Maja Robbers

Orientation and motion in the world’s languages

From field studies to cross-linguistic comparison
Dissertation presented at Uppsala University to be publicly examined in Humanistiska Teater, Thunbergsvägen 3C, Uppsala, Saturday, 14 October 2023 at 10:00 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Anetta Kopecka (Université Lumière Lyon 2).

Abstract

Human life frequently involves spatial orientation and motion, and natural languages express manifold aspects of spatial perception in diverse ways. The articles included in this thesis delve into several of these aspects and explore space and motion from multiple perspectives, ranging from a dedicated field study of within-language variation systems to cross-linguistic comparisons of various orientation and motion aspects. A field study of Yine, an Arawakan language spoken in the Peruvian Amazonas, points to relative reference frame use by younger speakers, as opposed to riverine orientation in the speech of senior community members. The central questions addressed in two comparative studies revolve around the distinctions among languages in terms of which form classes are employed to express Place/Goal/Source coding and related motion aspects with different types of Ground arguments. A study of basic location and motion expressions in 35 languages demonstrates that motion verbs are more often pivotal than grammaticized markers for Goal and Source, and that coding asymmetries arise from the use of both also within languages. A dedicated paper on the novel computer-assisted sampling technique employed to obtain the sample of 35 languages discusses how informativity loads in grammatical descriptions can be pre-screened to facilitate data assessment. It is shown how this method can be employed to create genealogically balanced samples which give access to the variety of coding strategies present in the world’s languages. Another part of the cumulative dissertation addresses the question of how these form classes interact with various contextual factors and describes previously underresearched Source expressions based on iconic ordering of elements in detail. The findings of the thesis address the nuanced nature of orientation and motion expressions across languages. The exploration of these topics is underpinned by an array of typologically and areally diverse languages, and descriptive gaps in lesser-explored languages are highlighted throughout the thesis. The research underscores the importance of scrutinizing languages that have previously received only limited attention due to a lack of descriptions, and it offers insights into underresearched languages that are hoped to contribute to the development of a general typology of motion and orientation.

Keywords: spatial relations, motion events, diversity linguistics, language typology, frames of reference

Maja Robbers, Department of Linguistics and Philology, Box 635, Uppsala University, SE-75126 Uppsala, Sweden.

© Maja Robbers 2023

ISSN 1652-1366
URN urn:nbn:se:uu:diva-509567 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-509567)
Für Lothar
List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


(III) Robbers, Maja. Inferred Source of motion and bipartite Source expressions. *Submitted*.

(IV) Robbers, Maja and Harald Hammarström. Bibliographic bias and information-density sampling. *Submitted*.

(V) Robbers, Maja. A variety-driven approach to Goal-Source (a)symmetries. *Submitted*.

Reprints were made with permission from the respective publishers.
Acknowledgements

The journey of writing this thesis has been influenced by diverse projects that have both enriched and challenged my path, especially in the tumultuous global landscape created by the COVID-19 pandemic during the peak of my studies. Navigating through these intricate patterns of academic pursuits and unforeseen global shifts would not have been possible without the support and contributions of many, and I am fortunate to have been part of a department where colleagues and students foster a stimulating atmosphere enriched by their diverse interests and boundless enthusiasm for linguistics. I am deeply grateful for the support provided by my supervisor Harald Hammarström who empowered me to shape my own scholarly path throughout the entire journey, and who provided profound answers and effective help whenever needed. My sincere thanks also go to my co-supervisor Martin Haspelmath, who joined my journey much later but who has been an exceptional source of extremely clear insights. I am also beyond thankful for the guidance provided by Thomas and Christel Stolz during my studies in Bremen, and by José Antonio Flores Farfán and José Santiago Francisco in Mexico City, all of whom sparked and fed my interest for diversity linguistics in the first place. I also wish to thank Henrik Bergqvist and Jürgen Bohnemeyer for their meaningful input, and Michael Dunn for all the kind help throughout the years. Needless to say, I am extremely grateful to the Marcus and Amalia Wallenberg Foundation for making the ‘From Dust to Dawn: Multilingual Grammar Extraction from Grammars’ project possible, and to Olof Gjerdmans stipendiefond for the additional support. Of course, I also want to express my gratitude to all the staff at the Department of Linguistics and Philology in Uppsala, and especially to Artur Kilmizev, Marc Allassonnier-Tang, Philipp Rönchen, Erik Elgh, Freja Lindgren, Oscar Billing, Emil Mannby, Deepthi Gopal, Tiago Tresoldi, Mahmut Agbaht, and Buket Öztekin for their support and companionship. Anna Ingves of Scandinavian Languages deserves a special mention for her enthusiastic friendship and all the kind help she offered on countless occasions. My heartfelt appreciation also goes to my colleagues and friends Julia Nintemann and Nicole Hober in Bremen for their dedicated contributions and hard work throughout our collaborative project. Por su ayuda en Perú, gracias a Juan Rubén Ruiz Zevallos y poyagnu, Miller, Eva, Adrien, Dencina, Ely, Taki gawa Twi. Danke an Adriana und Susy für die Freundschaft. Tausend Dank an meine Familie, an meine Eltern, an meine Brüderchen und an Kyra.
Tack för allt, Jesper. Gracias a Carolina, Iván y Bruno por ser mi familia en Uppsala, y a Alejandro por siempre ir un poco más allá para estar conmigo. Vladlena, ты ушла слишком рано и я скучаю по тебе.
Contents

