Typological challenges, cognitive insights, pragmatic explanations: Investigations on Movima, a linguistic isolate of lowland Bolivia

Katharina Haude

Université de Paris Ecole Doctorale "Sciences du language", ED 622

MÉMOIRE DÉPOSÉ EN VUE DE L'HABILITATION À DIRIGER DES RECHERCHES

Typological challenges, cognitive insights, pragmatic explanations: Investigations on Movima, a linguistic isolate of lowland Bolivia

Katharina HAUDE

JURY:

Mme. Claudine CHAMOREAU, DR, CNRS-SEDYL (examinatrice)

M. Spike GILDEA, PU, Université d'Oregon (rapporteur)

Mme. Françoise ROSE, DR, CNRS-DDL (présidente)

M. Stavros SKOPETEAS, PU, Université de Göttingen (rapporteur)

M. Otto ZWARTJES, PU, Université de Paris (garant)

Mme. Stéphane ROBERT, DR, CNRS-LLACAN (examinatrice)

3 décembre 2020

Table of Contents

DOCUMENT DE SYNTHESE

RESEARCH SUMMARY

Acknowledgements

This thesis sums up my main research activities and research topics since 2006, when I defended my doctoral dissertation at the Radboud University Nijmegen. First of all, I wish to express my gratitude to the directors of the institutions that hosted me during these years and provided me with excellent work conditions. This was, from 2006 to 2010, the Department of Linguistics of the University of Cologne, where my postdoctoral projects benefited greatly from the appreciation by the directors Hans-Jürgen Sasse and, later, Nikolaus P. Himmelmann. In 2010, after my recrutement at the CNRS, I joined the newly-founded research lab *Structure et Dynamique des Langues* (SeDyL), whose successive directors Anaïd Donabédian, Isabelle Léglise, and Sophie Vassilaki have always provided unconditional support for the way in which I carried out my research.

Becoming part of the French academic system meant a great change in my professional and personal life. I discovered a country that had seemed inaccessible to me because I didn't speak the language well enough, and discovered the beauty of a language that before had seemed impossible to speak or understand. Both the country and the language strike me now as wonderful. I would never have had this opportunity if it hadn't been for Françoise Rose, who invited me to the CELIA (the *Centre d'Études de Langues Indigènes d'Amérique*, the predecessor of the SeDyL) in 2007. Its director, Francesc Queixalós, then invited me to his monthly programme *Ergativité* and suggested that I apply for a CNRS position, which I would never have thought of otherwise. For his help in succeeding I will always be grateful. I also thank Duna Troiani, whose care after the difficult job interview is unforgettable; and Isabelle Léglise for keeping me updated on how things worked at the CNRS.

Having a permanent research position and doing the kind of research one wants to do is invaluable. However, an at least equally big gain I got through my employment at the CNRS were the new colleagues. At the SEDYL, besides the above-mentioned Anaïd, Duna, Isabelle, and Sophie, I am lucky to be (or have been) in the company of Christine Bonnot, Claudine Chamoreau, Denis Costaouec, Outi Duvallon, Laurence Guernalec, Jean-Michel Hoppan, Victoria Khurshudyan, Stefano Manfredi, Alexandru Mardale, Bettina Migge, Annie Montaut, Enrique Palancar, Hélène de Penanros, Deth Thach (†), and Nicole Tersis. The campus of Villejuif has still more to offer: the LaCiTO, where I had the luck to make the acquaintance of Isabelle Bril, Zlatka Guencheva, Marc-Antoine Mahieu, Alexis Michaud, Claire Moyse-Faurie, Lameen Souag, and last but not least Lila Adamou, who was the first to encourage me to do

my *HDR*; and the LLACAN, always welcoming and supportive. Here, I have had the pleasure to collaborate, in a variety of contexts, with Isabelle Alanièce, Christian Chanard, Dmitry Idiatov, Mark van de Velde, Amina Mettouchi, Tanya Nikitina, Nicolas Quint, Martine Vanhove, Jeanne Zerner. Since my former fellow student Yvonne Treis has arrived at the LLACAN, my academic life has become much more fun! Of all the colleagues in Villejuif, I am particularly indebted to Stéphane Robert for her advice and support during the *HDR* process. Among the Paris colleagues outside Villejuif, my very special thanks go to Aimée Lahaussois of the HTL lab, who helped me through all the ups and downs of "having to write a thesis" in the past half year. I don't know how I would have coped without the *enKAHDRment*! The long list above consists almost exclusively of people I did not know before putting a foot in the French research landscape; this may show how much I have gained from this step.

The habilitation process was my first personal experience with the French university system. I am deeply grateful to Otto Zwartjes, who accepted without hesitation to be my HDR supervisor and to accompany me through this complicated procedure. My deep gratitude further goes to the jury members: Françoise Rose, who acted both as a rapporteur and as jury president, to Stavros Skopeteas, who also acted as rapporteur, to Claudine Chamoreau, and to Stéphane Robert.

Unrelated to the institutions, many other people have accompanied me through the part of my professional life that is described on the following pages, be it through collaborative projects, advice, or inspiring discussions. In the wake of the DoBeS programme, which has established a closely-tied community of field linguists, Nikolaus Himmelmann encouraged me to apply for a EuroBABEL project, which led to a very enjoyable collaboration with Balthasar Bickel, Spike Gildea (who hosted me in Eugene during two unforgettable weeks in winter 2011), Eva van Lier, Robert Schikowski, Anna Siewierska (†), Alena Witzlack-Makarevich, Fernando Zúñiga, among others, as well as Elena Skribnik from another EuroBABEL project. During my regular stays at the Cologne Linguistics department I particularly appreciate the discussions with Werner Drossard, Henrike Frye, Sonja Gipper, Birgit Hellwig, Felix Rau, Sonja Riesberg, and Gabriele Schwiertz, and Claudia Wegener. I am grateful to Stavros Skopeteas for letting me use his office in Göttingen in October 2019, where I started writing this thesis.

The research described here is based almost exclusively on the speech data I collected from the Movima speakers in Santa Ana del Yacuma, Bolivia. I wish to take this occasion to thank all the speakers who contributed to the data corpus, be it with just a short anecdote or as

participants in a conversation, be it with many hours of intensive work. The many friendly people who have made this work possible are listed here in alphabetical order, and probably some more have sadly disappeared meanwhile without my knowing it. *Merek solopa:ye iy'bikweł alra*, Balvina Almaquio Saucedo, Alcira Amblo Malale Ovales, Corpos Amblo Ovales, Esaltación Amblo Ovales(†), Jovina Amblo Ovales(†), Natalio Amblo Ovales(†), Germán Barba Mole(†), Leonilda Caumol Chirimani, María Caumol Cujuy, Nataniel Caumol Gualujna(†), Griselda Caumol Mayapo, Masimina Cayalo Almaquio, Peregrina Cayú Masaro(†), Eligardo Chirimani Malue, Natividad Chori Ovales, Cándida Chori Tereba, María Doris Zelady Machado, Santo Federico Zelady Machado, Juan Gualima Dolea, Etelvina Gualujna Amblo, Modesta Gualujna Amblo, Hernancia Gualusna Amblo(†), Nelly Gualusna Saucedo, Hipólito Gualusna Saucedo, Santos Guasinave Cáumol, Fabián Humasa Almaquio, Ajdón Humasa Amblo, Benito Humasa Amblo, María Machado Powoslo, Julia Malale Humasa(†), Francisco Malue Chori, Delmira Masaro Yoqui(†), Baldomero Mole Cáller, Pastora Mole Poso, Eleuterio Mole Vargas, Arnulfo Mole Yalauma, Beatriz Onarri Malue, Pedro Onarri Yalauma, Herlan Rojas Rossel, Ela Rossel Chori(†), Matilde Rossel Humasa, Flora Saucedo Gualima, Dionicio Tonore Cayalo, Angela Tonore Limaica, Brunilda Vaca Ovales, Dora Vargas, Concepción Vargas Mole(†), Laureano Yalauma Onarry, Marcelina Yoqui Cayú, and Juan Zelady Humaza(†).

Last but not least, I wish to thank my little family: Harald, for letting me know, each time I expressed doubts, how much fun it must be to write this kind of thesis, and Pauline, for asking almost every day "On which page are you now?" (I won't tell her that I'm back on page 4).

Thanks to you all!

