De Gruyter Mouton.
- Embick, David. 2021. The motivation for roots in Distributed Morphology. *Annual Review of Linguistics* 7. 69–88. (doi:10.1146/annurev-linguistics-040620-061341)
- Embick, David & Marantz, Alec. 2008. Architecture and blocking. *Linguistic Inquiry* 39(1). 1–53.
- Evans, Nicholas & Osada, Toshiki. 2005. Mundari: The myth of a language without word classes. *Linguistic Typology* 9(3). 351–390. (doi:10.1515/lity.2005.9.3.351)
- Gebhardt, Lewis. 2023. *The study of words: An introduction*. London: Routledge. (doi:10.4324/9781003030188)
- Grestenberger, Laura & Kastner, Itamar. 2022. Directionality in cross-categorial derivations. *Glossa: A journal of general linguistics* 7(1). (doi:10.16995/glossa.8710)
- Harley, Heidi. 2006. *English words: A linguistic introduction*. Malden, MA: Blackwell.
- Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40(3–4). 225–276. (doi:10.1515/tl-2014-0010)
- Haspelmath, Martin. 2018. How comparative concepts and descriptive linguistic categories are different. In Daniël Van Olmen, Tanja Mortelmans & Frank Brisard (eds.), *Aspects of linguistic variation: Studies in honor of Johan van der Auwera*, 83–113. Berlin: De Gruyter Mouton. (<https://zenodo.org/record/3519206>)
- Haspelmath, Martin. 2020. The morph as a minimal linguistic form. *Morphology* 30(2). 117–134. (doi:10.1007/s11525-020-09355-5)
- Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *Journal of Linguistics* 57(3). 605–633. (doi:10.1017/S0022226720000535)
- Haspelmath, Martin. 2023a. Defining the word. *WORD* 69(3). 283–297. (doi:10.1080/00437956.2023.2237272)
- Haspelmath, Martin. 2023b. Word class universals and language-particular analysis. In Eva van Lier (ed.), *The Oxford handbook of word classes*, 15–40. Oxford: Oxford University Press. (<https://ling.auf.net/lingbuzz/005899>)
- Haspelmath, Martin. 2023c. Types of clitics in the world's languages. *Linguistic Typology at the Crossroads* 3(2). 1–59. (doi:10.6092/issn.2785-0943/16057)
- Haspelmath, Martin. 2024a. Inflection and derivation as traditional comparative concepts. *Linguistics* 62(1). 43–77. (doi:10.1515/ling-2022-0086)

- Haspelmath, Martin. 2024b. Four kinds of lexical items: Words, lexemes, inventorial items, and mental items. *Lexique: Revue en Sciences du Langage* (34), 71–95. (doi:10.54563/lexique.1737)
- Hengeveld, Kees & Eva van Lier. 2010. An implicational map of parts of speech. *Linguistic Discovery* 8(1), 129–156.
- Himmelmann, Nikolaus. 2008. Lexical categories and voice in Tagalog. In Peter Austin & Simon Musgrave (eds.), *Voice and grammatical relations in Austronesian languages*, 247–293. Stanford: CSLI Publications.
- Hudson, Grover. 1986. Arabic root and pattern morphology without tiers. *Journal of Linguistics* 22(1), 85–122. (doi:10.1017/S0022226700010574)
- Jackendoff, Ray & Jenny Audring. 2020a. Relational Morphology: A cousin of Construction Grammar. *Frontiers in Psychology* 11. (<https://www.frontiersin.org/articles/10.3389/fpsyg.2020.02241>)
- Jackendoff, Ray & Jenny Audring. 2020b. *The texture of the lexicon: Relational Morphology and the Parallel Architecture*. Oxford: Oxford University Press.
- Lehmann, Christian. 2008. Roots, stems and word classes. *Studies in Language* 32(3), 546–567. (doi:10.1075/sl.32.3.04leh)
- Lehmann, Christian. 2013. The nature of parts of speech. *STUF: Language Typology and Universals* 66(2), 141–177. (doi:10.1524/stuf.2013.0008)
- Lichtenberk, Frantisek. 1991. Semantic change and heterosemy in grammaticalization. *Language* 67(3), 475–509. (doi:10.2307/415035)
- Lieber, Rochelle. 2009. *Introducing morphology*. Cambridge: Cambridge University Press.
- Lohndal, Terje. 2020. Syntactic categorization of roots. *Oxford research encyclopedia of linguistics*. (doi:10.1093/acrefore/9780199384655.013.257)
- Malkiel, Yakov. 1977. The social matrix of Palaeo-Romance postverbal nouns. *Romance Philology* 31(1), 55–90.
- Marouzeau, J. 1961. *Lexique de la terminologie linguistique: francais, allemand, anglais, italien*. 3rd ed. Paris: Geuthner.
- Mithun, Marianne. 1997. Lexical affixes and morphological typology. In Joan L. Bybee, John Haiman & Sandra A. Thompson (eds.), *Essays on language function and language type: Dedicated to T. Givón*, 357–371. Amsterdam: Benjamins.
- Mugdan, Joachim. 2015. Units of word-formation. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation: An international handbook of the languages of Europe*, vol. Vol. 1, 235–301. Berlin: De Gruyter Mouton. (doi:10.1515/9783110246254-017)
- Nida, Eugene A. 1949. *Morphology: The descriptive analysis of words (2nd edition)*. Ann Arbor: University of Michigan Press.
- Plag, Ingo. 1999. *Morphological productivity: Structural constraints in English derivation*. Berlin: Mouton de Gruyter. (doi:10.1515/9783110802863)
- Urbanczyk, Suzanne. 2011. Root-affix asymmetries. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, 2490–2515. Oxford: Wiley-Blackwell. (doi:10.1002/9781444335262.wbctp0104)
- Valera, Salvador. 2015. Conversion. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-Formation: An International Handbook of the Languages of Europe (Volume II)*, 322–339. Berlin: De Gruyter Mouton. (doi:10.1515/9783110246278)
- van Lier, Eva (ed.). 2023. *The Oxford handbook of word classes*. Oxford: Oxford University Press.
- Whitney, W. D. 1885. The roots of the Sanskrit language. *Transactions of the American Philological Association* 16, 5–29. (doi:10.2307/2935779)
- Wiltschko, Martina. 2005. The syntax of precategorial roots. *Proceedings of the 10th Workshop on the Structure and Constituency of Languages of the Americas*, 245–258. UBCWPL. Vancouver.

Michael Spagnol

# *Taf x'naf?* A closer look at stative verbs in Maltese

**Abstract:** The purpose of this paper is twofold. First, it classifies Maltese stative verbs into four distinct groups based on their morphological characteristics, especially their ability to form the perfect. Second, it demonstrates that, despite their formal differences, these verbs share particular morphological and syntactic traits. Notably, they have an actual present time reference in the imperfect, frequently rely on auxiliary verbs to express tense and modality, and are generally incongruent with progressive constructions, setting them apart from dynamic verbs.

**Keywords:** lexical aspect, stative verbs, morphosyntax, pseudo-verbs, Maltese

## 1 Introduction

This work<sup>1</sup> examines the classification and grammatical behaviour of stative verbs in Maltese. Unlike dynamic verbs that express actions or events, stative verbs denote unchanging conditions, qualities, or relationships. Four distinct categories of Maltese stative verbs are identified: (a) perfectless verbs, (b) pseudo-verbs, (c) semi-perfectless verbs, and (d) verbs with a perfect form. This typology reveals the morphological and syntactic diversity within the class of stative verbs in Maltese. Key characteristics of stative verbs include their dependence on auxiliary verbs like *kien* 'to be' to express tense and modality, and their incompatibility with the progressive aspect. The study sheds light on how Maltese encodes stativity through specific linguistic means, providing insights into the interplay between morphology and syntax.

The rest of the paper is organised into six sections. Section 2 provides an overview of aspectology, situating Maltese stative verbs within a broader framework. Section 3 focuses on verb classification, elaborating on the distinctions between

---

<sup>1</sup> This study revisits and builds upon findings previously discussed by the author (Spagnol 2007, 2009). An earlier version of this research, written in Maltese, was published as Spagnol (2023). The present version offers an expanded analysis of the original with the goal of engaging a wider scholarly community. All examples, unless otherwise attributed, have been formulated by the author.

---

**Michael Spagnol**, University of Malta, e-mail: michael.spagnol@um.edu.mt

states and events and the features that define each category. Section 4 explores the defining characteristics of stative predicates, highlighting their persistence, lack of change, and interpretation in the imperfect. Section 5 presents the main typology of stative verbs in Maltese, identifying four categories, while section 6 establishes morphosyntactic criteria for identifying stative verbs, notwithstanding their formal differences. Finally, Section 7 summarises the key findings of the study and proposes directions for future research.

## 2 Aspectology

Tense and aspect are two primary means of expressing temporality in language. While tense locates an event in time relative to another point (often the moment of speaking), aspect focuses on the event's internal temporal structure. Comrie's (1976: 3) often-cited definition of aspect as "different ways of viewing the internal temporal constituency of a situation" highlights how aspect enables us to perceive a situation either as a complete whole (perfective) or as an ongoing process (imperfective). For instance, in Maltese, sentences (1a) and (2a) show a perfective view, while sentences (1b) and (2b) reflect a progressive, imperfective view.