1 Introduction and aims of the thesis ............................................................ 13
   1.1 Previous and related studies ............................................................... 15
   1.2 From spatial interrogatives to general coding tendencies .............. 16

2 Overview and results of the compilation thesis ......................................... 24

3 Outlook ...................................................................................................... 29

References ..................................................................................................... 30
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>first person</td>
</tr>
<tr>
<td>3</td>
<td>third person</td>
</tr>
<tr>
<td>ACC</td>
<td>accusative</td>
</tr>
<tr>
<td>APPL</td>
<td>applicative</td>
</tr>
<tr>
<td>BG</td>
<td>background marker</td>
</tr>
<tr>
<td>DAT</td>
<td>dative</td>
</tr>
<tr>
<td>DEF</td>
<td>definite</td>
</tr>
<tr>
<td>F</td>
<td>feminine</td>
</tr>
<tr>
<td>FREQ</td>
<td>frequentative</td>
</tr>
<tr>
<td>IMP</td>
<td>imperative</td>
</tr>
<tr>
<td>INF</td>
<td>infinitive</td>
</tr>
<tr>
<td>IPFV</td>
<td>imperfective</td>
</tr>
<tr>
<td>LOC</td>
<td>locative</td>
</tr>
<tr>
<td>M</td>
<td>masculine</td>
</tr>
<tr>
<td>N</td>
<td>neuter</td>
</tr>
<tr>
<td>NF</td>
<td>non-finite</td>
</tr>
<tr>
<td>NOM</td>
<td>nominative</td>
</tr>
<tr>
<td>NONDUR</td>
<td>non-durative</td>
</tr>
<tr>
<td>NPF</td>
<td>noun-prefix</td>
</tr>
<tr>
<td>OBL</td>
<td>oblique</td>
</tr>
<tr>
<td>PART</td>
<td>partitive</td>
</tr>
<tr>
<td>PAST</td>
<td>past-tense</td>
</tr>
<tr>
<td>PFV</td>
<td>perfective</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>PRES</td>
<td>present tense</td>
</tr>
<tr>
<td>QUOT</td>
<td>quotative</td>
</tr>
<tr>
<td>SBJ</td>
<td>subject</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>TOPO</td>
<td>toponym/placename</td>
</tr>
<tr>
<td>VCL</td>
<td>verb-stem closure</td>
</tr>
</tbody>
</table>
1 Introduction and aims of the thesis

Moving and orienting oneself in space are essential functions in human thought, action and speech. Motion and spatial perception, much like overlaps in the coding of spatial and temporal relations (Haspelmath 1997), have garnered attention in the linguistic literature. Scrutinizing a variety of languages, it has been shown that meaning components of location and motion events are coded by and distributed across different parts of speech, depending on languages’ individual tendencies and restrictions (Bohnemeyer et al. 2007). Furthermore, what can be outlined as the ‘spatial system’ of a natural language may consist of more closely or loosely connected individual subsystems. In the Mesoamerican Sprachbund, for example, we often find spatial relations coded by body part lexemes and by grammatical forms derived from body part roots (inter alia, Hollenbach 1995; Levy 1999; MacLaury 1989). The productive use of body part roots for spatial functions, such as topological relations and motion verb derivation, is paralleled by a preference for geocentric and especially intrinsic spatial reference frames (Levinson 2003), as opposed to relative, ego-centered frames (employing, for instance, terms relating to left and right), which are familiar from Standard Average European (SAE) languages.\(^1\) Relative frame types found in Mesoamerican languages have been hypothesized to surface due to language contact (Bohnemeyer et al. 2014; O’Meara and Pérez Báez 2011). Across languages in general, egocentric orientation in space tends to arise in varieties spoken in urban settings, whereas a preference for geocentric frames can be observed in varieties spoken in environments with salient landmarks (Palmer et al. 2017; Mishra et al. 2003; Pederson 2006).

Expression types giving evidence of certain reference frame use can be coded by grammatical as well as lexical items, which may belong to open or closed classes and to smaller or larger inventories. Quantitative studies have attempted to link larger inventories of grammaticized directional or otherwise deictic items with cultural ‘complexity’ (Perkins 1992), or morphological ‘complexity’ in general with smaller group size (Lupyan and Dale 2010). Researchers of spatial language are thus called to discuss what ‘complexity’ of a

\(^1\) Lum et al. (2022) discuss the history of the research on spatial frames of reference in detail. As to non-egocentric orientation, Bohnemeyer et al. (2015) provide the most detailed account of subtypes of the geocentric frame type, that is, ABSOLUTE (e.g., north), GEO MORPHIC (e.g., downriver) and LANDMARK-BASED (e.g., mountainward) (see also Lum et al. 2022: 5).
spatial system may mean when this bias towards languages that employ morphological coding is removed, but coding strategies, repertoire sizes of motion verbs and other lexical classes, deixis, pragmatics and variation are taken into account. More in-depth research on individual languages and smaller-scale comparisons are still required to work towards a definition of ‘complexity’, or challenge the very notion of it on the basis of the remarkable variety found in the world’s languages. The present thesis is a compilation of studies that aims to address these and related issues by filling descriptive gaps in underresearched languages (II and III) and compare specific spatial functions across typologically diverse languages (I and V), partly by employing a novel sampling technique (IV and V). All of the five studies strongly emphasize the persistent need to document expressions of spatial relations and motion events in natural languages to uncover the variety of systems and links between putative subsystems.