Cologne, May 2020

Symbols and abbreviations

-	affix
--	external clitic
~	reduplication
< >	infix
=	internal or neutral clitic
1	first person
2	second person
3	third person
AB	absential
ABS	absolutive
ACC	accusative
ADJ	adjective
AGT	agentive
ANT	anterior
ART	article
AV	actor voice
BE	bound nominal element
BEG	begun aspect
CAUS	causative
CLF	classifier
CO	co-participant
COP	copula
D	dummy
DEM	demonstrative
DES	desiderative
DET	determiner
DR	direct
DSC	discontinuous
DUR	durative
EMPH	emphatic
OBV	obviative
EPIST	epistemic marker
ERG	ergative
EV	evidential
EVT	event
EXTL	external argument
F	feminine
GEN	genitive
HOD	hodiernal
HYP	hypothetical
IJ	interjection
IMP	imperative
INSTR	instrumental
INTR	intransitive
INV	inverse
ITN	intentional
LK	linker
LOC	locative
LV	linking vowel
M	masculine
MD	middle
MLT	multiple
MOV	moving
N	neuter
N	noun
NEG	negation
NMZ	nominalization
NOM	nominative
NONFUT	non-future
NTR	neutral
OBL	oblique
PFV	perfective

PL	plural
PRC	process
PRO	free pronoun
PST	past
PV	patient voice
R/R	reflexive/reciprocal
REAS	reason
REL	relativizer
RES	resultative
SG	singular
SPK	near speaker
ST	stative
SUB	subordinate
UV	undergoer voice
V	verb
VALDECR	valency decrease
VBZ	verbalizer

List of Figures

List of Tables

1. Introduction

My motivation to study linguistics was my fascination at decoding words and sentences from an unfamiliar language and trying to identify, understand, and explain the patterns found. I felt a particular sympathy for Native American languages, which dated back to a slightly romantic passion for Native American cultures and languages that I had developed during my adolescence and which culminated, back then, in a 400-item word list of Lakhota assembled from all the books about Lakhota Indians that I could find. (I had no idea that grammars and dictionaries of this language existed and were even available at the university library of my home town.)

Because of this early passion for Native American languages, right after finishing school I spent a year in La Paz, Bolivia, where I took courses in Aymara and Quechua at the *Universidad Mayor San Andrés* and in private lessons. It was also the reason why (once I found out that the university subject "Linguistics" existed) I enrolled at the University of Cologne: Here, at least I was not received with a frown when I announced my interest in Native American languages, since the head of the department, Hans-Jürgen Sasse, had himself worked on the Iroquoian language Cayuga. When choosing my courses I preferred those that took into account language description and the structure of non-Indo-European languages. I particular liked typologically oriented courses that focused on specific problems such as parts of speech, nominalization, grammatical relations or lexical typology.

However, due to my stay in Bolivia, I wanted to focus on South American languages, and so, I started to look for a suitable place elsewhere in Europe where I could go as an exchange student – which was not as easy in pre-Internet times as it is now. To my surprise, the only Linguistics department in Europe that I could find whose staff included experts in more than one language of South or Central America was the Department of Comparative Linguistics in Leiden. The head of the department, Willem Adelaar, was one of the few linguists in Europe at the time who specialized in the area. Luckily, the department of Dutch language and literature at Cologne University had an ERASMUS exchange programme with Leiden University, and in 1996 I was awarded a one-year grant. In Leiden, I followed courses on Quechua, Nahuatl, Cariban, and Guaraní, and additionally participated in an intensive fieldwork course on Bété, a tone language of Ivory Coast, at he Africanist department.

Back in Cologne, I graduated with a Master's thesis on a directional suffix in Aymara, for which I did my first little field work in the country itself: I spent 2 weeks in La Paz doing

elicitation sessions with my former Aymara teacher, Filomena Miranda. When I approached my Master, one of the contacts I had made during my ERASMUS year, Mily Crevels, suggested that I apply for a PhD position at Pieter Muysken's *Spinoza* project in Nijmegen. The topic of the PhD thesis would be exactly the kind of endeavour I had envisaged ever since my first stay in Bolivia: studying one of the so far undescribed languages of the Bolivian Amazon area. This PhD project corresponded perfectly with my passion for analyzing unfamiliar linguistic structures: It gave me the opportunity to describe from scratch a linguistic isolate on which hardly any descriptive material existed so far. The result was a rather comprehensive grammar that covered the most important phonetic, phonological, morphological, and syntactic phenomena of the linguistic isolate Movima.

As a descriptive linguist I could have left it there and moved on to another language, as many linguists do. However, Movima has remained at the center of my scientific interest. It appears to be a typologically unusual language, and I wish to understand its properties better and to make them accessible to the scientific community. In part, this is because I like to go into details. But apart from that, I also see it as my responsibility to continue working on this language, since I am in the best position to do it. Fieldwork has become more difficult with the speakers of this language becoming fewer, and over the last 14 years I have built up a rich data corpus that I am very well acquainted with, having recorded, transcribed, and translated most of it myself.

If I had not worked on Movima, I would not be where I am now. The syntactic phenomena I described raised the interest of Francesc Queixalós, the director of the CELIA (which later became the SeDyL). He invited me to participate in the research programme *Ergativité* of the *Fédération Typologie et Universaux Linguistiques* and to co-edit, together with Gilles Authier, a volume compiling papers by the participants in the programme (Authier & Haude 2012). Queixalós proposed that I participate at the *concours* for the CNRS, through which I got my present position as *chargée de recherche*. This position has allowed me to continue the in-depth study of Movima, independently of current trends that I would be obliged to follow in other research environments. Analyzing the typologically rare and highly interesting properties of Movima from a typological perspective is what I have made my main task. Thus, from someone who enjoys decoding the structure of an unfamiliar, unwritten language I have developed into someone doing in-depth analysis of the now familiar structures of this language. During this process, I have become more typologically (rather than "merely" descriptively) oriented because I try to find the right typological contexts that help assessing the data.

During the years that followed my doctoral exam, I carried out individual and collective research projects, participated in scientific networks, co-edited several collective volumes, wrote scientific papers, and taught or organized seminars for undergraduate students. Therefore, I believe that now I am in a position to advise and take responsibility for graduate students who plan to carry out a dissertation project in one of the areas of my expertise, including the morphosyntax of lesser-described languages, especially American Indian languages, and language description from a non-formalist perspective. This is why I hope to receive my *Habilitation à Diriger les Recherches* on the basis of my previous work and my future research plans, which are outlined in the remainder of this document.

The following chapters describe the research I carried out after acquiring my doctoral degree. In Chapter 2 I give a detailed description of the projects I directed and/or participated in, both international (2) and in France (2.2).

Some general facts about Movima and the data are provided in Chapter 3, with Section 3.1 giving a brief introduction to the situation of the language, Section 3.2 presenting the fieldwork during which I collected the data, and Section 3.3 explaining my work with the data corpus. Section 3.4 describes the materials I created to support the Movima speaker community in their efforts to revitalize their language.

Based on the data I collected in the field, four major topics can be identified that have occupied me most during the last fourteen years. I tried to describe them in Chapters 4–7 in a way that is self-contained; more detailed information can be gained from the articles summed up there. A central topic, detailed in Chapter 4, is the morphosyntactic alignment system of Movima, i.e. the encoding patterns of the arguments of transitive and intransitive clauses. As is detailed in Section 4.1, the Movima alignment system combines traits of ergative, direct-inverse, and symmetrical voice systems. Section 4.2 describes my arguments for explaining this puzzle: Maybe Movima transitive clauses originate from an equational clause pattern with a predicative participle, in which a proposition like *I love my sister* is expressed as *My sister is my loved one*. The problem of grammatical relations and alignment still fascinates me, and I hope to be able to work on it in the future from a broader typological perspective and in the context of more discourse data (see Chapter 8).

While alignment deals with the parallels between transitive and intransitive clauses and the way their arguments are encoded, I have also been interested in the structure of Movima transitive clauses alone. What is the basis for the choice of either the direct or the inverse construction? As in other languages with inverse systems, the choice of either construction depends on interacting factors involving the properties of the discourse referents. These can be

discourse-internal (such as "givenness" of a discourse referent) or ontological (such as humanness): When the actor in the event is given and/or human and the undergoer is new and/or inanimate (e.g. *My sister bought a car*), the direct construction is chosen, and when the actor is new and/or inanimate and the undergoer is given and/or human (e.g. *A car has hit my sister*), the inverse is chosen. However, there are many possibilities in between these extreme endpoints, and often it seems unpredictable when a speaker will choose one construction rather than the other. The three corpus-based studies I published on this issue are described in Chapter 5.

In the studies presented in Chapters 4 and 5 I primarily analyzed basic sentences, i.e., the verb-initial sentences that constitute the most frequent sentence type in actual Movima discourse. Inspired by my participation in the collaborative research programmes *L'énoncé et ses composantes* of the SeDyL and *Information Structure and Grammatical Relations* of the LABEX EFL, I started to work also on non-basic constructions, i.e. sentence types that are less common and seem to be reserved for specific pragmatic purposes, as described in Chapter 6. These constructions usually involve nonverbal predicates, i.e. nouns and adjectives, but also pronouns (Section 6.1). After having identified the syntactic make-up of these constructions, I analyzed their function in discourse (Section 6.2). This led me to revise some of my previous analyses, and non-basic constructions are a matter of ongoing research.

In Chapter 7 I review my work on aspects of Movima grammar that go beyond main-clause syntax and that have occupied me in an on-and-off manner since the beginnings of my post-doctoral career. One is the internal structure of the "referential phrase" (a term that can usually be paraphrased as "noun phrase", but is broader than the latter; see 4.2.2). I worked in some detail on the tense-marking effect of the determiner that marks referential phrases (7.1), and have started to work on the properties of complement and adverbial clauses, which have the form of referential phrases in Movima. I claimed that Movima embedded clauses contain more information than matrix clauses and therefore, contradict what is generally assumed about subordination (7.2).