(1a) *Il-vann għadda mill-mina.*  
 DEF-van pass.PFV.3SG.M from.DEF-tunnel  
 'The van drove through the tunnel.'

(1b) *Il-vann għaddej mill-mina.*  
 DEF-van pass.PROG.3SG.M from.DEF-tunnel  
 'The van is driving through the tunnel.'

(2a) *It-tifel niżel it-taraġ.*  
 DEF-boy descend.PFV.3SG.M DEF-stairs  
 'The boy went down the stairs.'

(2b) *It-tifel nieżel it-taraġ.*  
 DEF-boy descend.PROG.3SG.M DEF-stairs  
 'The boy is going down the stairs.'

Following Binnick (1991, 2012, 2020), this study distinguishes between three types of aspect: grammatical, lexical, and phasal. Grammatical aspect, also termed

*viewpoint aspect* (Smith 1991), is concerned with how the event's temporal structure is framed. It may be expressed through verbal inflectional morphology, as Italian imperfective *lui cucinava* 'he was cooking', and periphrastically, as Dutch progressive *hij is aan het koken* 'he is cooking'. Lexical aspect, sometimes called *situation aspect* (Smith 1991), *Aktionsart* or *Aristotelian aspect* (Binnick 1991), classifies verbs based on intrinsic features like dynamicity (stative, *to own* vs. dynamic, *to eat*), telicity (goal-oriented, *to die* vs. with no defined endpoint, *to laugh*), and durativity (durative, *to sleep* vs. punctual, *to blink*).<sup>2</sup> Lastly, phasal aspect focuses on the internal phases of an event, such as its initiation, progression, or completion, and is often expressed by Maltese verbs such as *beda* 'to start' (*beda jiekol* 'he started eating') or *kompl* 'to continue' (*kompl jaqra* 'he kept on reading').

Maltese has been characterized as an aspect-prominent language, with tense and modal associations stemming from its aspectual nature (Borg 1981: 141; Ingham 1983; Bhat 1999: 171–172, 178). The study of verbal aspect in Maltese has garnered significant attention, particularly following Borg's (1981) pioneering investigation of a set of motion verbs within the localist theoretical framework of the journey paradigm. This foundational work, along with subsequent studies such as Fabri (1995) and Ebert (2000), focus on grammatical aspect. Some of the verbs used to analytically express phasal aspect are examined by Vanhove (1987, 1993), under the broad terminological heading of 'verbal auxiliaries'. Phasal constructions are discussed by Borg and Azzopardi-Alexander (1997), Spagnol (2007), Stolz and Ammann (2007, 2008), who offer more detailed explorations of ingressives and continuatives, and Camilleri (2016), who investigates temporal and aspectual auxiliaries, including pseudo-verbs. Lexical aspect was first systematically addressed by Spagnol (2007, 2009), who demonstrated that the intrinsic temporal properties of verbs play a fundamental role in Maltese's verbal system.<sup>3</sup>

Building upon previous research on Maltese aspectology, the rest of this study concentrates on lexical aspect, with an emphasis on the morphosyntactic properties of stative verbs. It further examines the relationship between lexical aspect and grammatical aspect, tense, and modality.

---

2 The development of the concepts of aspect and *Aktionsart*, along with the terminology related to them, has a complex and often perplexing history. For brief historical accounts, see Binnick (1991: 139–149), Kortmann (1991), and Młynarczyk (2004: Ch.2).

3 Further studies touching on various elements of Maltese's aspectual system include Aquilina (1965, 1973, 1979), Bhat (1999), Binnick (1991), Borg (1988), Comrie (1985, 1991), Cremona (1961, 1964, 1966), Cuvalay-Haak (1997), Grech (1977), Olsen (1997), Saydon (1935), Schinas (1977), Sutcliffe (1936), Thieroff (1995), and Zammit Mangion (1977).

### 3 Classifying verbs

The traditional characterization of verbs as words that denote actions or events is inadequate, not only because many nouns like *ġiri* ‘running’, *qabża* ‘a jump’, and *sparar* ‘shooting’ also convey events, but also because some verbs, as shown in sentences (3)–(6), do not denote actions but rather states.

- (3) *Il-basket jżen nofs kilo.*  
 DEF-bag weigh.IPFV.3SG.M half kilo  
 ‘The bag weighs half a kilo.’

- (4) *It-tifel jixbah lil zitu.*  
 DEF-boy resemble.IPFV.3SG.M OBJ aunt.SG.M  
 ‘The boy resembles his aunt.’

- (5) *It-tifla ma tiflaħx.*  
 DEF-girl NEG feel.well.IPFV.3SG.F.NEG  
 ‘The girl is unwell.’

- (6) *Fejn toqgħod?*  
 Where live.IPFV.2SG  
 ‘Where do you live?’

A fundamental distinction is made between states and events, or between stative predicates and events, which are further divided into different types. Events are divided into two types based on durativity. Durative events, like *studja* ‘he studied’ and *lagħab bil-ballun* ‘he played with the ball’, take time to occur, while punctual events happen instantaneously. According to Comrie (1976: 42), a punctual event is “a situation that does not last in time (is not conceived of as lasting in time), one that takes place momentarily”, such as *sab iċ-ċavetta* ‘he found the key’ and *wasal id-dar* ‘he arrived home’. In Maltese, the distinction between the two types of events becomes evident in expressions involving duration, like *għal siegħa sħiħa* ‘for a whole hour’ and with the durative verb *dam* ‘to take time’, as these indicate prolonged action. This incompatibility with instantaneous predicates is seen in sentences (7)–(8).

- (7a) **Lagħab**            *bil-ballun*        *għal siegħa*    *shiha*.  
 play.PFV.3SG.M with-DEF-ball for hour whole.SG.F  
 'He played with the ball for a whole hour.'
- (7b) *Dam*                            *kwarta*        **jistudja**.  
 take.time.PFV.3SG.M quarter study.IPFV.3SG.M  
 'He spent a quarter of an hour studying.'
- (8a) \***Sab**                        *iċ-ċavetta*    *għal siegħa*    *shiha*.  
 find.PFV.3SG.M DEF-key for hour whole.SG.F  
 'He found the key for a whole hour.'
- (8b) \**Dam*                        *kwarta*        **jasal**                        *id-dar*.  
 take.time.PFV.3SG.M quarter arrive.IPFV.3SG.M DEF-home  
 'He spent a quarter of an hour arriving home.'

Events are further classified based on their telicity, i.e., whether they reach a natural endpoint (telic), such as *il-vapur għereq* 'the ship sank' and *kiteb ittra* 'he wrote a letter', or can stop at any arbitrary point (atelic), for example *lagħab bil-ballun* 'he played with the ball' and *dañak* 'he laughed'. The implications of telicity vary in progressive constructions, as shown in sentences involving the progressive marker *qed*. While *It-tifel kien qed jilgħab bil-ballun* 'The boy was playing with the ball' implies he spent time playing, *It-tifel kien qed jikteb ittra* 'The boy was writing a letter' does not imply the letter was completed, because it is goal-oriented and culminates in a specific outcome.

The situation types mentioned above are summarised in Table 1. It is worth noting that verbs commonly shift from one category to another, and that such shifts in verb meanings often adhere to predictable patterns influenced by context (Vendler 1957; Pustejovsky 1995; de Swart 1998). For instance, while the verb *wiżen* 'to weigh' in sentence (3) above expresses a state, referring to the weight of the bag, in sentences like *il-qabla tiżen lit-trabi* 'The midwife weighs the babies' or *Se niżen il-bagalja* 'I'm going to weigh the suitcase', the verb expresses an action.

**Table 1:** Types of situations.

Situation	Description	Examples	Gloss
stative predicate	state, rather than action or process	<i>jaf</i> <i>jixbah</i> <i>ma jiflahx</i>	to know to resemble to be unwell
durative event	action that takes time	<i>studja</i> <i>ppassiġġa</i> <i>nħasel</i>	to study to stroll to wash oneself
punctual event	action that occurs instantaneously	<i>sab iċ-ċavetta</i> <i>wasal id-dar</i> <i>sfronda</i>	to find the key to arrive home to collapse
telic event	action with an endpoint that must be reached	<i>kiteb ittra</i> <i>ghereq vapur</i> <i>miet</i>	to write a letter to sink (ship) to die
atelic event	action that may stop at any arbitrary point	<i>lagħab bil-ballun</i> <i>dahak</i> <i>tkellem</i>	to play with the ball to laugh to speak

Put differently, the three features, that is dynamicity (change of state), telicity (end, limit or boundary), and durativity (temporal extent), should theoretically define eight distinct possibilities. However, most studies identify between three and six classes of eventualities. Despite some criticism and attempts at reclassification, most discussions of actionality begin with Vendler's (1957) four-way classification, as illustrated in these examples.