The main aims of the compilation thesis are to explore topics surrounding spatial language from different angles and perspectives to point out potential oversights in previous research. Despite a rich research tradition, studies in spatial orientation and motion are far from coming close to generalizations, defining universals of spatial language or exhausting the description of the plethora of strategies and form classes that are relevant to the coding of location and motion. The overarching battery of questions of this compilation of studies can be summarized as follows:

How do languages differ in terms of form classes, or parts of speech, that code spatial relations and motion? How do these forms interact with each other, and with other mechanisms of language, such as element order, inference and utterance context? Which functional categories are affected by preferences for geocentric orientation, and which contact phenomena can we observe?

These questions may not be new, but must be tackled continuously by taking into account the variety of languages to rule out any potential areal or genealogical biases. Moreover, and beyond this thesis, it is paramount to continue to study languages and areas that have not received much attention in the previous literature due to lack of or unavailability of written descriptions. Researchers of spatial language can now profit from the study of newer descriptions as well as from collaborative research. This has partly been tackled by the present thesis. Following the summary of relevant previous literature (Section 1.1), Section 1.2 introduces related research, and includes a short introduction to the variety of coding strategies for basic motion events in the world’s languages. Section 2 provides an overview of the topics and results of the individual studies in this compilation thesis. Section 3 provides an outlook and proposes some pathways for future studies.
1.1 Previous and related studies

Due to the great variety of lexical and grammatical strategies to express spatial relations, linguistic studies on space and motion tend to focus on individual languages, areas or families, as well as on individual aspects of spatial systems in detailed, small-scale and cross-linguistic comparisons. Goal-Source (a)symmetries have been a subject of considerable interest (see, for example, the studies in Kopecka and Vuillermet 2021; Wälchli and Zúñiga 2006; Nikitina 2009; Creissels 2006, 2009; Pantcheva 2009, 2010, 2011; Lestrade 2010, Nintemann et al. 2020), as well as topological relations (Piaget and Inhelder 1956) and demonstratives (Himmelmann 1997; Diessel 1999; Levinson et al. 2018). Other topics such as spatial interrogatives (Stolz et al. 2017) have only recently gained more attention. In previous larger-scale studies on spatial language, especially South American and Papunesian languages tend to be underrepresented. These areas attest to great linguistic variety, the documentation of which has only started to catch up in the last years. As there is no general typology of spatial language previously undescribed features may arise via new descriptions and challenge existing definitions. Gerstner-Link (2018: 734), in a description of Kilmeri, a Border language of Papua New Guinea, addresses this and states the following on the issue of topological relations:

Now, there doesn’t exist a full ‘typological inventory’ of topological features so far; rather, in the current state of research one needs to partly rely on features that may turn out to be language specific. It will then be necessary later on to adjust such language specific features to a more general inventory.

This also holds true for the greater picture of spatial language such as motion event coding. Traditional accounts of motion events in relation to a Ground entity are often based on SAE languages such as English. For instance, Jackendoff’s (1983: 165) notion of “… bounded paths, [which] include[…] source-paths, for which the usual preposition is “from”, and goal paths, for which the preposition is “to”, which may lead the reader to infer that the relevant parts of speech are grammatical items and not lexical items.2 Similarly, Frawley (1992: 173) states that “[i]n all languages it is possible to encode both the origin of the motion, the source, and the destination of the motion, the goal. The following expression makes this clear: […] John ran from the back door to the front door”. The English sentence cited by Frawley (1992: 173) is somewhat unfortunate in the light of ‘radically verb-framing’ languages (see example (5) below) which require two VPs to express Source and Goal of the same motion event, respectively (Wälchli and Zúñiga 2006; Bohnemeyer 2003; Bohnemeyer et al. 2007). The next section sheds light on the issue of

2 Jackendoff (1983), however, does not claim that such formal specifications are true for languages other than English.
diverging coding patterns and coding preferences, and discusses the role of verbs as opposed to grammatical markers in spatial language.

1.2 From spatial interrogatives to general coding tendencies

In their large-scale comparative study on Place/Goal/Source interrogative paradigms, a precursor to (I) and Nintemann et al. (2020), Stolz et al. (2017: 596) state that “[w]herever there are different degrees of complexity […], there is an overwhelming probability that the complexity increases from WHERE via WHITHER to WHENCE”. They further conclude that maximally syncretic behavior of the three spatial interrogative forms, which correlates strongly with Place/Goal/Source distinctions being expressed by Path-conflating verbs, is virtually absent from Europe. This is in line with Creissels’ (2006: 20) study on Place/Goal/Source paradigms in declarative expressions, which points to two dominant coding patterns in Europe, i.e. Place=Goal syncretic forms as the first option, and distinct forms for all three relations as the second option. Stolz et al. (2017: 635–641) further discuss whether there tends to be symmetry or asymmetry between spatial interrogatives and their declarative equivalents across languages. Nintemann et al. (2020) approach this question by comparing spatial interrogative forms and their ‘declarative equivalents’, such as spatial adverbial forms, in 250 languages. In Europe, these declarative equivalents are often analyzed as adverbs, but in their study, the relevant forms are defined functionally as translational equivalents of English (to/from) here, (to/from) there, and so on, since a formal definition of ‘adverb’ is difficult to maintain in the light of variation across languages.