In each of the chapters 4–7 I list the articles that deal with the topics treated there in more detail. In some cases one subsection corresponds closely to one single article (e.g. Sections 4.1.4, 5.1, or 7.2), but since most of the topics are interrelated, the content of a single paper is often spread over several subsections. In that case, the papers dealing with a larger topic are listed at the beginning of the chapter or larger section (e.g. Chapter 4). References to the different studies are given throughout the text and can be found in the reference list at the end of this document.

Chapter 8 starts with a summary of my research activities (**Erreur ! Source du renvoi introuvable.**) and gives an outlook of the ways in which I intend to broaden my perspective on the basis I have built up so far. A larger research question, which still requires more fine-tuning before it can develop into a more concrete research project, regards the interesting correlation of some typological features in predicate-initial languages (8.2). In 8.3, I describe planned and ongoing research activities that deal with the interaction between discourse and the choice of morphosyntactic constructions.

2. Research activities since 2006

After finishing my doctorate, I successfully applied for several international research projects, for which I got full-time post-doctoral research contracts (Section 2). Since 2007 I have also continuously been involved in different collaborative research programmes in France (Section 2.2).

2.1 International projects

2.1.1 The DobeS documentation project (Volkswagen Foundation)

My first research project after I finished my doctoral thesis was financed by the DobeS programme (*Dokumentation bedrohter Sprachen*/Documentation of Endangered Languages) of the Volkswagen Foundation. Taking the two phases together, *Documenting Movima, an unclassified language of the Moxos region (Bolivia)* (2006-2009) and *Making Movima visible: documenting a linguistic isolate in the Moxos cultural complex* (2010-2013), this project had a duration of 7 years.

The main goal of the DobeS project was to record the speech of those who had learned Movima as their first language in order to create a large, structured database of as many different types of spoken texts as possible produced by as many speakers as possible. The interdisciplinary team included, besides myself, a doctoral student in social anthropology supervised by Nikolai Grube at Bonn University, and three student assistants specializing in General Linguistics, Computational Linguistics, and Media Arts, respectively. The anthropologist and I recorded over 50 Movima speakers, both men and women, on audio and, whenever possible, video. In the interest of the PhD project (which unfortunately never came to a conclusion), we prioritized autobiographical narratives that could serve as an empirical basis for investigating the concept of identity in the Movima community. This resulted in

approximately 100 hours of recorded material (Haude and Beuse 2006–), of which 21 hours were annotated and form over two thirds of the corpus that constitutes the basis of my research. The data collection and the resulting corpus, as well as the materials for the speaker community that issued from this project, are reported on in Sections 3.2–3.4.

Being the coordinator of this project has allowed me to gain experience beyond the domain of linguistic research, such as project management, personnel management, and financial administration. I also got thoroughly acquainted with the at that time most up-to-date linguistic annotation and archiving software ELAN, IMDI, and LAMUS, which were developed at the Max-Planck-Institute for Psycholinguistics in Nijmegen. Furthermore, the over 50 teams that constituted the DoBeS programme formed a solid community, which came together once a year for a workshop at the Max-Planck-Institute. At these workshops, practical issues of fieldwork, the interaction with the speaker community, and results of the documentation projects, as well as ethical issues of archiving and public access were discussed. Participating in this large collaborative enterprise was an excellent opportunity to meet some of the most important linguists in the field and to establish solid contacts with colleagues in Germany and abroad. I also got more deeply involved on the organizational level when I was elected into the scientific committee of the programme. I was also part of a "task force" for establishing a Code of Conduct for DobeS teams; this latter was founded because the documentation of a language involves complicated technical and social issues for teams with an intricate hierarchical structure (PI, doctoral students, postdocs, student assistants, etc.).

Aside from the collection of data, an important goal of the DobeS project was to provide materials that would support efforts in the language community to revitalize the Movima language. Throughout the project I did my best to communicate the idea of language documentation to the speaker community. Among other things, I brought a poster (see Figure 1) in order to explain the DoBeS programme, the recording and annotation procedure, and the storage in an online archive.

Figure 1. Poster explaining the DobeS project

2.1.2 Referential Hierarchies in Morphosyntax (EuroBABEL-ESF) and The Movima Inverse From A Typological Perspective (DFG)

In 2008, I successfully submitted a project for the 3-year programme EuroBABEL ("Better Analyses Based on Endangered Languages"), which was part of the EuroCores programme of the European Science Foundation and whose intention was to combine the documentation of endangered languages and linguistic analysis. The setting was rather complex and the application process highly complicated: Through the European Science Foundation, five collaborative research projects (CRP) would be assembled, each of which should consist of individual projects from at least 2 pre-selected European countries or the US, each of them financed by their individual national funding institutions. Of the 20 proposed projects, only seven were selected for EuroBABEL.

The RHIM project assembled 5 individual projects based in Germany, England, Switzerland, and the U.S., which were led by experts in the domain of morphosyntactic typology. The individual projects, each with their own research team comprising PhD students and/or

postdocs, are listed in Table 1. My individual project also included a doctoral candidate, Robert Keller, who prepared a dissertation on the Movima inverse within the RRG framework and under the supervision of Robert D. Van Valin (HHU Düsseldorf). Unfortunately the dissertation did not come to a conclusion.

Table 1. Individual projects in the RHIM project

Individual Project	**PI and team**	**University and funding Institution**
IP 01: Differential agreement vs. differential case	B. Bickel (with G. Iemmolo, R. Schikowski, A. Witzlack-Makarevich)	Leipzig/Zurich (DFG)
IP 02: Sahaptian and the evolution of hierarchical systems	S. Gildea (with J. Jansen)	Oregon (NSF)
IP 03: The Movima inverse from a cross-linguistic perspective	K. Haude (with R. Keller)	Cologne/Paris (DFG)
IP 04: Hierarchical ranking and argument encoding in three-participant clauses	A. Siewierska (with E. van Lier)	Lancaster (AHRC)
IP 05: Mapudungun and Blackfoot: inverse alignment with special attention on three-participant clauses	F. Zúñiga (with A. Herdeg)	Zurich/Bern (SNF)

The individual projects functioned largely independently, but we had regular project meetings, sometimes with additional invited speakers, and different constellations of project members organized workshops for a broader public and with refereed abstracts. In this way, we became a closely-knit team. Anna Siewierska's death in summer 2011 was a terrible blow, however, and it took us time to gain back our previous joyful spirit.

The most tangible collective outcome of this project was a special issue of *Linguistics* that I co-edited with Alena Witzlack-Makarevich; see Table 2 for the contents of this special issue.

Table 2. Table of Contents of *Linguistics* 54(3)

2.1.3 The encoding of three-participant events crosslinguistically (DobeS)

Another typologically-oriented project I participated in was the DobeS project *Cross-linguistic patterns in the encoding of three-participant events* directed by Anna Margetts (Monash University, Melbourne), whose German counterpart was the Linguistics Department of Cologne and of which Nikolaus Himmelmann and I were co-applicants. The idea of this project came in 2011, when I read an article by Margetts & Austin (2007) that discusses the various ways in which different languages express events with three participants, such as 'give'. Unlike other studies on the topic, the paper is not restricted to languages that have dedicated "ditransitive" constructions for expressing three-participant events. This seemed to me the ideal approach, since many Native American languages do not have dedicated ditransitive constructions. I contacted Anna Margetts, whom I knew from the DobeS programme, and together we decided to submit a project for the last round of the DobeS programme, which funded projects dealing with the scientific exploration of existing DobeS corpora.

The participating languages and language experts (with their affiliation at that time) that formed the project team are listed in Table 3. Each of the eight project members had compiled a documentary corpus in a previous DobeS project.

Table 3. Languages and researchers participating in the "3Participant" project

Language (family/region)	Participating researcher
Beaver (Athabascan)	D. Jung (Zurich University)
Bora (isolate, Amazonia)	F. Seifart (University of Amsterdam)
Movima (isolate, Amazonia)	K. Haude (SeDyL-CNRS)
Saliba-Logea (Oceania)	A. Margetts (Monash University, Melbourne)
Savosavo (Papua)	C. Wegener (University of Bielefeld)
Totoli (Austronesian)	S. Riesberg (University of Cologne)
Vera'a (Austronesian)	S. Schnell (Melbourne University)
Waima'a (Austronesian)	J. Hayek (Monash University, Melbourne)

The principal goal of this project, which lasted over four years (2013-2017), was to identify the expression of three-participant events independently of whether a language has a ditransitive construction (with three core arguments) or not. We started out from the hypothesis proposed by Margetts & Austin (2007) that roughly 12 types of three-participant events can be idenified semantically and that there are approximately 7 main strategies that languages employ to express these. The list of the main "target events", i.e. of the events whose expression we annotated in our corpora, is given in , and the abbreviated list of the major encoding strategies is given in Table 5. At each of our yearly project meetings, both lists were discussed and adjusted according to our findings.