State:	<i>It-tifla taf il-Frañċiż</i> 'The girl knows French'
Activity:	<i>It-tifel daqq il-pjanu</i> 'The boy played the piano'
Accomplishment:	<i>It-tifla qasmet it-triq</i> 'The girl crossed the street'
Achievement:	<i>It-tifel rebaħ it-tellieqa</i> 'The boy won the race'

Subsequent studies, such as Dowty (1979), have refined Vendler's framework, categorising these classes using the features [ $\pm$ durative], [ $\pm$ telic], and [ $\pm$ dynamic], as in Table 2. States are unchanging situations that are durative but not goal-oriented. Activities are dynamic actions without a defined endpoint, while accomplishments involve processes with a clear endpoint, characterised as durative, dynamic, and

telic. Lastly, achievements are instantaneous events with a specific result, making them telic and dynamic but not durative.

**Table 2:** Classification of situations by features.

Class	Dynamic	Telic	Durative	Example
States	-	-	+	<i>taf</i> 'know'
Activities	+	-	+	<i>daqq il-pjanu</i> 'play the piano'
Accomplishments	+	+	+	<i>qasmet it-triq</i> 'cross the street'
Achievements	+	+	-	<i>rebaħ it-tellieqa</i> 'win the race'

## 4 States

This section focuses on the first type of situation: states, “that puzzling type in which the role of the verb melts into that of predicate, and actions fade into qualities and relations” (Vendler 1967: 152). The most common stative situations express possession, location, and mental states, encompassing verbs such as *għandu* ‘to have’, *jismu* ‘to be named’, *seta* ‘can’, *ried* ‘to want’, and *xtaq* ‘to wish’. Three key characteristics are central to defining states.

First, lack of change. States, as Lyons (1977: 483) puts it, exist rather than occur, as they remain the same over time without change. Comparing *jaf* ‘to know’ (stative) with *għera/jiġri* ‘to run’ (dynamic), we see that all phases of the situation in (9) are linguistically identical, unlike the phases in (10). While knowing involves a constant mental state, so each moment is indistinguishable from the next, each moment in the running process involves different positions, movements, and energy levels, contributing to a constantly evolving sequence of actions.

- (9) *It-tifel jaf il-verità.*  
 DEF-boy know.IPFV.3SG.F DEF-truth  
 ‘The boy knows the truth.’

- (10) *It-tifel jġri ġirja kuljum.*  
 DEF-boy run.IPFV.3SG.F run.SG.F every day  
 ‘The boy runs every day.’

The second characteristic is persistence. States persist, unlike events, which tend to cease if not actively continued. The phasal verbs *baqa* ‘to remain’ and *kompl* ‘to continue’ highlight this distinction: the former tends to associate with states and sometimes with events, while the latter only fits with events, as sentences (11)–(12) show.

- (11a) *It-tifel baqa’ jixbah lil zitu.*  
 DEF-boy keep.PFV.3SG.M resemble.IPFV.3.SG.M OBJ aunt.3.SG.M  
 ‘The boy continues to resemble his aunt.’

- (11b) *\*?It-tifel kompl jixbah lil zitu.*  
 DEF-boy continue.PFV.3SG.M resemble.IPFV.3.SG.M OBJ aunt.3.SG.M  
 ‘The boy continues to resemble his aunt.’

- (12a) *It-tifel baqa’ joqghod Ghawdex.*  
 DEF-boy keep.PFV.3SG.M live.IPFV.3.SG.M Gozo  
 ‘The boy continues to live in Gozo.’

- (12b) *\*?It-tifel kompl joqghod Ghawdex.*  
 DEF-boy continue.PFV.3SG.M live.IPFV.3.SG.M Gozo  
 ‘The boy continues to live in Gozo.’

Third, the interpretation of verbs changes in the imperfect. Dynamic verbs in the imperfect have a habitual or frequentative reading (13a). Conversely, stative verbs in the imperfect have an actual present time reference, even without explicit temporal context. This is evidenced by their inability to license habitual adjuncts like *kull filghodu* ‘every morning’ (13b–c).

- (13a) *It-tifel idoqq il-pjanu kull filghodu.*  
 DEF-boy play.IPFV.3SG.M DEF-piano every morning  
 ‘The boy plays the piano every morning.’

- (13b) *\*It-tifel jismu Peter kull filghodu.*  
 DEF-boy be.named.IPFV.3SG.M Peter every morning  
 ‘The boy is called Peter every morning.’

- (13c) \**It-tifel jaf il-verità kull filgħodu.*  
 DEF-boy know.IPFV.3SG.M DEF-truth every morning  
 ‘The boy knows the truth every morning.’

## 5 Typology of stative verbs in Maltese

In Maltese, stativity spans a spectrum, ranging from strongly stative verbs like *jaf* ‘to know’ to those with more dynamic characteristics, such as *xtaq* ‘to wish’. In this section, I present a classification of Maltese stative verbs into four main types: (a) perfectless statives, which generally lack a perfect morphological form in standard Maltese; (b) pseudo-verbs, linguistic expressions exhibiting verb-like morphosyntactic properties; (c) semi-perfectless statives, having a perfect form that is usually not used in stative contexts; and (d) stative verbs with a standard perfect form. I will now provide a brief overview of each category and introduce diagnostic tests to differentiate them from dynamic verbs and establish their coherence as a natural class.

### 5.1 Type A | Perfectless verbs

- (14) *It-tifla taf il-verità.*  
 DEF-girl know.IPFV.3SG.F DEF-truth  
 ‘The girl knows the truth.’
- (15) *It-tifel jixbah lil zitu.*  
 DEF-boy resemble.IPFV.3SG.M OBJ aunt.SG.M  
 ‘The boy resembles his aunt.’
- (16) *X'nambihom ir-ritratti?*  
 what-need.IPFV.1SG.3PL DEF-photos  
 ‘What do I need those photos for?’
- (17) *Huma jaħtu għal li ġara.*  
 They blame.IPFV.3PL for what happen.3SG.M  
 ‘They are to blame for what happened.’
- (18) *It-tifel jismu Peter.*  
 DEF-boy be.named.IPFV.3SG.M Peter  
 ‘The boy’s name is Peter.’



*lagħab* ‘to play’, and *deher* ‘to appear’ follow regular patterns in the formation of their perfect forms.

## 5.2 Type B | Pseudo-verbs

- (21) **Għandha** *xagħarha twil.*  
 have.3SG.F hair.3SG.F long  
 ‘She has long hair.’
- (22) *Il-ktieb* **fi** *ħafna stampi.*  
 DEF-book in.3SG.M many pictures  
 ‘The book has many pictures in it.’
- (23) **Bini** *l-ġuħ.*  
 with.1SG DEF-hunger  
 ‘I am hungry.’
- (24) **Għadni** *d-dar.*  
 still.1SG DEF-home  
 ‘I am still at home.’
- (25) *Kemm* **ilek** *hawn?*  
 how.much last.2SG here  
 ‘How long have you been here?’
- (26) *It-tifla* **qisha** *rieqda.*  
 DEF-girl as.if.3SG.F sleepy.3SG.F  
 ‘The girl seems sleepy.’
- (27) *It-tifel* **donnu** *ma jiflaħx.*  
 DEF-boy seem.3SG.M NEG endure.3SG.M.NEG  
 ‘The boy seems unwell.’
- (28) *Ħu-k* **ghoddu** *waqa’.*  
 brother2SG almost.3SG.M fall.PERF.3SG.M  
 ‘Your brother nearly fell.’

- (29) **Ħaqqu**            *aħjar minn hekk.*  
 deserve.3SG.M better than this  
 ‘He deserves better than this.’
- (30) **Moħħni**    *fihom lejl u nhar.*  
 mind.1SG in.3PL night and day  
 ‘They’re always on my mind.’
- (31) *Dejjem*    **ħsiebni**            *hemm.*  
 always thought.1SG there  
 ‘I’m always focused.’
- (32) *Imtlejt*            **kollni**    *trab.*  
 fill.PRF.1SG all.1SG dust  
 ‘I got completely covered in dust.’
- (33) *Wara l-ikel*            *kont*            **nofsni**    *rieqed.*  
 after DEF-food be.PST.1SG half.1SG asleep.SG.M  
 ‘After eating, I was half asleep.’

Verbal expressions like those in (21)–(33) are referred to as *pseudo-verbs*, that is prepositions (*għand-* ‘at’, *fi-* ‘in’, *bi-* ‘with’, *għad-* ‘still’, *il-* ‘ago’), synchronic or historical imperatives (*qis-* ‘measure’, *donn-* ‘consider’, *għodd-* ‘count’), nouns (*ħaqq-* ‘justice’, *moħħ-* ‘mind’, *ħsieb-* ‘thought’), and quantifiers (*koll-* ‘all’, *nofs-* ‘half’) that, in certain respects, behave like verbs. Pseudo-verbs share three key characteristics with verbs. Firstly, they agree with the subject in number, person, and gender. Unlike regular verbs, which achieve agreement through affixes, pseudo-verbs rely on bound pronouns, e.g., *It-tifel ħaqq-u rigal* ‘The boy deserves a gift’, *It-tifla ħaqq-ha rigal* ‘The girl deserves a gift’, and *It-tfal ħaqq-hom rigal* ‘The children deserve a gift.’