As an introductory example, Standard German illustrates how closely spatial interrogatives and spatial adverbs can be formally related. The paradigm of spatial adverbs, despite being “highly lexicalized” (Diessel 2003: 644), is largely morphologically transparent and regular. It reflects the coding pattern of the corresponding spatial interrogatives (1) as long as only the distal forms are considered (2).

(1) Standard German spatial interrogatives (Nintemann et al. 2020: 3)

a. Place

---

3 Stolz et al. (2017) employ the three Early Modern English interrogative words as terminological placeholders for Place/Goal/Source interrogative forms and take constructional length as a proxy for formal complexity.

4 Forms corresponding to English spatial adverbs can be homonymous with, or derived from, demonstratives, but a cross-linguistically valid definition of ‘adverb’ per se seems difficult or impossible (Schachter and Shopen 2007: 20; Ramat and Ricca 1994).
Wo ist er?  
where be:3SG.PRES 3SG.M  
‘Where is he?’

b. Goal  
Wohin geht er?  
where.to go:3SG.PRES 3SG.M  
‘Where does he go?’

c. Source  
Woher kommt er?  
where.from come:3SG.PRES 3SG.M  
‘Where does he come from?’

(2) Standard German distal deictic answers  
(Nintemann et al. 2020: 3)

a. Er ist dort.  
3SG.M be:3SG.PRES there  
‘He is there.’

b. Er geht dorthin.  
3SG.M go:3SG.PRES to.there  
‘He goes there.’

c. Er kommt dorther.  
3SG.M come:3SG.PRES from.there  
‘He comes from there.’

The interrogative forms wohin and woher are derived from the spatial interrogative stem wo ‘where’ and contain the directional clitics hin and her, which may be interpreted as Goal and Source markers also in combination with the distal deictic adverb dort ‘there’. However, taking into account the proximal spatial adverb hier ‘here’ in combination with hin and her, both forms hierhin and hierher are parsed as Goal-oriented forms meaning ‘to(wards) here’ in most varieties of contemporary Standard German. An alternative to the potentially ambiguous hierher is von hier ‘from here’, employing the Source preposition von ‘from’, which also functions as a Source marker with the distal

---

5 In fact, the Atlas zur deutschen Alltagssprache (Atlas of German everyday language) shows that her is more common than hin in imperative expressions, such as Komm’ hier her! ‘Come here!’, but hin is more common in some parts of Western Germany (Elspaß and Möller 2003: Heft III, <https://www.atlas-alltagssprache.de/runde-3/f04a-f10d/>; accessed May 16, 2023).
adverb (3a) and the spatial interrogative (3b). The preposition von is the default strategy to express motion from a nominal Source Ground (3c).

(3) Standard German Source expressions

a. *Sie kommt von dort.*
   3SG.F come:3SG.PRES from there
   ‘She comes from there.’

b. *Von wo kommt sie?*
   from where come:3SG.PRES 3SG.F
   ‘Where does she come from?’

c. *Sie kommt von der Schule.*
   3SG.F come:3SG.PRES from the:DAT.F school(F)
   ‘She comes from school.’

The Standard German paradigm is therefore irregular due to the syncretism of the proximal deictic adverbs and due to several solutions for Source, where also regional variation must be taken into account (Elspaß and Möller 2003: Heft III). Also Goal expressions may vary in Standard German, mostly due to Ground type and fixed expressions. Nominal Grounds receive overt marking via the preposition zu ‘to’ and a gender-sensitive article in dative case, which frequently contracts to a gender-sensitive form such as zum in (4a). In contrast, most toponyms behave like spatial adverbs and receive the preposition nach (4b). The preposition nach ‘to(wards)’ also derives adverbs from dative-case nominals such as nachhause ‘home(wards)’ from Hause ‘house.DAT’ in (4c), which are paradigmatic with spatial adverbials and do not combine with Goal prepositions. Consequently, the Goal argument in (4c) can also be expressed with the adverb heim ‘home’. There is no corresponding Goal interrogative form involving either of these two prepositions.8

(4) Standard German Goal expressions

a. *Sie geht zum Haus.*
   3SG.F go:3SG.PRES to.the:DAT.N house(N)
   ‘She goes to the house.’

---

6 There are exceptions, such as some toponyms that have grammatical gender such as in die Schweiz ‘(lit.) into the(F) Switzerland’, especially with rivers and street names (Julia Nintemmann, p.c.).

7 The same applies to English adverbs derived from nominals, cf. She goes home versus She goes to the house.

8 However, the interrogative expression Von wo nach wo? ‘From where to where?’ is acceptable in informal use of several German varieties.
b. **Sie fährt nach Uppsala.**
   3SG.F drive:3SG.PRES to TOPO
   ‘She drives to Uppsala.’

c. **Sie geht nachhause/heim.**
   3SG.F go:3SG.PRES home(wards)
   ‘She goes home.’

The Standard German examples also illustrate that the same deictic motion verbs, such as *kommen* ‘come’, can appear in Goal and Source expressions, and that Path is expressed primarily by adpositions, in a satellite-framed way (Talmy 1985). The presence of dedicated spatial markers for Goal and Source is a prerequisite to connect two Grounds in the same syntactic unit with a single motion verb only, such as English *She ran from here to there*, whereas RADICALLY VERB-FRAMING languages (Bohnemeyer 2003) require a VP for every Ground argument. Example (5) of the Mixe-Zoque language Soteapaneco illustrates this pattern, despite the intrusion of a loan marker. In the absence of a Goal marker in (5), both Ground arguments *yiim* ‘here’ and *jeexic* ‘there’ occur in distinct, juxtaposed clauses and must occur with dedicated Goal and Source verbs, despite the overt marking of the Source Ground via the Spanish loan preposition *de* ‘from’.