Table 4. The main 3-participant "target events"

Set A	Agent causes recipient to receive theme
Set B	Agent causes theme to move to/be located at location
Set C	Agent acts to communicate information to recipient
Set D	Agent/recipient takes possession of theme from source
Set E	Agent removes theme from R-type possession or location
Set F	Agent intends to cause recipient to receive theme (and it is the activity which *creates* or *makes available* the Theme for the Recipient)
Set G	Agent uses non-body part instrument to impact on or make change to patient
Set H	Agent uses body part instrument to impact on or make change to patient
Set I	Conditions of satisfaction imply that agent causes recipient to receive theme
Set J	Agent acts to cause recipient to receive theme at some future point in time

Set K	Agent allows recipient to receive theme: verbs of permission and enablement
Set L	Agent causes recipient not to receive or have access to theme: verbs of refusal

Table 5. “Encoding strategies” of three-participant-event expressions

1. Three-place predicate strategy 1a) Direct argument strategy: All three arguments are expressed as direct arguments of the verb (which does not carry valence increasing morphology). 1b) Causative strategy: i)The verb root is restricted to two arguments, a third argument is added by a causative affix. ii) The verb root is restricted to one argument, a transitive stem is derived by an applicative, a third argument is added by a causative affix. 1c) Applicative strategy i) The verb root is restricted to two arguments, a third argument is added by a applicative affix. ii) The verb root is restricted to one argument, a transitive stem is derived by a causative affix, a third argument is added by an applicative affix.
2. Oblique and adjunct strategies 2a) R-type obliques and adjuncts i) “Oblique Applicative”: The verb includes an applicative-like marker:2a-APPL ii) “Oblique Causative” : The verb includes a causative-like marker: 2a-CAUS 2b) T-type obliques and adjuncts i) “Oblique Applicative” : The verb includes an applicative-like marker: 2b-APPL ii) “Oblique Causative” : The verb includes a causative-like marker: 2b-CAUS
3. Serial verb strategy 3a) R-type serialized P 3b) T-type serialized P
4. Incorporation strategy 4a) The incorporated noun is a syntactic argument of the verb (= subtype of direct argument strategy).

i) R-type incorporated P ii) T-type incorporated P 4b) The incorporated noun is not a syntactic argument of the verb. i) R-type incorporated P ii) T-type incorporated P
5. Adnominal strategy 5a) Possessive strategy: The R-type participant is expressed as the possessor of the theme. 5b) Proprietive strategy: The T-type participant is expressed as the dependent of the agent.
6. Directional strategy
7. Absorption strategy 7a) Direct lexicalization 7b) Zero derivation 7c) Denominal derivation 7d) Absorbed classifiers or object markers 7e) Participant-based event classification

A special feature of the project was that we did not just look at lexical items and their meanings. The explicit aim was to do corpus-based research, and we examined our corpora in detail in order to find out how three-participant events are expressed in spontaneous discourse, independently of the lexical expression used. A verb with the meaning 'take (something somewhere)', for instance, was only counted if the Goal was either overtly expressed or implied through the context; by contrast, a verb meaning 'give' was not counted if it did not encode a three-participant event (as e.g. in the English expression *Give me a hand* for *Help me*).

The "target events" to code for were characterized as neutrally as possible to allow for their identification in different languages; for instance, English *give* usually expresses an event of Set A, in which an agent causes a recipient to receive a theme, and the same is true for English *bring*. However, *bring* can also express a situation of Set B, in which an agent causes a theme to move to or to be located at a location; this depends on the context.

Figure 2 illustrates how we annotated our data, which most of us did in Toolbox. Under the tier for the free translation, we created is a tier "\3Pant" in which we systematically entered the event type (in Figure 2: "A"), a code for the verb in parentheses (here: "give-kayaLe"), the

coding strategy (here: “2b”, indicating that the Theme is encoded as an adjunct), and indicated if one argument is unexpressed (here: “-T”, meaning that the theme argument is not overtly expressed in the clause). An additional tier, “3Pant comment”, allowed for any commentary that would help understand the annotation. Here, I noted that the presence of a theme participant is understood from the context.

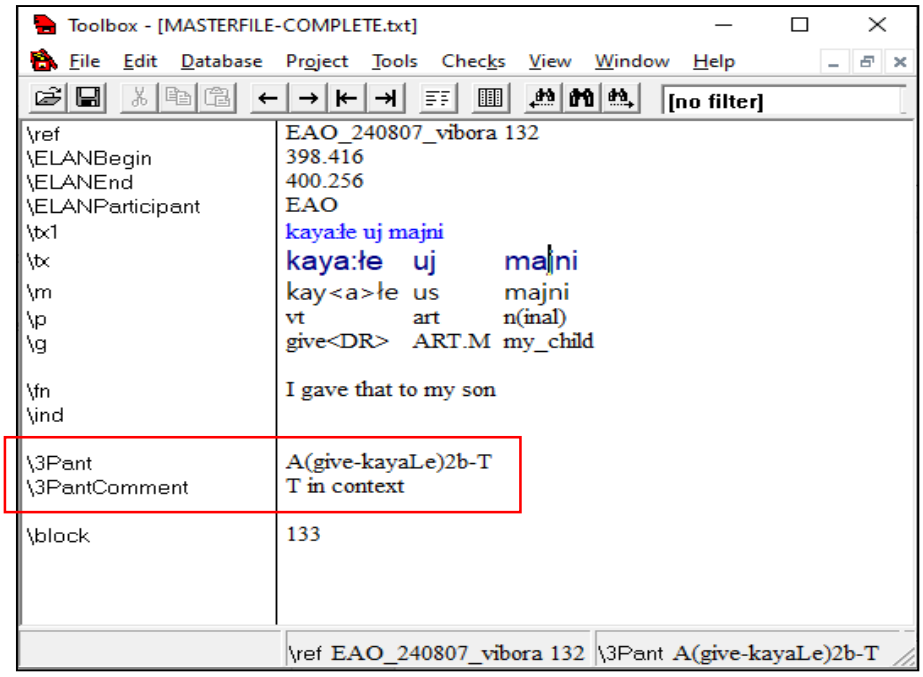


Figure 2. Example of annotation of “3Pant” features in the Movima Toolbox corpus

The languages represented in our project show different patterns regarding the encoding of three-participant events. Most importantly, the languages of the Pacific, Saliba-Logea, Vera’a, and Savosavo show productive verb serialization, whereas the Amerindian languages in the project, Beaver, Bora and Movima, show abundant use of applicative morphology and incorporation, in line with the well-known polysynthetic character of many Amerindian languages.

While our annotations showed many potentially interesting results, which were analyzed statistically by Andrew Margetts (Monash University), this project suffered from the fact that none of the main participants, especially the PI (who had not been allocated a teaching relief), could work full-time on its coordination and on published outcomes. We then limited the topic to one particular type of three-participant event, which does not seem to have been examined in detail before: “caused accompanied motion in a direction” (expressed in English by the verb

bring, but expressed in very different ways in other languages). On this topic, we submitted a co-authored journal paper (Margetts et al. To appear).

2.2 Projects in France

2.2.1 *Ergativité* (*Fédération TUL – Typologie et Universaux Linguistiques*)

After my first two-week visit to the CELIA in April 2007, the director of the CELIA, Francesc Queixalós, invited me to become a regular member of the programme *Ergativité* financed by the *Fédération Typologie et Universaux Linguistiques*. The monthly meetings with experts on ergativity and ergative languages (Gilles Authier, Denis Creissels, Guillaume Jacques, Aurore Monod-Becquelin, Francesc Queixalós, Valentina Vapnarsky, and others, plus occasional invited guests) were extremely inspiring for my research. My eagerness to understand and participate in the discussions was an enormous motivation to improve my French, which, up to then, I only remembered from some years at highschool.

Together with Gilles Authier I took over the task of compilating the volume that combined the results of the members of this project (Authier & Haude 2012). The authors and their contributions are listed in Table 6.

Table 6. Contents of Authier & Haude (2012a)

***Ergativity, Valency and Voice*, ed. by Gilles Authier and Katharina Haude**
Introduction (Gilles Authier and Katharina Haude)
Ergativity and voice in Mayan languages: a functional-typological approach (Colette Grinevald and Marc Peake)
Ergativity and the passive in three Mayan languages (Valentina Vapnarsky, Cédric Becquey, and Aurore Monod Becquelin)
A tale of two passives in Cavineña (Antoine Guillaume)
The detransitive voice in Kryz (Gilles Authier)
Laz middle voice (René Lacroix)
Argument demotion in Japhug Rgyalrong (Guillaume Jacques)
The Katukina-Kanamari antipassive (Francesc Queixalós)
Undergoer orientation in Movima (Katharina Haude)
Case patterns and verb classes in Trumai (Aurore Monod Becquelin and Cédric Becquey)
Ergativity in the Adyghe system of valency-changing derivations (Alexander Letuchiy)
The evolution of transitive verbs in Basque and emergence of dative-marked patients (Céline Mounole)

2.2.2 *Saillance* and *Référence et repérage* (SeDyL)

In January 2010 I started my position as a *chercheur 1ère classe* at the CNRS. As such, I became a member of the lab *SeDyL* (Structure et Dynamique des Langues), which was newly created from a fusion of the CNRS-IRD lab *CELIA* (Centre d'Études de Langues Indigènes d'Amérique) and the *Circle Linguistique de l'INALCO* (CLI).