Secondly, some of them take negation using the circumfix *ma . . . -x* like regular verbs, e.g., *M’għandux tfal* ‘He doesn’t have children’, *Il-bramel ma fihomx ilma* ‘There’s no water in the buckets’, *Ma ħaqqnix iżjed minn hekk* ‘I don’t deserve more than this’, and *M’ilux wisq hawn* ‘He hasn’t been here long.’ Others take the final *-x* (e.g., *Qisekx mudell* ‘Thinking you’re a model now?’) or *mhux* (e.g., *Illum mhux moħħu hawn* ‘Today his mind isn’t here’).

Thirdly, pseudo-verbs take the first-person singular bound pronoun *-ni*, which is specifically used with verbs (e.g., *fittix-ni* ‘he looked for me’, *ried-ni* ‘he wanted me’) but not with nouns (e.g., *wiċċ-i* ‘my face’, *omm-i* ‘my mother’) or prepositions (e.g., *fuq-i* ‘on me’, *taħt-i* ‘under me’). This is why a distinction is made between

prepositional *fi* and *bi* (e.g., *Saq għal go fi-ja* ‘He drove into me’, *Qed jidhāk bi-ja* ‘He is laughing at me’) and the pseudo-verbs (e.g., *Fi-ni tifel* ‘I’m well-built’, *Bi-ni l-ghatx* ‘I’m thirsty’). Some speakers also distinguish between nominal *moħħ* and *ħsieb* (e.g., *Telagħli musmar fuq moħħ-i* ‘I have a pimple on my forehead’, *Qrali ħsieb-i* ‘He read my thoughts’) and their use as pseudo-verbs (e.g., *Moħħ-ni fit-televixin* ‘My mind is on the television’, *Dejjem ħsieb-ni biex noħroġ* ‘I’m always thinking about going out’).

Pseudo-verbs, therefore, express stative situations related to possession (*għand-*, *bi-*), inclusion (*fi-*, *koll-*), appearance (*donn-*, *qis-*), and mental states (*moħħ-*, *ħsieb-*), and they have an actual present time reference. Their perfect is constructed analytically using the verb *kien* ‘to be’, e.g., *Il-bott kien fih iz-zokkor* ‘The jar had sugar in it’, *Dak il-ħin kienet għadha d-dar* ‘At that moment, she was still at home.’

### 5.3 Type C | Semi-perfectless verbs

- (34) *Attent għax it-taġen jaħraq.*  
 careful.SG.M because DEF-pan burn.IPFV.3SG.M  
 ‘Be careful because the pan is hot.’
- (35) *Kuġinti toqġhod Birkirkara.*  
 cousin.1SG live.IPFV.3SG.F Birkirkara  
 ‘My cousin lives in Birkirkara.’
- (36) *Il-basket jżen nofs kilo.*  
 DEF-bag weigh.IPFV.3SG.M half kilo  
 ‘The bag weighs half a kilo.’
- (37) *Mhux sejjer skola għax ma jiflaħx.*  
 NEG going school because NEG endure.IPFV.3SG.M-NEG  
 ‘He’s not going to school because he’s unwell.’
- (38) *Il-madum jizloq ħafna.*  
 DEF-tiles slip.IPFV.3SG.M much  
 ‘The tiles are very slippery.’
- (39) *Il-ġirien tiegħek jiġu minni.*  
 DEF-neighbors your come.IPFV.3PL from.1SG  
 ‘Your neighbors are related to me.’

- (40) *Tilbes kuluri jgħajtu.*  
wear.IPFV.3.SG.F color.PL shout.IPFV.3PL  
'She wears bright colors.'
- (41) *Il-kutra xxewwek.*  
DEF-blanket itch.IPFV.3SG.F  
'The blanket is itchy.'
- (42) *It-triq tagħmel ma' Marsaxlokk.*  
DEF-road make.IPFV.3SG.F with Marsaxlokk  
'The road falls within the remit of Marsaxlokk.'
- (43) *Ħuk jifhem fil-politika.*  
brother-2SG understand.IPFV.3SG.M in.DEF-politics  
'Your brother is knowledgeable about politics.'
- (44) *Għandek ilsienek iniggeż.*  
have.2SG tongue.2SG sting.IPFV.3SG.M  
'You have a sharp tongue.'
- (45) *Sab xogħol jgħodd għalih.*  
find.PFV.3SG.M job suit.IPFV.3SG.M for-3SG.M  
He found a job that suits him.
- (46) *L-imqass jaqta' ħafna.*  
DEF-scissors cut.IPFV.3SG.M much  
'The scissors are very sharp.'

Verbs in this category, exemplified by sentences (34)–(46) have a perfect form which is, however, not used when the verb is employed in a stative sense. For instance, *ħaraq* 'to burn', *wiżen* 'to weigh', and *għadd* 'to count' as transitive verbs are used in both the perfect (47a), (48a), (49a) and the imperfect (47b), (48b), (49b). However, as intransitive verbs (34), (36), (45), they acquire a stative meaning and, therefore, the perfect form is not used. Instead of the perfect form – e.g., \**It-taġen ħaraq* 'The pan burned,' \**Il-basket wiżen nofs kilo* 'The bag weighed half a kilo,' or \**Ix-xogħol għadd għalih* 'The job suited him'—the construction with *kien* (to be) is used, that is, *It-taġen kien jahraq* 'The pan was hot,' *Il-basket kien jiżen nofs kilo* 'The bag weighed half a kilo,' and *Ix-xogħol kien jgħodd għalih* 'The job suited him.'

- (47a) *Dis-sena ħarqu ħafna murtali.*  
 this-year burn.PFV.3PL many fireworks  
 'This year, they set off many fireworks.'
- (47b) *Kull sena jaħarqu ħafna murtali.*  
 every year burn.IPFV.3PL many fireworks  
 'Every year, they set off many fireworks.'
- (48a) *Qabel ma siefer wiżen il-basket.*  
 before REL travel.PFV.3SG weigh.PFV.3SG.F DEF-bag  
 'Before traveling, he weighed the bag.'
- (48b) *Qabel ma jsiefer jiżen il-basket.*  
 before REL travel.IPFV.3SG weigh.IPFV.3SG.F DEF-bag  
 'Before traveling, he weighs the bag.'
- (49a) *Il-flus għoddejthom darbtejn.*  
 DEF-money count.PFV.1SG.3PL twice  
 'I counted the money twice.'
- (49b) *Il-flus ngħoddhom darbtejn.*  
 DEF-money count.IPFV.1SG.3PL twice  
 'I count the money twice.'

These verbs need to be distinguished from dynamic verbs that refer to general states or universal truths in the imperfect, such as *Ix-xemx titla' mil-Lvant*. 'The sun rises in the East.' and *Is-silġ jinħall fis-šana*. 'Ice melts in the heat.', or typical characteristics of natural entities like *Is-serduq jidden*. 'The rooster crows.' and artificial ones like *L-arloġġ itektek*. 'The clock ticks.' These are not stative verbs, for one thing, because in the imperfect they have a habitual or frequentative reading, as shown by their compatibility with adverbial expressions of habit, e.g., *Is-serduq jidden kull filghodu*. 'The rooster crows every morning.' and *Ix-xemx titla' kuljum mil-Lvant*. 'The sun rises every day in the East.'

## 5.4 Type D | Verbs with perfect forms

- (50a) *In-nanna tħobbu ħafna lin-nannu.*  
 DEF-grandma love.IPFV.3SG.F.3SG.M much to.DEF-grandpa  
 'Grandma loves Grandpa very much.'

- (50b) *In-nanna      ħabbitu                      ħafna    lin-nannu.*  
 DEF-grandma love.PRF.3SG.F.3SG.M much to.DEF-grandpa  
 ‘Grandma loved Grandpa very much.’
- (51a) *Is-sala    tesagħna                      kollha.*  
 DEF-hall hold.IPFV.3SG.F.1PL all  
 ‘The hall has enough space for all of us.’
- (51b) *Is-sala    wesgħetna                      kollha.*  
 DEF-hall hold.PRF.3SG.F.1PL all  
 ‘The hall accommodated all of us.’
- (52a) *Il-ħwienet    jistgħu                      jifthu.*  
 DEF-shops can.IPFV.3PL open.IPFV.3PL  
 ‘The shops can open.’
- (52b) *Il-ħwienet    setgħu                      jifthu.*  
 DEF-shops can.PFV.3PL open.IPFV.3PL  
 ‘The shops could open.’
- (53a) *Ma    jaħmilx                                      id-dizonestà.*  
 NEG tolerate.IPFV.3SG.M.NEG DEF-dishonesty  
 ‘He does not tolerate dishonesty.’
- (53b) *Ma    ħamilx                                      id-dizonestà.*  
 NEG tolerate.PFV.3SG.M.NEG DEF-dishonesty  
 ‘He did not tolerate dishonesty.’
- (54a) *It-tifel    irid                                      ġelat.*  
 DEF-boy want.IPFV.3SG.M ice.cream  
 ‘The boy wants ice cream.’
- (54b) *It-tifel    ried                                      ġelat.*  
 DEF-boy want.IPFV.3SG.M ice.cream  
 ‘The boy wanted ice cream.’
- (55a) *Sieħbek    jixtieq                      isiefer                      magħna.*  
 friend.2SG want.IPFV.3SG.M travel.IPFV.3SG.M with.us  
 ‘Your friend wants to travel with us.’