(5) Soteapaneco (Robbers and Hober 2018: 418)

```
Caay-i de yiim, nje-i jeexic
remove-IMP from here go-IMP there
‘Move from here to there.’
```

Previous studies (Creissels 2006; Wälchli and Zúñiga 2006: 289; Nintemann et al. 2020; Stolz et al. 2017: 596) point out that especially Mesoamerican and Sub-Saharan African languages tend to not employ grammatical markers for Place/Goal/Source functions but rely on verb-conflation. The pattern is also not uncommon in other areas of the world, but is absent or rarely attested in Europe (Creissels 2006; Stolz et al. 2017; Nintemann et al. 2020, inter alia). The following example (6) shows a radically verb-framed solution to express Goal and Source in Yine (Arawakan). The strategy to express Place, Goal and Source by verb-conflation consistently applies across different Ground types such as nouns (6b) and spatial adverbs (6a). Atelic direction is also expressed

---

9 As a sole exception, Nintemann et al. (2020: 245) discuss verb-conflating coding of Place/Goal/Source with spatial adverbiaI Grounds in Adyghe (Abkhaz-Adyge), a language which also relies on verbal semantics as well as on reversive and tempora-aspectual morphology to express these three relations (Smeets 1984).
by geocentric river-based motion verbs and spatial adverbs (Matteson 1965; Hanson 2010).10

(6) Yine (Hanson 2010: 57, 205, original boldface omitted)

a. hewi n-hapoka-tka
   here 1SG-arrive-PFV
   ‘I arrived here.’

b. wane hima r-hifpaka-çe-m-ta-ya
   there/thus QUOT 3-come.out-FREQ-NONDUR-VCL-APPL

   kanawa
   canoe
   ‘A canoe would always emerge from it, reportedly.’

The radically verb-framing strategy can be described as consistent and symmetric coding pattern but detailed and language-specific studies may reveal differing repertoire sizes for Goal versus Source verbs. It is also questionable whether potential ‘complexity’ in the sense of Lupyan and Dale (2010) can be assessed for these languages which express man spatial functions with verbs rather than with grammatical forms. Taking zero-coding of Place, Goal and Source expressions in interrogatives and adverbial equivalents as a proxy for Goal and Source expressions coded by verb-conflation11, Nintemann et al. (2020: 292) present the global distribution of (a)synergetic coding patterns in 250 languages (Figure 1). The high frequency of Pattern I may be a result of convenience sampling and the tendency in grammatical descriptions to discuss relevant forms whenever they show morphological distinctions in the three domains (as opposed to radical verb-centric coding).

10 (II) and additional field notes indicate that all of the riverine orientation terms are more present in the usage of senior speakers of Bufeo Pozo Yine, but this must be assessed with quantitative studies involving more speakers and several Yine-speaking communities.

11 Place may likewise be coded by verbs, however, this is more often subject to zero-coding and zero-anaphora (V).
Figure 1 (taken from Nintemann et al. 2020: 292): Pattern I (red) refers to distinct interrogative and adverbial forms for Place, Goal and Source; Pattern I/Pattern II (yellow) refers to doculects with distinct forms but optional Place=Goal syncretic coding; Pattern II (green) refers to Place=Goal syncretic coding; Pattern V (blue) refers to Place=Goal=Source syncretic forms which strongly hints at radical verb-framing; Pattern V as equal option (purple) refers to doculects that attest to fully syncretic paradigms with optional overt grammatical coding; Other (turquoise) refers to Goal=Source and Place=Source syncretic paradigms.

Figure 1 shows that Goal=Source and Place=Source syncretic interrogative and/or spatial adverbial forms are globally rare. A language that pervasively employs this pattern across all domains is Skolt Saami (Feist 2010), as well as other North and Eastern Saami varieties (Olle Kejonen, p.c.; Nintemann et al. 2020: 195-197). Example (7) shows the Place=Source syncretic deictic forms in contrast to a dedicated Goal form.

(7) Skolt Saami (Feist 2010: 227, 226, 354)

a. Place
   to'ben  jälstiim  mängg  piârrjed
   there  live.PAST.IPL  many.SG.NOM  family.PART
   ‘there we lived, many families’

b. Goal
   tok  mâ’ne  ñkâcâd  jiânnai
   thither  go.PRES.3PL  look.INF  much

   oummu
   person.PL.NOM
   ‘a lot of people went there to look’
Skolt Saami Place and Source expressions are thus differentiated by the semantics of the accompanying verbs, and potentially by inter-sentential, sentential or situational context. Also Creissels (2006) and Nikitina (2009) observe that radically verb-framing languages do not only attest to subsets of motion verbs that inherently refer to Source-oriented or Goal-oriented motion, but other strategies involving context and inference. This insight matches results of (III) and (V) which indicate that tendentially radically verb-framing languages do not always employ distinct sets of dedicated Goal verbs and Source verbs to code the two relations, but that both expression types can be headed by Goal verbs. Iconic ordering (Schapper 2011) of Place and Goal forms, as well as lexemes denoting punctual events such as ‘start’ verbs, lexical semantic contrast, deictic location, and context play crucial roles for Source expressions across languages (III).