After joining the SeDyL I was asked to coordinate, together with Annie Montaut, the publication of a special issue of *Faits de Langues* that would bring together contributions from the CELIA and the CLI working groups on the topic of *Saillance* (see 5.2). In accepting, I took the opportunity to also include contributions from members of the project *Referential Hierarchies in Morphosyntax* (see 2.1.2), which dealt with the same phenomenon. The twelve papers compiled in the special issue (Haude & Montaut 2012) are listed in Table 7.

Table 7. Table of contents of *Faits de Langues (39 :1)* (Special issue on *La saillance*)

Annie Montaut & Katharina Haude	Présentation générale
Frédéric Landragin	La saillance – questions méthodologiques autour d'une notion multifactorielle
Spike Gildea	The referential hierarchy and attention
Christine Bonnot	Deixis et saillance d'événement – le cas des énoncés à accent non final en russe contemporain
Anaïd Donabédian	Evidentiel et progressif – quel statut grammatical pour la saillance prédicative ?
Annie Montaut	Saillance et antisaillance en hindi – du constituant de l'énoncé à l'énoncé saisi dans son ensemble
Francesc Queixalós	Saillance en sikuani
Jean-Michel Hoppan	Questions de saillance et épigraphie maya
Alexandru Mardale	Le trait [+ personne] comme facteur de saillance en roumain
Anna Siewierska & Eva van Lier	Ditransitive constructions with two human non- agentive arguments
Odile Renault-Lescure	Personnes et saillance en kali'na
Katharina Haude	Saillance inhérente et saillance discursive en movima
Fernando Zúñiga	What do we (not) know about Blackfoot inversion?

I then became a member of the research strand of the SeDyL *L'énoncé et ses composantes: Rréférence et repérage* directed by Chr. Bonnot and O. Duvallon, which analyzed the internal structure of sentences in different languages. I enjoyed the discussions with colleagues who had a different linguistic background than myself, mostly based on Culioli's *Théorie des*

opérations énonciatives. In particular, I appreciated the very valuable feedback I got regarding pragmatically marked constructions in Movima, on which I report in Section 6.1.3.

2.2.3 The Typology and Corpus Annotation of Information Structure and Grammatical Relations (LABEX EFL, GD1)

Shortly after my arrival at the CNRS, the Excellence Laboratory LABEX EFL (*Empirical Foundations of Language*) was created. Within its research strand 3, *Typology and dynamics of linguistic systems*, I was involved in the proposal for the collaborative project (*opération de recherche*) GD1, *The Typology and Corpus Annotation of Information Structure and Grammatical Relations*, of which I later took over the coordination together with Martine Vanhove. The 8 permanent members of this project came from the three research labs in Villejuif that are associated with the LABEX: A. Mettouchi, T. Nikitina, S. Robert, and M. Vanhove from the LLACAN; C. Chamoreau, E. Palancar, and myself, from SeDyL; and E. Adamou from the LaCiTO). First, at the monthly meetings of the group, we discussed the grammatical relations and information structure in our respective languages of research, and we organized the second international conference *Information Structure in Spoken Language Corpora 2* (ISSLAC2) in 2015. In the second phase, we concentrated on one particular topic, clefts, on which we organized an international workshop (Clefts and related focus constructions, 2018). This very fruitful collaboration had particular influence on my work described in Section 6.

Throughout the project, and particularly in the final phase, we discussed possible ways to annotate linguistic corpora for information structure. Some decided to annotate their corpora for specific pre-defined categories (such as “focus”, “topic”, “afterthought”), while others, including myself, preferred to annotate for the different structures found in the data in order to identify their information-structural meaning later.

In 2018, I was elected as the co-director of the research strand (“Axe”) 3 of the LABEX EFL, and as the main responsible of this research strand in the second phase of the LABEX (2020–2025).

2.2.4 CorTypo: Creating corpora for cross-linguistic analysis (ANR)

The ANR project *Designing Spoken Corpora for Cross-Linguistic Research* (*CorTypo*, 2013-2017, PI Amina Mettouchi) had the aim of creating comparable annotations for language corpora of different languages. The main task for us was to identify the ways ("predications") in which different communicative functions (e.g. reference) are expressed in our languages of research, and to identify these on the basis of with 30 minutes of spoken discourse from our corpora annotated according to a fixed scheme. The theoretical assumption on which this project was based, and which was particularly advocated by Z. Frajzyngier, was that typology should be carried out from a "non-aprioristic" perspective, i.e. without assuming that any given function would be coded in the grammars all languages. In other words, each construction found in a language can be assumed to have a particular function of its own, even if it can be difficult to identify it. I wish to point out here that this view was also the one I had been brought up with. My university teacher H.-J. Sasse formulated it very nicely: "One of the fundamental heuristic strategies in linguistic analysis should be the attempt to find a uniform function or meaning for each formal linguistic phenomenon, i.e., not to assume *a priori* that there are irregular mapping relations between form and function (homonymy, etc.), but to proceed from the assumption that there is, *in principle*, a 1:1 relation, each form having one and only one basic function, the subfunction being determined by and explainable in terms of the environment. Assuming of homonymy is acceptable only as the result of the inability to discover a uniform meaning, i.e., as a last resort, so to speak, but never as an acceptable working principle" (Sasse 1991: 94).

Thus, the CorTypo project was quite in line with my own view on typological and descriptive research. I found it problematic, however, that the goal was to identify "all" predications in the language. I still do not know how this should be possible. A lot of my struggle when studying my data is due to the difficulty to identify a structure as a construction of its own, as a single construction, or as representing more than one construction (which may be distinguished by prosody, for instance). The differences between two constructions can be extremely subtle, and their identification can be a matter of debate. An example of a late discovery of a "new" construction is given in Section 6.2.4, where I describe a cleft that is only subtly distinct from a non-cleft focus construction in Movima (Haude 2019c).

3. The basis of my research: the Movima language

3.1 Background on Movima and previous studies

My research is based on first-hand data from Movima, an endangered linguistic isolate of lowland Bolivia. Movima is spoken in Santa Ana del Yacuma, the capital of the province Yacuma in the Bolivian Beni department, which is today one of the main cattle-breeding areas of Bolivia (see Figure 3). The region consists of natural savannah and swamps, interspersed with scattered forest islands and numerous smaller rivers and streams. The town can be reached from the department capital Trinidad either with a small Cessna plane in a 40min flight or, during the dry season, by car, which usually takes between 4 and 7 hours depending on the conditions of the road and the vehicle.

Santa Ana del Yacuma was founded by the Jesuits at the beginning of the 18th century, and it is the place where most Movima speakers live today. The town's population is split up into a large upper class of white land-owners, a growing middle class (teachers, nurses, self-employed) of mixed origin, and a lower class, consisting almost exclusively of Movima Indians who usually work as domestic employees, washer women or cowherds. In addition, there is a large number of Quechua- and Aymara-speaking merchants who have migrated from the highlands.

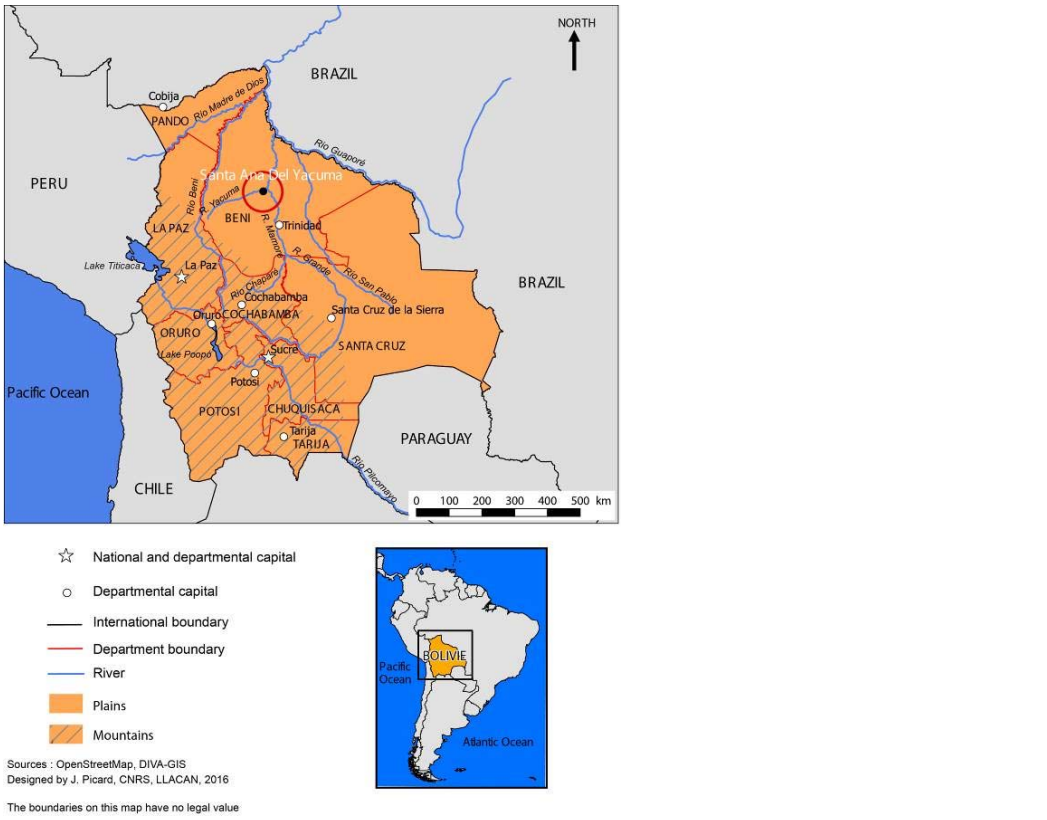


Figure 3. The location of Santa Ana del Yacuma, the capital of the Movima language

The region where Movima is spoken used to be of extremely high linguistic diversity, as shown by the map of languages in Figure 4. Two large South-American language families are represented here, Arawak (Baure) and Tupi (Guarayo, Sirionó, Guaraní), while other languages belong to regionally limited and often very small families (e.g., Mosetenan has only two members, Yuracarean is most probably a single language with different variaties rather than a family). In addition, there are several isolates, most of which are concentrated in the area of the Mamoré river, and one of which is Movima. The neighbouring isolates Canichana and Cayuvava are now extinct, according to the most recent edition of the Ethnologue (Eberhard, Simons & Fennig 2019).