- (55b) *Sieħbek xtaq isiefer magħna.*  
 friend.2SG want.PFV.3SG.M travel.IPFV.3SG.M with.us  
 ‘Your friend wanted to travel with us.’

Some stative verbs do occur in the perfect (50)–(55), placing them at the opposite end of the stativity spectrum, closer to dynamic verbs. As with the verbs discussed in the previous section, the meaning of a verb can affect its use in the perfect tense. For instance, *ħabb* ‘to love’, when used to express a liking for a type of food (e.g., *Inħobbu l-ghaġin* ‘I like pasta’), is less commonly found in the perfect (?*ħabbejtu l-ghaġin* ‘I liked pasta’). Instead, the construction with *kien*, such as *Kont inħobbu l-ghaġin* ‘I used to like pasta’, is more natural. Similarly, the verb *wesa* ‘to hold, have capacity’ without an attached pronoun is generally not used in the perfect. Compare sentences (51a–b) with the following examples:

- (56a) *Is-sala tesa’ elf ruħ.*  
 DEF-hall hold.IPFV.3SG.F thousand soul  
 ‘The hall can accommodate a thousand people.’
- (56b) *\*?Is-sala wesgħet elf ruħ.*  
 DEF-hall hold.PRF.3SG.F thousand soul  
 ‘The hall accommodated a thousand people.’
- (56c) *Is-sala kienet tesa’ elf ruħ.*  
 DEF-hall be.PRF.3SG.F hold.PRF.3SG.F thousand soul  
 ‘The hall used to accommodate a thousand people.’

## 6 Criteria for stative verb identification

In summary, stative verbs in Maltese can be grouped into four main categories, as shown in Table 3. Although morphologically different, the verbs in these categories behave syntactically in similar ways, particularly in relation to tense, modality, and certain adverbial expressions on one hand, and grammatical aspect on the other. Below, I outline two criteria: the first presents a stative syntactic construction, while the second identifies a context in which stative verbs typically do not occur – namely, the progressive construction formed with *qed* (shortened form of *qiegħed*, active participle of *qagħad* ‘to stay’) and a verb in the imperfect.

**Table 3:** Types of stative verbs in Maltese.

Class	Description	Examples	Translation
perfectless verbs	verbs which, in most varieties of Maltese, do not have a perfect form	<i>jaf</i>	'know'
		<i>ifuħ</i>	'smell good'
		<i>jismu</i>	'be named'
pseudo-verbs	elements that share some morphosyntactic characteristics with verbs	<i>fiħ</i>	'contain'
		<i>ħaqqu</i>	'deserve'
		<i>ghadu</i>	'be still'
semi-perfectless verbs	verbs that have a perfect form but do not use it when the meaning is stative	dynamic <i>żelaq</i> 'to slip' is used in both the perfect:	
		a) <i>It-tifel żelaq fil-bitha</i> and the imperfect:	a) 'The boy slipped in the yard'
		b) <i>It-tifel jżloq spiss fil-bitha</i> while stative <i>żelaq</i> 'to be slippery' is used only in the imperfect	b) 'The boy often slips in the yard'
		c) <i>Il-madam jżloq</i> or	c) 'The tiles are slippery'
		d) <i>Il-madam kien jżloq</i> but not	d) 'The tiles were slippery'
e) * <i>Il-madam żelaq</i>	e) 'The tiles slipped'		
verbs with perfect forms	unlike other statives, these verbs have a perfect form	<i>ħabb</i>	'love'
		<i>seta'</i>	'can'
		<i>xtaq</i>	'wish'

## 6.1 Criterion 1: The use of the verb *kien*

At the heart of the first criterion, which distinguishes stative from dynamic verbs in Maltese, lies the verb *kien* 'to be', and occasionally *ġie* 'to come' or *sar* 'to become', in the imperfect, within three syntactic contexts exemplified here using the dynamic verb *wasal* 'to arrive' and two stative verbs: *jaf* 'to know' from the first category (perfectless verbs), and a verb from another category (pseudo-verbs, semi-perfectless verbs, verbs with perfect forms). Stative verbs require this marker when placed in three constructions:

- (i) The future tense construction formed with future marker *se*, *ser*, *ħa*, or *ghad* and a verb in the imperfect

- (57) *Naħseb se jasal tard.* dynamic  
'I think he'll arrive late.'
- (58a) \**Kif se naf x'hin jibda l-film?*  
'How will I know what time the movie starts?' stative
- (58b) *Kif se nkun naf x'hin jibda l-film?*  
'How will I know what time the movie starts?'
- (59a) \**Jekk jarak bil-ġelat, se jrid wiehed.*  
'If he sees you with the ice cream, he'll want one.' stative
- (59b) *Jekk jarak bil-ġelat, se jkun irid wiehed.*  
'If he sees you with the ice cream, he'll want one.'
- (ii) Deontic modal constructions with the pseudo-verb *għandu* 'should' or 'must'
- (60) *Għandu jasal fil-ħin.* dynamic  
'He should arrive on time.'
- (61a) \*?*La jaħdem hawn, għandu jaf x'hin jibda l-film*  
'Since he works here, he should know what time the movie starts.' stative
- (61b) *La jaħdem hawn, għandu jkun jaf x'hin jibda l-film.*  
'Since he works here, he should know what time the movie starts.'
- (62a) \*?*Il-madum m'għandux jizloq.*  
'The tiles should not be slippery.' stative
- (62b) *Il-madum m'għandux ikun jizloq.*  
'The tiles should not be slippery.'
- (iii) Constructions with an adverbial expression of purpose, such as (*sa*)*biex* 'in order to'
- (63) *Biex jasal fil-ħin, irid iqum kmieni.* dynamic  
'To arrive on time, he has to wake up early.'
- (64a) \*?*Biex taf x'hin jibda, aħjar issaqsi.*  
'To know what time it starts, you'd better ask.' stative
- (64b) *Biex tkun taf x'hin jibda, aħjar issaqsi.*  
'To know what time it starts, you'd better ask.'



illustrated by the oddness of such a sentence as ‘\*?Julie is having blue eyes’ (Filip 2012: 730), since eye colour is generally a permanent feature.

Therefore, when dynamism or change is implied, stative verbs take on certain characteristics of dynamic verbs, allowing them to fit within the progressive construction. Consider, for instance, (66a) compared to (66b). The latter, being about a child, implies a sense of dynamism, reflecting the way children’s features often change as they grow. The progressive is also used with stative verbs when telicity is implied, turning the states into temporary ones. Compare, for instance, (67a) with (67b). The progressive in the latter sentence implies a temporary situation, such as, *Kuġinti qed toqgħod man-nannu sa ma tbajjad id-dar* ‘My cousin is staying with grandpa until she paints her house’ or *sa ma ssib akkomodazzjoni oħra* ‘until she finds other accommodation.’

(66a) *It-tifel jixbah lil zitu.*

‘The boy resembles his aunt.’

(66b) *It-tifel lil zijuh qed jixbah.*

‘The boy is resembling his uncle.’

(67a) *Kuġinti toqgħod Birkirkara.*

‘My cousin lives in Birkirkara.’

(67b) *Kuġinti qed toqgħod man-nannu.*

‘My cousin is staying with Grandpa.’

## 7 Conclusion

In this paper, which presents an investigation of stative verbs in Maltese, I sought to achieve two main goals. First, by classifying these verbs into four distinct categories – perfectless verbs (*jaf* ‘to know’, *jixbah* ‘to resemble’), pseudo-verbs (*filh* ‘to contain’, *haqqu* ‘to deserve’), semi-perfectless verbs (*jaħraq* ‘to be hot’, *jaqta* ‘to be sharp’), and verbs with a perfect form (*xtaq* ‘to wish’, *ħabb* ‘to love’) – I show the diverse ways in which Maltese encodes stativity. Second, the study demonstrates that despite their formal differences, these verbs share certain morphosyntactic traits, namely they often rely on auxiliary verbs to express tense and modality and are generally incompatible with progressive constructions. A key finding is the pivotal role of auxiliary verbs like *kien* ‘to be’ in expressing tense and modality in stative verbs. These features highlight their fundamental distinction from dynamic verbs and the specific mechanisms Maltese uses to express stativity.

The paper further demonstrates that stative verbs generally resist contexts requiring dynamism, such as the progressive aspect or specific modal and purpose constructions, reinforcing their association with persistence and lack of change. The analysis also highlights the interplay between lexical and grammatical aspect, particularly in how stative verbs can adopt dynamic interpretations in specific contexts, such as temporary or contingent states.