In (III), these types of Source expressions are discussed in detail, considering cross-linguistic variation but with the exception of nonverbal cues such as gesture due to lack of data. Example (8) shows an expression from a description of the Mande language Wan which, in isolation and without situational or discourse context, can be parsed as either Goal or Source. In (9), Ese Ejja (Pano-Takanan) shows a non-ambiguous and cross-lingually attested coding strategy to express Source (III), which consists of a Place-denoting component, here a locative-marked Ground argument and a non-finite posture verb, which is followed by a motion verb. The strategy relies on two components, the Place-denoting form and the motion verb, none of which express Source in isolation, and which must be iconically ordered to identify the Place-marked Ground argument as the original location before the location change expressed by the motion verb. Hence, parsing of the Ground argument as Source of motion is inferred.

Goal and Source can also be distinguished via lexical contrast. (10) shows an expression in Yine in which Source is coded via the lexical contrast between the spatial adverb *teno* ‘high’ and the verb *hiçrika* ‘fall, descend’. *Teno* can also express ‘deep’ in the sense of ‘deep down’ (Hanson 2010), for instance referring to an object underwater. However, due to semantic restrictions in the combination of lexemes in (10) it is parsed as ‘high (up)’.
(8) Wan (Nikitina 2009: 1125)

è zō bā è lé
3SG.SBJ come.PST field DEF at

bé é p5 l5
then 3SG thing eat.PAST

‘He came to the field and had a meal’ or ‘He came from the field and had a meal.’

(9) Ese Ejja (Vuillermet 2021: 245; 265)

[E-iy=jo neki](=0) towaa-ani.
NPF-hill=LOC sit.NF(=ABS) jump-PRES

‘He is jumping from the hill.’ (Lit. Standing on the hill he’s jumping.)’

(10) Yine (Hanson 2010: 278)

tenö r-hičrika-ya-tka
high 3-descend-APPL-PFV

‘He fell from high up.’

The importance of context, deictic location and inference is outside the scope of (I) but is highlighted in (III), the results of which indicate differences in orientation frame use shown by young speakers’ and senior speakers’ of Yine. Although not enough information on orientation frame use could be collected from grammatical descriptions to be included and compared in the final study spanning 35 languages (V), all of the included studies point to the salience of languages that primarily employ verb-conflation to express motion, location and orientation and to the importance of perspective, be it deictic or situational. Among relevant verb stems and verb forms, marker-like behavior is attested especially for Source expressions (III), and research conducted in the course of this thesis points to a larger distribution of verb-centric coding strategies than previously assumed.
2 Overview and results of the compilation thesis

The present dissertation aims to fill descriptive gaps and addresses several topics that relate to motion and orientation in space from different viewpoints. All of the compiled studies shed light on underresearched topics and languages, and insights gained over the course of the doctoral studies conflate in the comparative study in (V). As introduced with the examples of Standard German above (Section 1.2), an unsurprising finding is that, across languages and areas, within-language variation poses a problem for per-language typologies of orientation, motion and location. This section discusses the articles of the compilation thesis in chronological order. However, experimental sampling and data evaluation for (V) were conducted over the course of the entire doctoral studies.

Connecting to previous and collaborative research, (I) is part of a large-scale comparative project on (a)symmetric coding of Place, Goal and Source with spatial adverbial Ground arguments in 250 languages, the results of which have been published in Nintemann et al. (2020) as a follow up to Stolz et al. (2017) on spatial interrogative forms in 437 languages.¹² (I) provides a qualitative discussion of the data which precedes the quantitative evaluation in Nintemann et al. (2020) and discusses several types of morphological syncretism of Place, Goal and Source as shown by spatial adverbial forms. Nintemann et al. (2020) and other publications from the project (Robbers and Hober 2018; Robbers 2020) approach coding (a)symmetry by assessing morphology of spatial adverbial forms, such as English here and there, and conclude that Source constructions tend to be longer than Goal constructions, which in turn exceed Place constructions in terms of constructional length. However, (I) and Nintemann et al. (2020: Chapter 6) address several issues concerning the comparability of these deictic forms, as languages with dominant geocentric tendencies may show preferences towards landscape-specific

¹² Place, Goal and Source are terms based on Jackendoff’s (1983) semantic primitives which are defined by Stolz et al. (2017: 1) as “the location of an entity in space”, “the endpoint of the movement of an entity in space” and “the starting point of the movement of an entity in space”, respectively. Figure and Ground are terms taken from Talmy (2000) which receive a handy definition in Admiraal’s description spatial language in Baure (Arawakan): “The Figure is the object that undergoes the motion, with respect to the reference object, the Ground. The Path is the trajectory that the Figure follows, or the site that it occupies” (Admiraal 2016: 8).
forms. Different intuitions came together during the writing of (I) and of Nintemann et al. (2020), as the authors were responsible for the sampling of languages from different macro areas. For Nintemann et al. (2020), of which (I) is a precursor, Julia Nintemann (University of Bremen) sampled and analyzed 50 languages from Africa and 50 languages from Asia, Maja Robbers (Uppsala University) sampled 50 languages from Oceania and 50 languages from the Americas, and Nicole Hober (University of Bremen) sampled 50 languages from Europe. All three authors conducted the statistical analyses for Nintemann et al. (2020) and wrote the manuscript. For (I), both authors selected and analyzed linguistic examples from their subsamples and wrote the manuscript with equal division of labor. Some of the main results of Nintemann et al. (2020) are summarized in Figure 1 above.