In comparison with many other indigenous languages of the Amazon area, the Movima speaker community is still relatively large, with probably around 200 fluent speakers left (Movima language activists counted 500 speakers in 2006, but this number also included those who did not speak the language fluently). However, the language has not been transmitted to children for at least two generations, and so, most fluent Movima speakers are over sixty years old. This has probably been due to spread of Spanish in rural areas through the Bolivian

educational reform in 1952, as well as to the low prestige of indigenous languages. Given this problematic situation alongside with the relatively large number of remaining speakers, Movima was an ideal case for carrying out a language documentation project.

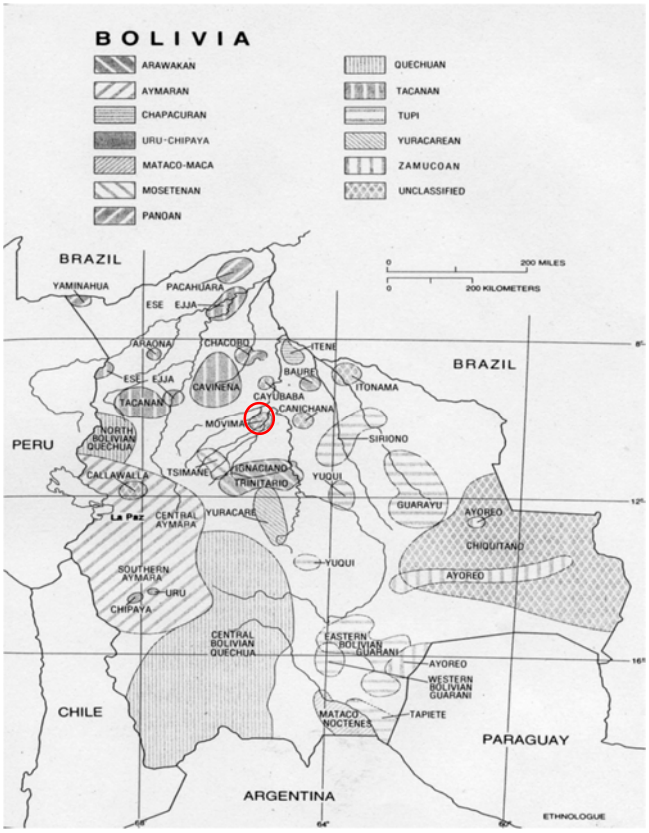
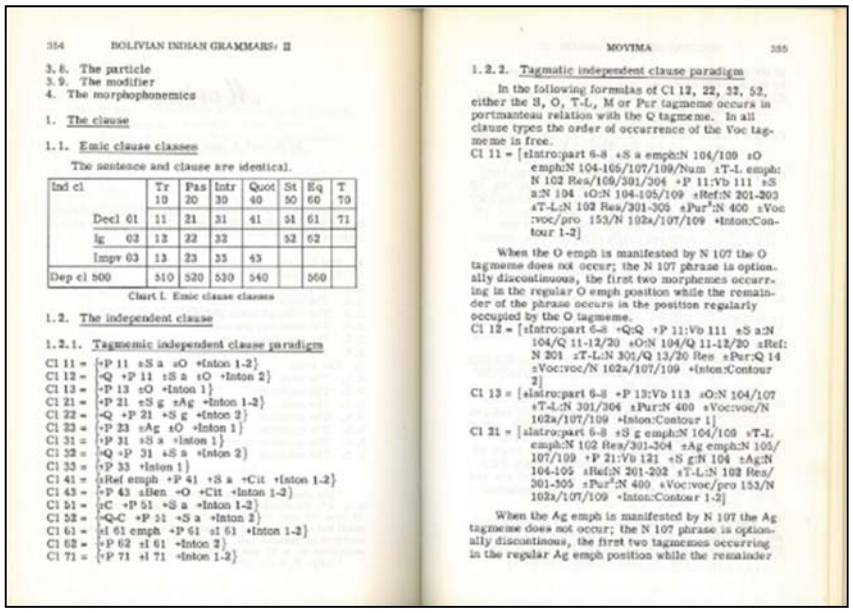


Figure 4. Languages of Bolivia (Ethnologue 1996)

The first materials of the Movima language are lists of words and phrases ((Hervás y Panduro 1787; d'Orbigny 1839; Heath 1883; Cardús 1886; Chamberlain 1910; Créqui-Montfort & Rivet 1914; del Castillo 1929). Of these older lists, the one produced by the missionary Cardús (1886) is the most reproduced one in the literature.

In the 1960s, Movima was studied extensively by the couple Robert and Judith Judy, of the Summer Institute of Linguistics. Their publications include a word list with a grammatical sketch (Judy & Judy 1962a), an outline of the phoneme inventory (Judy & Judy 1962b), two articles on specific topics (Judith E. Judy 1965; Robert A. Judy 1965), and a grammar sketch (Judy & Judy 1967). Unfortunately, these publications are extremely concise, and the grammar sketch, like most of the SIL publications of the 1960s, is based on the rather inaccessible tagmemic framework. Figure 5 gives an example of this type of linguistic description, which

is not very useful as a basis for further investigation. Therefore I had to start nearly from scratch when I began my research on this language.



354 BOLIVIAN INDIAN GRAMMARS: II

3.8. The particle
3.9. The modifier
4. The morphophonemics

1. The clause

1.1. Emic clause classes

The sentence and clause are identical.

Ind cl		Tr 10	Pas 20	Intr 30	Quot 40	St 50	Eq 60	T 70
	Decl 01	11	21	31	41	51	61	71
	Ig 02	12	22	32		52	62	
	Impv 03	13	23	33	43			
Dep cl 500		510	520	530	540		560	

Chart I. Emic clause classes

1.2. The independent clause

1.2.1. Tagmemic independent clause paradigm

Cl 11 = [+P 11 ±S a ±O +Inton 1-2]
Cl 12 = [+Q +P 11 ±S a ±O +Inton 2]
Cl 13 = [+P 13 ±O +Inton 1]
Cl 21 = [+P 21 ±S g ±Ag +Inton 1-2]
Cl 22 = [+Q +P 21 +S g +Inton 2]
Cl 23 = [+P 23 ±Ag ±O +Inton 1]
Cl 31 = [+P 31 ±S a +Inton 1]
Cl 32 = [+Q +P 31 ±S a +Inton 2]
Cl 33 = [+P 33 +Inton 1]
Cl 41 = [±Ref emph +P 41 +S a +Cit +Inton 1-2]
Cl 43 = [+P 43 ±Ben +O +Cit +Inton 1-2]
Cl 51 = [±C +P 51 +S a +Inton 1-2]
Cl 52 = [+Q-C +P 51 +S a +Inton 2]
Cl 61 = [±l 61 emph +P 61 ±l 61 +Inton 1-2]
Cl 62 = [+P 62 ±l 61 +Inton 2]
Cl 71 = [+P 71 ±l 71 +Inton 1-2]

MOVIMA 355

1.2.2. Tagmatic independent clause paradigm

In the following formulas of Cl 12, 22, 32, 52, either the S, O, T-L, M or Pur tagmeme occurs in portmanteau relation with the Q tagmeme. In all clause types the order of occurrence of the Voc tagmeme is free.

Cl 11 = [±Intro:part 6-8 ±S a emph:N 104/109 ±O emph:N 104-105/107/109/Num ±T-L emph: N 102 Res/108/301/304 +P 11:Vb 111 ±S a:N 104 ±O:N 104-105/109 ±Ref:N 201-203 ±T-L:N 102 Res/301-305 ±Pur[1]:N 400 ±Voc:voc/pro 153/N 102a/107/109 +Inton:Contour 1-2]

When the O emph is manifested by N 107 the O tagmeme does not occur; the N 107 phrase is optionally discontinuous, the first two morphemes occurring in the regular O emph position while the remainder of the phrase occurs in the position regularly occupied by the O tagmeme.