In sum, this study provides an analysis of aspectology and verb classification in Maltese, shedding light on the characteristics of stative verbs. It also opens avenues for further research, including cross-linguistic comparisons and studies into the diachronic development of stativity in Maltese. Future research is necessary to (a) conduct a corpus-based analysis of stative verbs to provide a more data-driven understanding of their distribution, and (b) extend the scope of the study to include dynamic verbs with an examination of their morphosyntactic behaviour. Such work would facilitate a comparison between stative and dynamic verbs, offering deeper insights into their interaction and role within the Maltese verbal system.

## Abbreviations

DEF	definite
F	feminine
IPFV	imperfect(ive)
M	masculine
OBJ	object marker
NEG	negation
PFV	perfect(ive)
PROG	progressive
PL	plural
SG	singular

## References

- Aquilina, Joseph. 1965. *Teach Yourself Maltese*. London: English Universities Press.
- Aquilina, Joseph. 1973. *The Structure of Maltese*. Malta: Royal University of Malta.
- Aquilina, Joseph. 1979. *Maltese-Arabic Comparative Grammar*. Malta: University of Malta.
- Aquilina, Joseph. 1987. *Maltese-English Dictionary*, vol.1. Malta: MidSea Books.
- Bhat, Darbhe Narayana S. 1999. *The Prominence of Tense, Aspect and Mood*. Amsterdam/Philadelphia: John Benjamins.
- Binnick, Robert I. 1991. *Time and the Verb: A Guide to Tense and Aspect*. University of Chicago Press.
- Binnick, Robert I. (ed.). 2012. *The Oxford Handbook of Tense and Aspect*. MIT Press.

- Binnick, Robert I. 2020. Aspect and Aspectuality. In Bas Aarts & April McMahon (eds.), *The Handbook of English Linguistics*, 183–207. 2nd edn. Hoboken, NJ: Wiley-Blackwell.
- Borg, Albert. 1981. *A Study of Aspect in Maltese*. Ann Arbor: Karoma.
- Borg, Albert. 1988. *Ilsienna: Studju Grammatikali*. Malta: Self published by the author.
- Borg, Albert & Marie Azzopardi-Alexander. 1997. *Maltese: Descriptive Grammars*. London/New York: Routledge.
- Camilleri, Maris. 2016. *Temporal and Aspectual Auxiliaries in Maltese*. University of Surrey, PhD dissertation.
- Carlson, Greg N. 1977. A Unified Analysis of the English Bare Plural. *Linguistics and Philosophy* 1(3). 413–456.
- Chierchia, Gennaro. 1995. Individual-Level Predicates as Inherent Generics. In Gregory N. Carlson & Francis Jeffrey Pelletier (eds.), *The Generic Book*, 176–223. Chicago: University of Chicago Press.
- Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.
- Comrie, Bernard. 1985. *Tense*. Cambridge: Cambridge University Press.
- Comrie, Bernard. 1991. On the importance of Arabic to general linguistic theory. In Bernard Comrie & Mushira Em (eds.), *Perspectives on Arabic Linguistics III: Papers from the Third Annual Symposium on Arabic Linguistics*, 3–30. Amsterdam: John Benjamins.
- Cremona, Anthony. 1961. The classification of the Maltese verb. *Journal of Maltese Studies* 1. 117–129.
- Cremona, Anthony. 1964. *Tagħlim fuq il-Kitba Maltija II*. 4th edn. Malta.
- Cremona, Anthony. 1966. The phraseological use of the verbs 'Qies', 'Għadd', and 'Dann'. *Journal of Maltese Studies* 3. 75–78.
- Cuvalay-Haak, Martine. 1997. *The Verb in Literary and Colloquial Arabic*. Berlin: Mouton de Gruyter.
- de Swart, Henriëtte. 1998. Aspect shift and coercion. *Natural Language and Linguistic Theory* 16(2). 347–385.
- Dowty, David R. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Reidel.
- Ebert, Karen. 2000. Aspect in Maltese. In Östen Dahl (ed.), *Tense and Aspect in the Languages of Europe*, 753–785. Berlin: Mouton de Gruyter.
- Fabri, Ray. 1995. The tense and aspect system of Maltese. In Rolf Thieroff (ed.), *Tense Systems in European Languages*, 327–343. Tübingen: Niemeyer.
- Filip, Hana. 2012. Lexical Aspect. In Robert I. Binnick (ed.), *The Oxford Handbook of Tense and Aspect*, 721–751. Oxford: Oxford University Press.
- Grech, J.J. 1977. *A study of aspect in English and Maltese*. University of Malta B.A. thesis.
- Ingham, Bruce. 1983. Review of *A Study of Aspect in Maltese*. *Bulletin of the School of Oriental and African Studies, University of London* 46(2). 353–354.
- Korpus Malti, Maltese Language Resource Server. 2024. *MLRS Corpus* (Version 4.2). University of Malta. Retrieved from <https://mlrs.research.um.edu.mt>.
- Kortmann, Bernd. 1991. The triad 'Tense-Aspect-Aktionsart': Problems and possible solutions. In Carl Vetters & Willy Vandeweghe (eds.), *Perspectives on Aspect and Aktionsart*, 9–29. (Belgian Journal of Linguistics 6). Bruxelles: Editions de l'Université de Bruxelles.
- Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Młynarczyk, Anna. 2004. *Aspectual Pairing in Polish*. Utrecht: LOT.
- Olsen, Mari Broman. 1997. *A Semantic and Pragmatic Model of Lexical and Grammatical Aspect*. New York: Garland.
- Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge, MA: MIT Press.
- Saydon, Pietru Pawl. 1935. Il-Kelmiet "sa", "ha", "qed". *Il-Malti*, June, 44–45.
- Schinas, M. 1977. *Futurity in Maltese and English*. University College of North Wales M.A. thesis.
- Smith, Carlota. 1991. *The Parameter of Aspect*. Dordrecht: Kluwer Academic Publishers.

- Spagnol, Michael. 2007. *L-Aspett Lessikali fil-Verb Malti*. University of Malta M.A. thesis.
- Spagnol, Michael. 2009. Lexical and grammatical aspect in Maltese. *Ilsienna: Journal of Maltese Linguistics* 1. 51–86.
- Spagnol, Michael. 2023. Verbi Stativi fil-Malti. *Lehen il-Malti*. 185–196.
- Stolz, Thomas & Andreas Ammann. 2007. Beda u qabad: The Maltese inchoative/ingressive. *Romano-Arabica* 6/7. 149–158.
- Stolz, Thomas & Andreas Ammann. 2008. The Maltese continuative: A grammaticalization borderliner. In Elisabeth Verhoeven, Stavros Skopeteas, Yong-Min Shin, Yoko Nishina & Johannes Helmbrecht (eds.), *Studies on Grammaticalization*, 169–184. Berlin: Mouton de Gruyter.
- Sutcliffe, Edmund F. 1936. *A Grammar of the Maltese Language with Chrestomathy and Vocabulary*. London: Oxford University Press.
- Thieroff, Rolf. 1995. More on inherent verb categories in European languages. In Rolf Thieroff (ed.), *Tense Systems in European Languages*, 1–36. Tübingen: Niemeyer.
- Vanhove, Martine. 1987. Remarques sur quelques auxiliaires et particules verbales en maltais. *Matériaux Arabes et Sudarabiques* 1. 35–51.
- Vanhove, Martine. 1993. *La langue maltaise: Etudes syntaxiques d'un dialecte arabe 'périphérique'*. Wiesbaden: Harrassowitz.
- Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* 66(2). 143–160.
- Zammit Mangion, Dorothy. 1977. *Some Aspects of the Maltese Verb in Relation to the Teaching of English*. University of Malta M.A. thesis.

# Tabula gratulatoria

Greville Corbett  
Pierluigi Cuzzolin  
Daria Dermaku (Appelganz)  
Ray Fabri  
Francesco Gardani  
Bernd Heine  
Johannes Helmbrecht<sup>1</sup>  
Nikolaus Himmelmann  
Nicole Hober  
Sonja Kettler  
Paolo Miccoli  
Edith Moravcsik  
Nicole Nau  
Immacolata Pinto  
Ignazio Putzu  
Tamar Reseck  
Maja Robbers  
Julian Rott  
Benjamin Saade  
Jeannette Sakel  
Susanne Schuster  
Cornelia Stroh  
Anna Maria Thornton  
Aina Urdze  
Ingo H. Warnke  
Marina Wienberg

---

<sup>1</sup> Johannes Helmbrecht intended to contribute to this volume, however, to our great regret he passed away in May 2024.