During data collection for (I) and Nintemann et al. (2020) it became clear that especially many Papuan and South American languages, albeit underrepresented in their convenience sample, rarely attest to ‘unmarked’ spatial deixic coding such as ‘(to) here/there’ but tend to show geocentric preferences, such as for dedicated geography-based lexemes or markers such as ‘up/downriver’. Since comparative and typological studies usually deal with data that describe doculects (Cysouw and Good 2013), i.e. the specific variety of a language as documented in a grammatical description, for every linguistic topic it is worthwhile to test for variation within the same speaker communities. A case in point is the Arawakan language Yine, spoken by river-dwellers in the Amazon rainforest of Peru. Using the Ball-and-Chair stimulus (Bohnemeyer 2011; Pérez Báez 2008: 29–32) to elicit spatial reference frames, (II) deals with linguistic variation in a Yine community and hypothesizes that not age but contact with Spanish leads to the absence of river-based spatial forms in young people’s Yine. Yine is a strongly verb-centric language that employs formally unrelated spatial adverbs and motion verbs that denote river-based orientation with a strong upriver/downriver axis. Across languages, geocentric spatial orientation can manifest in both grammaticized forms such as directional markers and in lexical items. (II) strongly indicates that younger Yine speakers tend to employ egocentric reference frames, which may be caused by contact with Spanish. Therefore, contact as a contributing factor for within-language variation is raised as an important issue from the viewpoint of field experiments in (II) and is again part of the findings of the comparative study (V).

Yine employs Path-conflating verbs and no grammatical marking to express motion to a Goal and motion from a Source, a strategy which is unusual enough from a European perspective but in fact very widespread outside of Europe (Nintemann et al. 2020). Focussing on a subtype of verb-centrism, (III) deals with languages that attest to Source coding strategies that include neither Source-conflating verbs nor dedicated grammatical material such as English from. This BIPARTITE Source expression type builds on a locative element which is followed by a Goal-oriented element to express Source via inference
from the combination of the iconically ordered parts (Schapper 2011), i.e. ‘Figure is at a location’, followed by ‘Figure goes (elsewhere)’. Abui (Timor-Alor-Pantar) attests to this pattern which has not been noted in full detail in (I) due to the focus on morphological (a)symmetries, but receives dedicated treatment in (III). Data discussed in (III) partly stem from manual convenience sampling for (I) and Nintemann et al. (2020) but also from the novel computer-assisted sampling process (IV) for the comparative study of spatial language in 35 languages (V), and from additional personal communication with colleagues answering a dedicated query in the Lingtyp mailing list in February 2021. The parallel subsampling for bipartite Source expressions, previously acknowledged by only few researchers on spatial language (such as Bourdin 1997: 189; Kopecka and Ishibashi 2011), hints at a greater distribution of this pattern in the world’s languages than predicted by convenience samples such as Nintemann et al. (2020).

Bipartite structures and other Source coding strategies that neither employ lexical forms nor grammatical marking of Place/Goal/Source relations but strongly involve element ordering, inference and/or pragmatic context, are attested in almost a quarter of the genealogically balanced variety sample of 35 languages in (V) for which the sampling strategy is described in detail in (IV). (IV) addresses a novel computer-assisted sampling technique using the Gramfinder tool (Hammarström 2021a; 2021b) over the DReaM corpus (Virk et al. 2020) to select descriptions of natural languages which are ranked for informativity according to pre-selected keywords that signal discussions of the desired linguistic topic. Experiments with several keywords, such as *laryngeal voice/vowel* or *switch reference*, were ran by Harald Hammarström and described in (IV), along with the case study on motion which lays the base for the variety study (V) conducted by Maja Robbers. Place/Goal/Source (a)symmetries across Ground types are the focus of (V), a study which aims to provide an overview of coding patterns in the world’s languages based on the most ‘information-dense’ descriptions returned by keyword-based sampling (IV).

The keyword *motion* returned a sample of descriptions that give special attention to the topic, so that a plethora of variables relating to space and motion could be filled (see the online appendix of (V) for a conflation of all variables). In total, 35 grammatical descriptions sampled for informativity (IV) were screened for relevant discussions and linguistic examples. Apart from variables relating to Place/Goal/Source coding preferences with different Ground types, such as animate Grounds and toponyms, other variables concern, for instance, distance levels coded by demonstratives or deictic directional marking. One of the most important results of (V) is that only few languages can be described as consistent in their coding patterns across Place, Goal and Source, and even less so when Median Grounds and other Ground types such as toponyms or deictic adverbs are taken into account. Consistent,
i.e. predictable patterns are provided either by grammatical markers or by radical verb-framedness (see examples [5] and [6] above). Especially Source expressions are subject to coding asymmetries, perhaps reflecting lower frequency in speech as opposed to Goal and Place events. Studies on the frequency of these spatial relations have been conducted for European languages (Verkerk 2017; Stefanowitsch and Rohde 2004).