Cl 12 = [±Intro:part 6-8 +Q:Q +P 11:Vb 111 ±S a:N 104/Q 11-12/20 ±O:N 104/Q 11-12/20 ±Ref: N 201 ±T-L:N 301/Q 13/20 Res ±Pur:Q 14 ±Voc:voc/N 102a/107/109 +Inton:Contour 2]

Cl 13 = [±Intro:part 6-8 +P 13:Vb 113 ±O:N 104/107 ±T-L:N 301/304 ±Pur:N 400 ±Voc:voc/N 102a/107/109 +Inton:Contour 1]

Cl 21 = [±Intro:part 6-8 ±S g emph:N 104/109 ±T-L emph:N 102 Res/301-304 ±Ag emph:N 105/107/109 +P 21:Vb 121 +S g:N 104 ±Ag:N 104-105 ±Ref:N 201-202 ±T-L:N 102 Res/301-305 ±Pur[1]:N 400 ±Voc:voc/pro 153/N 102a/107/109 +Inton:Contour 1-2]

When the Ag emph is manifested by N 107 the Ag tagmeme does not occur; the N 107 phrase is optionally discontinous, the first two tagmemes occurring in the regular Ag emph position while the remainder

Figure 5. Example pages from Judy & Judy (1967)

3.2 Fieldwork

I collected all the data on which my research is based myself, through elicitation and text analysis with native speakers of Movima. For this purpose I carried out ten field trips to Santa Ana del Yacuma, of a total duration of more or less 15 months. My field trips are listed in

Table 8, as well as the projects by which they were financed and their main purposes.

Table 8. Field trips

Year	Duration (appr.)	Objective/activities	Project
2001	12 weeks	Acquaintance, elicitation, text annotation	Dissertation
2002	10 weeks	elicitation, text annotation	Dissertation
2003	3 weeks	elicitation, text annotation	Dissertation
2004	7 weeks	elicitation, text annotation	Dissertation
2006	7 weeks	video recordings	DobeS
2007	6 weeks	video recordings	DobeS
2008	5 weeks	video recordings	DobeS
2009	4 weeks	annotation	DobeS/RHIM
2010	3 weeks	correction of annotations	RHIM
2012	2 weeks	correction of annotations	RHIM
TOTAL	59 weeks		

During these field trips I collected approximately 10,000 elicited examples and about 180 text recordings. For the grammar, I relied heavily on elicitation, i.e. the question-answer procedure with native speakers that helps understand a particular grammatical pattern. In my fieldwork for the DobeS project I concentrated more on the collection of spontaneous discourse and recurred to elicitation only when I was working with a speaker on the translation of a text. My data analysis is now mostly based on these spontaneous text data.

The corpus collected and annotated for the DobeS project contains 75 texts from 32 speakers (12 male, 20 female), resulting in almost 90,000 transcribed words. Most of the texts are translated into Spanish; a small part (4,000 words) is translated into English. The texts together constitute 21:37h of audio recordings, of which 17:36h are video-recorded as well (see Table 9, which lists the recordings according to the acronyms of the speakers).

Table 9. List of annotated texts from the DobeS project

Text label	annotation units	words transcribed	words glossed (ca)	words SP	words ENG	audio (hh:)mm:ss	video (hh:)mm:ss
AMY_180806	300	1400	0	0	0	19:45	
ATL_230806-tb-3p	447	2059	400	2059	1000	27:00	27:00
BHA_280706_1-tb	22	110	0	110	110	01:06	01:06
BHA_280706_2-tb	7	17	0	17	17	02:55:00	02:55:00
Cabildo 2006-07-17 PMP	56	231	0	0	0	03:28	
Cabildo_020907	122	0	0	0		12:00	
CCT 120907-1	139	1002	0	1002	0	08:45	08:45
CCT_120907_2	259	1729	0	1729	0	15:00	15:00
CCT_120907_3	67	248	0	248	0	03:33	03:33
CCT_LC_NCO_020907	229	1125	0	1125	0	12:20	12:20
CVM_GCM_020906-1-tb	417	1479	0	800	0	18:18	18:18
EAO_120906_1-tb	307	1608	0	1608	0	15:08	15:08
EAO_120906_2-tb	34	214	0	214	0	02:40	02:40
EAO_120906_3	282	1828	0	1828	0	17:08	17:08
EAO_240807-vibora	226	1235	1235	1235	1235	11:34	
EAO-tigreyperro_150808	178	1056	0	0	1056	13:00	13:00
EGA_BVO_AAO_HRR_180706_1	261	963	0	0	0	15:48	15:48
EGA_MGA_DMY_060906_1	143	506	0	0	0	07:43	07:43
ERM_090706_1	32	134	134	134	134	01:57	01:57
ERM_090706_2	32	178	0	178	178	01:37	01:37
ERM_090706_3	30	150	0	150	0	01:28	01:28

ERM_140806_1	1109	5300	0	5300	0	58:00	58:00
ERM_140806_2	575	2473	0	2473	0	35:55	35:55
ERM_150806	769	3185	0	3185	0	48:00	
GBM_110910_1-1	307	1100	0	1100	0	18:00	18:00
GCM_290806_1	269	1051	0	1051		12:26	12:26
GCM_290806_2	211	754	0	754		09:40	09:40
GCM_290806_3	387	1683	0	1683		19:47	19:47
GCM_290806_4	182	645	0	645		07:45	07:45
GCM_290806_5	436	1817	0	1817		20:24	20:24
HRR_081009_isbijaw	285	1659	0	600		16:00	
HRR_120808_gloss_tb	802	4182	1800	4182		48:41	48:41
HRR_170510-1	30	158	0	0		01:42	01:42
HRR_170510-2	19	86	0	86		01:00	01:00
HRR_200510-1	34	198	0	198		01:55	01:55
HRR_2009_tape1_A	669	3447	16	3447		31:25	
HRR_2009_tape1_B	363	2102	16	2101		22:09	
HRR_2009_tape2_A	107	668	480	320		08:37	
JGD_130907_10_tortugas	174	702	0	702		05:50	
JGD_130907-01_El hombre y el gallo	209	938	0	938		10:40	
JGD_130907-06_El pueblo que se hundio	334	1319	1319	1319	1319	15:08	
JGD_130907-08_El lago seco	68	300	300	300		03:47	
JGD_130907-09_El pescador	219	988	988	988		12:13	
JGD_130907-13_El gato diablo	294	1412	0	1412		17:48	
JGD_160808_fundacion-02	516	2418	0	2418		30:30	30:30
JGD_160808-Fundacion-01_tb	843	3897	0	3897		49:41	49:41
JGD_HRR_080908_2	310	1500	0	1500		22:00	22:00
JMH_160806_1	313	1010	0	1010		16:30	16:30
JMH_160806_2	288	1023	0	1023		14:37	14:37
JZH-080807	223	1250	0	1250		20:00	20:00
LCC_010609_1	354	1757	0	1757		18:00	18:00
LTC_020906_1-POY	253	971	0	971		13:30	13:30
LTC_020906_2-POY	72	258	0	258		04:00	04:00
LTC_020906_3-POY	232	964	0	964		10:40	10:40
LTC_020906_4	312	1157	0	1157		14:50	14:50
LTC_020906_5	425	1687	0	1687		19:00	19:00
LYO_250808	385	1866	0	1866		25:15	25:15
LYO_250808_2	330	1705	0	1705		22:00	22:00
MCA_060906_1	234	862	0	0		12:21	12:21
MCA_060906_2	376	1739	0	1739		20:43	20:43
MCA_280806_2	376	2252	12	2252		22:07	22:07
MCA_280806-1	447	2059	0	0		22:18	22:18
MCC_250806-tb	219	979	0	0		11:46	11:46
MRH_220908_mov1_1	291	1946	0	1946		21:42	21:42
NAO-FSG_300706_1	598	2106	0	2106		42:00	42:00
NCG_240806_1	57	285		285		07:15	07:15
NCG_240806_2	61	299	0	299		07:02	07:02

NCG_240806_3	42	211	0	211		08:22	08:22
NCO_280706.eaf	50	163	0	163	163	01:45	01:45
NGS_210806_1	77	351	0	0		08:05	08:05
PMP_170706	48	194	0	0		03:00	03:00
PMP_HRR_etal_210908	277	1376	0	1376		12:20	12:20
SGC_020806	33	189	0	0		02:10	
SUM	**19,362**	**89,913**	**6,700**	**77,679**	**3,977**	**21:36:51**	**17:36:17**

The data corpus that forms the basis for my research is the combination of the above text corpus with the texts I had collected (in audio only) in the previous years for my doctoral project. This earlier corpus consists of 101 shorter texts from 12 speakers, resulting in 44,700 transcribed words, and has a total length of 8:38h. This part of the corpus is fully translated into Spanish and partly into English. Taken together, the entire corpus now consists of 30h15min of audio-recorded (and partly video-recorded) text and 130,000 transcribed words. In comparison with corpora on well-studied, national languages, this may seem little. However, for an unwritten language on which no data existed before 2001, such a corpus is rather large.

I transcribed all these data myself, and did most the translations either alone or in collaboration with a speaker (usually not the one who had produced the text). In the first two years of my dissertation project I used pen and paper for this; later, I transcribed with Microsoft Word and then converted the data into Toolbox file; finally, as required by the DobeS programme, I used ELAN, whose development had reached maturity by then so the programme had become stable and easy to use. I soon knew the language well enough to do also most of the translations myself, which I then checked with native speakers only when I was not sure.