# Language index

- !Ora 118, 129  
Achumawi 2, 5–6, 15–16  
Ainu 234, 275, 278  
Algic 15  
Arabic 200, 204, 225, 234, 313, 321  
Armenian 182, 184, 191  
Atsugewi 2, 5–6, 15–16  
Austrian Sign Language 175  
Ayutla Mixe 188  
Azeri 55
- Bisaya 23  
Bora 188  
Bulgarian 200, 210–211, 214, 234, 251–252, 260
- Catalan Sign Language 175  
Cayuga 200, 204, 206–207, 218, 228, 233–235  
Celtic 160, 182, 275  
Chabacano 21–38  
Chico 8  
Chimariko 2, 5–6, 15–16  
Chinese 93, 172–174, 184, 225  
Chol 188  
Chukchansi 14–16  
Chukchi 300  
Chukotko-Kamchatkan 275, 300  
Chumash 15  
Cochimí 15  
Costanoan 12–13, 15  
Crow 188  
Czech 301
- Dutch 50, 270, 327
- Emmi 266  
English 21–23, 31–32, 48, 50, 53, 56–57, 93–95,  
97–98, 100, 102–114, 150, 153, 156, 170,  
200, 212, 218, 220–223, 225, 231–232, 234,  
245, 248, 257, 259–260, 265, 268, 281–282,  
287–288, 293, 296–297, 307–310, 315,  
317–318, 320–321  
– Old 250–251, 259–260  
– Middle 259–260
- Esselen 15  
Estonian Sign Language 175
- Filipino 23  
Finnish 202–203, 206, 210–211, 220, 223,  
231–234, 236  
French 42, 64, 70, 77, 119, 200, 203, 209–210,  
215–217, 220, 231, 234–236, 248, 256, 260,  
265, 268–270, 309, 317–318  
Frisian 42
- G|ui 118, 122, 124–126, 128  
Garifuna 189  
Georgian 200, 204, 210–211, 214, 218–219,  
228, 233  
German Sign Language (DGS) 157, 165, 174,  
176–178  
German 41–43, 48, 50, 53–54, 56–57, 63–65, 70,  
73, 88, 95, 107, 225, 234, 248, 251, 256–257,  
260, 268, 270–271, 309, 312  
– Old High (OHG) 64–65, 67–68, 77  
Germanic 63, 65, 71, 73–74, 88, 153, 235, 244,  
251, 275  
Gothic 251, 260  
Greek 48, 200, 211, 220, 226, 229, 234–235, 254  
– Modern 253–254, 260  
– Ancient 244–245, 249–250, 253–255, 257, 260  
– Homeric 248–250  
– Classical 248, 253–255, 257–258  
Greenlandic 165, 167, 234
- Hebrew 313  
Hokan 15–16  
Hungarian 156, 158–159, 165, 180, 200, 212–213,  
224, 231, 234–235
- Iatmul 189  
Ilokano 23  
Imonda 189  
Indo-European 55–57, 153, 155, 159–160, 182,  
191, 275, 300, 309  
Indonesian 266  
Indo-Portuguese 25

- Irish 270  
 Italian 63–64, 70, 75, 77, 88, 215–217, 236, 251,  
 256, 260, 265–266, 268–271, 309, 317, 327  
 Itelmen 300  
  
 Jalkunan 189  
 Japanese 151, 180, 190, 200, 204, 219, 221,  
 229–230, 233, 235, 280–282, 287, 309  
 Japonic 151, 157, 190, 275  
  
 Kalaallisut 165–168, 184  
 Kalamang 190  
 Karaim 55  
 Kartvelian 200–201, 210, 228, 234, 236  
 Karuk 2, 6–7, 15–16  
 Ket 200, 222–223, 233–236  
 Khalaj 55  
 Khanty 300–301  
 Khoe 117–142  
 Khoekhoe 118, 126–129, 140  
 Khwe 118–123, 125–126, 128  
 Klamath-Modoc 2, 15  
 Kombio 190  
 Konkow 8–10, 15–16  
 Korean 44  
 Koreanic 275  
 Koromfé 191  
 Koryak 300  
 Kuot 94–95  
 Kurdish/Kurmancî 48  
 Kuvale 117, 134  
 Kwadi 117–141  
 Kx'a 140–141  
  
 Latin 107, 244, 252, 254–256, 260, 310–311  
 Latvian 267  
 Laz 200, 204, 211, 217–218, 221–222, 228–229,  
 232–233  
 Livvi 300–301  
  
 Maidu 2, 8, 10, 15–16  
 Makasar 266  
 Malayo-Portuguese 25  
 Maltese/Malti 168–170, 200–201, 219, 233, 266,  
 314, 325–346  
  
 Mehek 191  
 Melanesian Pidgin 95, 111  
 Miwok 12–16  
 Mongolic 275  
 Mundari 320  
  
 Na-Dene 15  
 Nalik 93–115  
 Nanaic 278  
 Naro 118, 122, 127–130  
 Ngan'gityemerri 200–201, 204–206, 215,  
 217–218, 225, 227–228, 233–234  
 Nigerian Sign Language 175  
 Nisenan 8–13, 15–17  
 Nivkh 199–200, 204, 207, 215, 217, 220, 223,  
 228–230, 233, 236, 275, 278  
 Nomlaki 10  
  
 Old Church Slavonic 252, 260  
  
 Palaihnihan 5, 15  
 Patwin 10–12, 15–16  
 Penutian 15–16  
 Persian 55, 250, 257  
 Piedmontese 64, 70, 75, 77  
 Polish 276–277, 280, 292, 301–302  
 Pomoan 2, 15  
 Portuguese 23, 38, 102, 265, 270  
  
 Rapanui 164–165  
 Reta 191  
 Romance 43, 63–64, 77, 200, 220, 244, 256, 260,  
 265, 275, 317  
 Romani 42  
 Romanian 200, 210  
 Russian 44–45, 48, 200, 203, 207–209, 211, 215,  
 220, 222, 228, 232–233, 236, 252, 260, 270,  
 276, 280, 281–282, 287, 309, 317  
  
 Salinan 15  
 Salishan 320  
 Sandawe 192  
 Sanskrit 309  
 Semitic 169–170, 200, 219, 307, 309, 313  
 Seri 15

- Shasta 2, 4–7, 15–16  
 Shipibo-Konibo 192  
 Shoshoni 182, 192  
 Shua 118, 126, 129  
 Slavic 55, 200–201, 225, 229, 234–236, 244, 251, 275, 300–302, 311  
 Sorbian 42, 312  
 Spanish 3, 17, 21–23, 27–29, 31–34, 36–37, 43–44, 200, 209–211, 214–215, 220, 230, 234–236, 256, 260, 265, 270  
 Swedish 56  
  
 Tabasaran 193  
 Tagalog 21–24, 27–31, 33–34, 36–37, 200–201, 204–206, 210–211, 219–220, 224, 226–228, 233–235, 320  
 Takelma 15  
 Thai Sign Language 175  
 Tjwao 126, 129  
 Tok Pisin 93–115  
 Transeurasian 275–276  
 Ts'ixa 120–123, 125–126, 129  
 Tungag 94  
 Tungusic 161–163, 275, 278  
 Turkic 53, 55–57, 200, 275  
 Turkish 41–57, 200, 233, 266  
  
 Tuu 140–141  
  
 Udihe 161–163, 165  
 Uilta 275–302  
 Ukrainian 301–302  
 Ulcha 278  
 Uralic 158–159, 275, 300–301  
 Utian 12  
 Uto-Aztecan 15, 182, 192  
 Uzbek 55  
  
 Walser German 63–88  
 Wandala 267  
 Wari' 170–172  
 Washo 2, 7–8, 10, 15–16  
 Welsh 155, 159–161, 182, 184  
 Wintu 10–11  
  
 Yana 2, 6, 10, 15–16  
 Yeniseian 200, 275  
 Yokuts 14–16  
 Yukaghir 147–148, 157, 190, 275  
 Yukian (Yuki-Wappo) 2, 15  
 Yuman 15  
 Yurok 6–7, 15–16  
 Yuwaalaraay 155, 193



# Subject index

- ablative 148, 155, 158, 161–162, 193, 247, 251  
accusative 27, 36, 202, 212, 228, 233, 243, 253,  
285, 296, 312  
acquisition 44, 46, 50, 94, 96  
active 121, 341  
adjective(s) 28–29, 34, 64, 73–74, 77, 101, 104,  
128, 160, 275–302, 307, 310, 315–316, 321  
– proprietive 277, 284, 198  
– attributive 276–280, 283–287, 292, 295,  
300–302  
– predicative 128, 297  
affix(es) 13, 28–32, 34, 36, 136, 149, 151, 158,  
161–162, 165, 201, 205, 223, 227, 307,  
309–314, 316–317, 321, 336  
agreement 63–64, 73–77, 82, 85–88, 285, 336  
allative 191, 247, 259  
allochthonous 42–45  
allomorphy/allomorphies/allomorphic 66, 69,  
72, 119–120, 125–126, 128, 131, 138, 159  
aorist 14, 214  
aspect 5, 7, 13–15, 25, 27, 49, 102, 199–236,  
325–327, 341, 344, 346  
aspectual(ity) 51–53, 199–204, 207–208, 212–214,  
216, 220–221, 225–231, 233–236, 327  
aspectoid 199–202, 204–205, 207, 212–219,  
222–224, 229, 231, 233–235  
atelic 226–227, 329–330  
autochthonous 42–43, 45  
  
benefactive 25, 120–121, 123, 125, 127  
bilingual(s)/bilingualism 16, 21, 24, 34, 36, 41–42,  
46–50, 52–57, 93, 98, 100, 110, 114  
  