As briefly touched upon in the introductory section with the example of body part roots in Mesoamerican languages, it is clear that certain parts of speech may play crucial roles for several spatial functions which can be unrelated in other languages. The results of (V) show that also more narrowly defined functions that are in paradigmatic distribution in certain languages, such as consistent or regular Goal-Source coding by grammatical items such as case-markers or adpositions, can be formally distinct and hence not paradigmatic in others, for instance due to bipartite Source expressions (III) (see [9] above). Based on (V) it is further argued that only few languages can be analyzed as belonging to a ‘type’, such as consistently case-marking languages or consistently radically verb-framing languages, but that many or most languages show both tendencies to certain degrees. Acknowledging a potential bias against previously better described (or better understood) coding strategies (IV), the evaluation of 35 sample languages (V) suggests that at least half of the sample relies primarily on coding Place/Goal/Source via verb-conflation and grammatical marking tends to be absent or optional. For the remaining languages, grammatical marking may be present at least for Source, but verb-conflation is still crucial for Goal, as roughly three-quarters of the sample attest to frequent or optional zero-coding of Goal. While most languages code stative Figure-Ground relationships overtly with a locative or oblique marker, a bare motion verb suffices in most instances to express Goal-oriented motion.

However, the results vary to Ground type. The results of (V), summarized in its online appendix, are pointers to fruitful future studies on several topics within the research on spatial language. For instance, roughly one-third of the sample shows distinct behavior when the Ground is a toponym. This may include omission of markers as well as overt coding with dedicated material, such as loan markers. Example (11) shows a part of a route description in the Afro-Asiatic language Barain, which is in contact with Chadian Arabic, from which the included form min ‘from’ is likely borrowed. This overtly marked Source expression with a toponym Ground follows a verb-centric Source expression with a spatial adverbial Ground argument kakkireŋ ‘there’ and a serial verb juk-o ‘stand up’ which expresses Source in combination with the following motion verb kol-o ‘go’.

(11) Barain (Lovestrand 2018: 2)

\[
\text{kakkireŋ} \quad \text{juk-eyi} \quad \text{na} \quad \text{ni} \quad \text{kol-eyi}
\]

\[
\text{there} \quad \text{stand.up-IPFV} \quad \text{BG} \quad \text{SBJ.3PL} \quad \text{go-IPFV}
\]
Radically verb-framing languages often attest to borrowing of grammatical Source markers, for example in Mesoamerica (Hober 2020). The intrusion of *min* ‘from’ may depend on several factors, so that future studies shall assess whether toponyms trigger overt marking by borrowed items rather than nouns, or loan items such as non-indigenous toponyms rather than original toponyms (Julia Nintemann, in preparation). Within-community variation is possible also concerning contact phenomena and contact-induced changes, as the degrees of exposition to contact languages with distinct coding patterns may differ within communities and across occupations and generations. This is paralleled by distinct preferences in orientation systems, relating to the same factors such as age or subsistence mode, which is explored in (II). Also due to ongoing contact-induced change, lesser-studied coding strategies, usually involving verbs instead of grammatical markers, received special attention in the present compilation thesis on spatial language. The presence of verbs reflecting certain anchors such as riverine terms in Yine (II) is seen as an equally strong indicator for general orientation preferences as grammaticized items in languages with other formal coding tendencies. Results from the field (II) as well as from smaller-scale comparison (III and I) and larger-scale cross-linguistic comparison (V) with an improved sampling technique (IV) all point towards the importance of working towards a better integration of potentially challenging factors such as within-language variation, contact phenomena, and open class forms such as verbs being pivotal in the expression of motion and orientation.
3 Outlook

In light of steadily emerging data from previously underdescribed linguistic macro-areas on the one hand, and the ongoing loss of local languages around the globe and shrinking linguistic diversity on the other, the research on spatial events is far from exhausted. A useful starting point when documenting spatial features of a given language or variety is Wälchli’s (2001) cross-linguistically informed checklist for motion events which also addresses functions relating to demonstratives, spatial adverbs and deixis in general, as well as animacy distinctions and frames of reference. The ongoing endeavor to work towards a typology of spatial language also profits from dedicated comparative studies on neglected aspects of motion event descriptions such as Kopecka and Ishibashi (2011) on multipartite Source expressions in Polish, Japanese, Yek‘wana and Ese Ejja. The variety of motion event coding strategies and other features of the spatial system of natural languages across the globe also calls for detailed studies on the exact functions and scopes of motion and location verbs in the style of Nakazawa (2007), Wilkins and Hill (1995) and Matsumoto et al. (2017).

So far, there is no consensus on the distribution and the variety of coding strategies of motion, location and orientation expressions when taking into account all relevant research, such as individual descriptions of specific languages. A comprehensive typology of space will gain much from taking into account element ordering, context and inference from lexical cues as building blocks of language. Moreover, language-specific or community-specific preferences for geocentric orientation can have effect on motion event expressions. Geocentric forms are often paradigmatic with (i.e., belong to the same class or subset as) ‘unmarked’ deictic forms, and are often preferred in local languages and varieties (V). Since languages attest to a great variety of coding strategies (V), it will be worthwhile to work towards shared definitions of fine-grained variables to facilitate the creation of shared databases and to create opportunities for detailed analyses, such as Bickel’s (2015) multivariate approach. To tackle areal and genealogical biases in typological and cross-linguistic samples, the knowledge gained from newer descriptions especially of Papunesian and South American languages must be integrated into the effort to understand and grasp variety in spatial language.


Hammarström, Harald. 2021b. Gramfinder: Human and Machine Reading of Grammatical Descriptions In Eduard C. Dragut and (ed.), 3rd Workshop on Data Science with Human in the Loop, DaSH@KDD, Virtual Conference, August 15, 2021, 1-6. DBLP.