canon/canonical 13, 32, 34, 51–54, 125, 149–153,  
159–160, 162–163, 165, 170, 178, 181,  
245–246, 248  
causative/causativization 21–37, 123, 127, 207,  
228, 309  
clause combining 41, 49–50, 54–55  
configuration(s) 125, 132, 147–184, 189, 193  
connector(s) 54–56  
constituent order 275, 279, 283  
  
converb(s) 49–54, 56, 229  
convergence 26, 36, 50, 55, 69, 93, 98, 100–101,  
103, 140  
Creole(s) 21–25, 93  
  
dative 122–123, 127, 253, 258  
differential object marking (DOM) 201–203, 206,  
208, 210, 221  
differential voice marking (DVM) 199–200, 206,  
210, 221, 235  
directionality/directionality 147–152, 154–155,  
157–159, 163–166, 170–172, 175–181, 189, 193  
durative 102, 327–328, 330–331  
  
evidentiality 4, 6, 27  
  
general location(s) (GL) 147, 151–155, 158–166,  
168–174, 178–183, 188–193  
genitive 128, 208, 228, 253, 257–258,  
267, 312, 315  
Goal 147–155, 158–174, 176–183, 188–193, 199,  
201–202, 205–206, 210, 212, 216, 218–219,  
221–236, 327, 327, 330  
gram(s) 129–130, 138, 140, 268  
grammaticalisation/grammaticalization 16,  
24–25, 34, 36, 53, 88, 121, 128–130, 175, 177,  
201, 219, 228, 244  
  
heritage language(s) 41–57  
Hispanization 21–24, 36  
hypotactic/hypotaxis 49–50, 54  
  
idiomatic expressions/idiom(s) 111–112, 265,  
269–271  
imperfective 26, 123, 135, 200–202, 204–210,  
214–216, 226–228, 311, 326–327  
inflection/inflectional 9, 11, 36, 63, 70, 78–79, 80,  
85–86, 88, 123, 200, 204–205, 209–210, 214,  
217, 251, 311–313, 317, 327  
– strong 78  
– weak 64, 73, 78, 82  
instrumental 25, 228, 244, 251, 255, 297, 309, 318

- intransitive(s) 22, 26, 28–29, 31–32, 36, 102, 110, 132, 136–137, 139, 141, 207, 212, 217–218, 223, 315, 338
- iteration 218, 265–268, 270
- junction 117–141
- language contact 1–2, 17, 22–24, 37, 41–42, 46–48, 53, 55, 57, 64, 88, 93–94, 117, 140–141
- lexicalisation/lexicalization 25, 199–201, 207, 216–222, 224–226, 233, 235–236
- lexicon 21, 44, 93, 102, 106–107, 114, 117–118, 270–271, 287, 309, 319
- lexifier 21–22, 26, 32–33, 37
- lingua franca 93, 97
- linguistic minority 64
- locative 17, 27, 102, 123, 127, 158, 161–162, 164, 193, 247, 257, 259
- metaphor(s) 243–260
- minority language(s) 42–45
- monolingual(s)/monolingualism 45–46, 48–50, 52, 54–55, 93–94
- morph(s) 179, 307–310, 312–314, 321–322
- morpheme(s) 7, 55, 117, 119, 126–128, 131, 137, 140, 148, 151–152, 156, 158, 162–163, 165, 167–168, 175–181, 183, 189, 229, 234, 269–270, 281–282, 293, 308, 310–311, 313, 316
- morphological complexity 147–148, 153, 179, 183–184
- morphology 15, 64, 101, 114, 117–118, 201, 266, 270, 286, 296, 310, 313, 319, 325, 327
- Multi Word Expression(s) (MWE) 265, 269–271
- multilingual(s)/multilingualism 1, 3, 21, 45–47, 94, 97–98, 113–114
- multi-verb constructions (MVC) 117, 120–125, 129, 137–140
- nominalization(s) 12, 49, 258, 285–286
- nominative 243, 312
- noun phrase (NP) 101, 173, 207, 270, 275, 284–285, 292, 297–298
- noun(s) 17, 31–33, 37, 101, 104, 109, 128, 150–151, 153–155, 157, 160–161, 163–168, 171–174, 189, 193, 203, 207–209, 212–213, 212, 245, 250, 256, 258, 260, 270, 275–277, 279, 282–301, 307–312, 315–316, 318–321, 328, 336
- numeral(s) 101, 109, 281–282, 284, 296
- paratactic 49–50, 54–56
- passive 25, 76–78, 81–82, 85, 104–105, 121, 130, 243–246, 249–253, 255–260
- path 122, 140, 176–177, 205, 216, 218–219, 222–223, 225–226, 229–230, 232–236, 243–244, 247, 254–255, 257–258, 260
- perfective 5–6, 14, 128, 200–202, 204–211, 214–215, 219, 226–228, 326
- phonology 98–99, 114, 119, 176, 319–321
- Place 147–155, 158–173, 176, 178–180, 183, 188–193, 259,
- plural/plurality 101, 203, 207–208, 212, 266, 268, 270, 276, 284–285, 292–293
- possession/possessive 12, 101–102, 128, 161–163, 166, 171, 284, 286, 295, 331, 337
- alienable 102, 184
- inalienable 101–102, 284
- possessor(s) 51–53, 284–285
- possessum 51–53
- postposition(s) 49, 148, 158, 161, 163, 227, 229, 235, 294
- pragmatic(s) 54, 93, 112, 114, 265, 269, 271
- predicate(s) 24, 29–30, 36, 102, 104, 106, 121, 125, 129–130, 136–137, 140–141, 191, 204–206, 220, 227, 229, 233, 279, 326, 328, 330–331, 344
- complex 24, 28, 37, 119, 121, 125–126, 136, 139–140
- preposition(s)/prepositional 103–105, 151, 156, 159–161, 164–165, 168–169, 172–174, 226–228, 230–231, 243, 245, 248–253, 256–260, 336–337
- preverb(s)/preverbal 7, 27, 32, 102, 106, 136, 199, 204, 207, 210, 213–214, 216–219, 224–226, 228–229, 231, 234–236
- progressive 123, 134, 200, 202, 212, 214–216, 219, 221, 224, 227, 325–327, 329, 232, 341, 344–346
- pseudo-verb(s) 325, 327, 333, 335–337, 342–345
- reconstruction(s) 10, 117–118, 123, 139, 141
- reduplication 119, 132–137, 139, 141, 201, 205, 226–227, 265–269

- root(s) 17, 65, 67–69, 71–73, 86–88, 95, 102, 117, 119–122, 124–126, 129, 131–134, 136–139, 141, 217, 219, 222, 307–322
- semantic(s) 11, 14, 16, 24–26, 32–34, 37, 47, 54, 57, 93, 98, 110–111, 121, 128, 139–141, 155, 164, 175, 217, 226, 243–245, 266–268, 270–271, 282, 288, 291–292, 294–295, 298–300, 307, 312, 315, 318–322, 334
- semantic role(s) 217, 243–247, 254, 257, 259–260, 284
- agent(s) 24, 28, 35–37, 243–260,
  - patient(s) 36, 210, 214, 244–246, 254–255
  - instrument(s) 205, 244, 246–247, 255, 259, 317–318
- serial verb (constructions) 21–22, 24, 26–27, 33–34, 140, 225
- sociolinguistic(s) 22, 24, 36, 41–42, 46–47, 49, 52–54, 57, 97–98
- Source 147–154, 158–173, 176–180, 183, 188–193, 205, 221, 226, 228–229, 243–244, 247–248, 251–253, 256, 258, 260
- spatial relation(s) 147–184, 244, 253, 260
- stacking 199, 201, 214–217, 230–231, 235–236
- stem(s) 11–12, 15, 66–69, 85, 128, 132, 134–137, 139, 141, 200, 226–228, 311, 315–316, 334
- subordinate/subordination/subordinator 12, 25, 49–52, 55–57, 125
- syntactic/syntax 22, 25–26, 36, 53–54, 85, 93, 103–104, 114, 125, 129, 132, 135–137, 139–140, 153, 157, 164, 172, 174, 233, 244, 268, 270–271, 283–284, 316, 319, 325, 341–342
- tautology/tautologies 9, 265, 267–271
- telic/telicity 218, 226–227, 327, 329–331, 345
- tense(s) 1, 3–17, 85, 102, 268, 311, 325–327, 341–342, 344–345
- past 4–15, 17, 65, 77, 102, 119, 313
- transitive/transitivity 22, 26, 28–29, 31–32, 34–37, 75, 102, 104, 110, 207, 210, 212, 217, 223, 229, 244–245, 249–250, 253, 258, 260, 315, 338
- valency 22, 24–26, 36, 199, 210, 212, 214, 217, 221, 233, 235, 243
- verb(s)
- stative 219, 325–346
  - perfectless 325, 333, 337, 342, 344–345
  - light 21, 32–33, 37, 201, 227, 234
  - motion 155, 174, 218–219, 221–224, 226, 229–232, 234, 327