I basically transcribed the texts in the orthography established for Movima, which is very close to the phonemic representation (Haude 2006: 65–66). Transcribing Movima is not difficult for me, since there are few phonetic intricacies and hardly any morphophonological fusion that would lead to a loss of audible segments; furthermore, Movima speakers tend to speak slowly and clearly. Transcription is also greatly facilitated by ELAN. ELAN is a specifically designed annotation programme developed at the MPI for Psycholinguistics in Nijmegen (https://tla.mpi.nl/tools/tla-tools/elan/), which links audio and video files to a surface for time-aligned transcriptions. In this way, one can select a segment of the recording, and type while listening to the segment as many times as necessary. Later, one can add as many lines as necessary for the annotation. Figure 6 shows a screenshot of an ELAN file with the

transcription, the English translation, glosses, and other annotations that I added later for specific purposes.

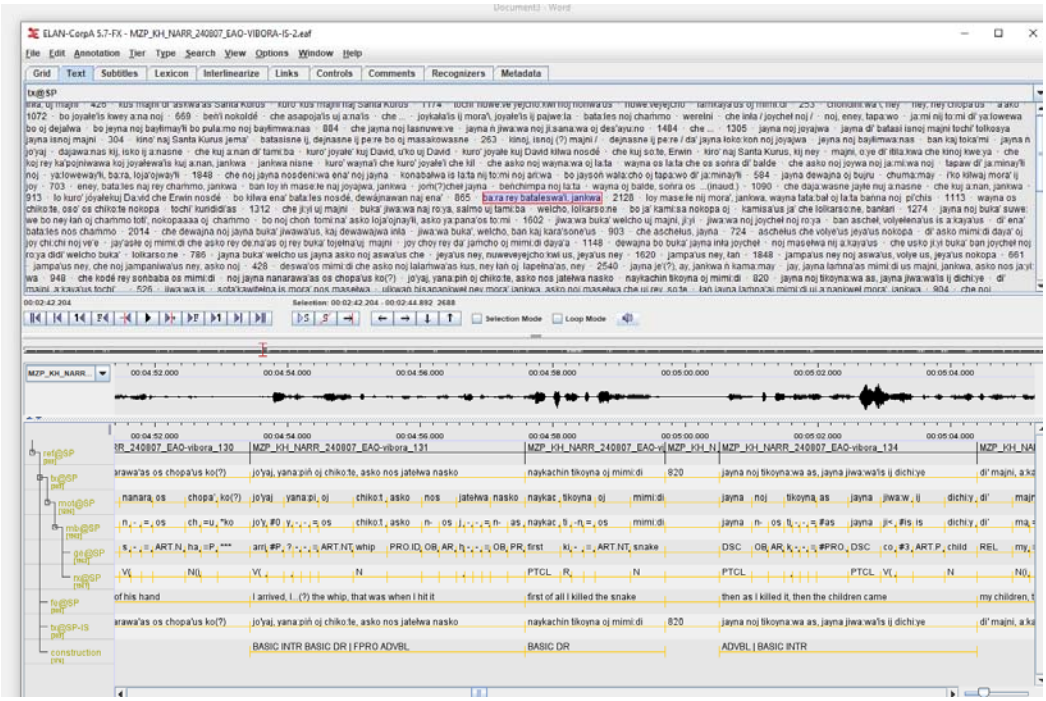


Figure 6. Example of an annotation in ELAN (screenshot)

But even with some fluency at transcribing and with a dedicated tool such as ELAN, linguistic annotation is an extremely time-consuming task. A 5-minute-segment of spoken Movima means for me at least one hour for transcription, one hour for translation, and one hour for glossing. Hence, the transcription of my 30 hours of recording alone has taken me over 360 hours, which corresponds to about 45 8-hour work days; translation and glossing, where it has been done, can be added to this at a similar rate. This explains why only a small part of the corpus is glossed (i.e. has interlinear morpheme-by-morpheme translation), as can be seen in Table 9 above. Given the amount of time that each annotation step takes, one has to make choices. My choice was to transcribe as much data as possible rather than invest the time in glossing a smaller amount of data.

The reason why I had to annotate the corpus all by myself was that I never managed to find an assistant in the field who would be both able and inclined to do this work. The speakers I worked with were all elderly and had usually no practice in writing. The younger semi-speakers and language activists I asked, who read and write and can in some cases also use a computer, and who I would of course had paid for their work, did not seem interested in doing

transcription; they preferred to write texts of their own. This was a bit disappointing for me; however, at the same time this shows how much the Movima speakers who are interested in the revitalization of their language are independent from money-driven decisions. And furthermore, there is no better way to getting familiar with one's data than by transcribing them oneself.

As required by the DobeS project, the data were integrated into the DobeS archive at the Max Planck Institute Nijmegen (now The Language Archive, https://tla.mpi.nl/). Many of the Movima data are openly accessible and can be freely downloaded, most are accessible after registration, and only few documents are closed to the public because they contain too personal information. In principle it is possible for registered users to find in the archive the texts from which the examples in my publications are taken. The TLA in its present appearance is not very transparent, however, and it is difficult to gain useful information about the material contained in it. It is also unclear if it will be maintained for much longer, since the Science and Cognition Group (dir. S. Levinson), which hosted the archive at the Max-Planck-Institute, does not exist anymore. For this reason, the Linguistics Department of the University of Cologne has taken the initiative to host the data at the Language Archive Cologne (https://lac.uni-koeln.de/). At the time at which I am writing this text, one of their staff members, Gabriele Schwiertz, is moving the Movima data to the LAC; I only have to re-check the files and their access rights.

The Movima corpus furthermore corresponds to the criteria required by projects on larger corpus analysis. For instance, with this corpus I participate in the German-French project *Language Documentation Reference Corpora* (ANR-DFG; http://doreco.info/) directed by Manfred Krifka and Frank Seifart (both ZAS, Berlin), where corpora have to correspond to the following minimal requirements in order to be considered:

- a minimum of 10,000 transcribed words distributed over various recording sessions/annotation files
- translations into a major language
- time-alignment of transcription and translation with audio files at the level of sentences, paragraphs, utterances, or intonation units (i.e., "annotation units" in ELAN)
- audio is of reasonable quality (not too much overlapping speech or background noise)

To sum up, thanks to the friendliness and generosity of the Movima speakers and the institutions representing them, and to the DobeS project, whose principal goal it was to collect natural discourse data, I was able to build up a corpus that is indeed a goldmine for research on grammar and discourse of a formerly nearly undocumented linguistic isolate. The research described in this thesis has only been possible because of this corpus, and it will serve as a basis for future research.

3.3 Analyzing the corpus data

The grammar I had written as my doctoral thesis was mostly based on direct elicitation with speakers who helped me understand the structure of their language. The DobeS project that followed, by contrast, concentrated on the collection of natural discourse data. This combination of knowing the language well and disposing of a rich corpus made it possible to base my post-doctoral research almost entirely on natural discourse data.

While I have all my transcriptions in ELAN, where they are linked to the audio and video files, I usually analyze my data in Toolbox, a programme that had been developed for field linguists by the Summer Institute of Linguistics. Despite the constant improvement of ELAN, to which additional features for analysis have been added, Toolbox was the programme I had grown accustomed to, and for my purposes it still has a few advantages over ELAN.

In Toolbox, I have put all my annotations together into one "Master file", which consists of 26,221 annotation units (or "Toolbox records"; Figure 8 below is an example). I get often asked what my annotation units are based on. In fact, they are a hybrid of prosodic and syntactic unit. On the one hand, sentences are clearly identifiable in Movima, since constituent order is relatively strict. This facilitates a segmentation by sentence, and often, a sentence coincides with an intonation unit. On the other hand, syntactic units are often interrupted by pauses or filler elements. In particular, Movima speakers show an interesting tendency (worth of further research e.g. in a Master's thesis) to separate the article from the content word in a noun phrase by a pause or a filler element. Furthermore, an intonation unit – just as a sentence – can be quite long, and splitting it up is sometimes necessary in order to arrive at manageable chunks. Since I carried out my transcriptions during a time span of almost 20 years, and I usually tried to transcribe as quickly as possible, there is no real consistency in what an annotation unit in my corpus actually is. For me, it must be a manageable chunk that allows further meaningful annotation.

I usually analyze my data by applying filters, which allow me to look at only those annotation units that share a particular property. Since most of my data are not glossed, I use a

tagging system adapted to my research questions, which can then be addressed through filters. I started out by creating an additional annotation tier (\ind for "index"), in which I note any phenomenon that is relevant for a paticular research question. So, for instance, one allomorph of the morpheme marking the "direct" voice of transitive verbs (see Section 4) has the form /a/, which obviously is difficult to search for without a gloss. Here, I note 'DR' in the "\ind" tier. When I want to see a list of transitive constructions, I can create a filter in Toolbox that gives me all the results with this annotation. Figure 7 below shows the result, in which the 3065 entries (as indicated at the bottom of the window) with this annotation are given in a list together with their translations. Clicking on one of these lines gives me the corresponding full record with all the tiers it contains.

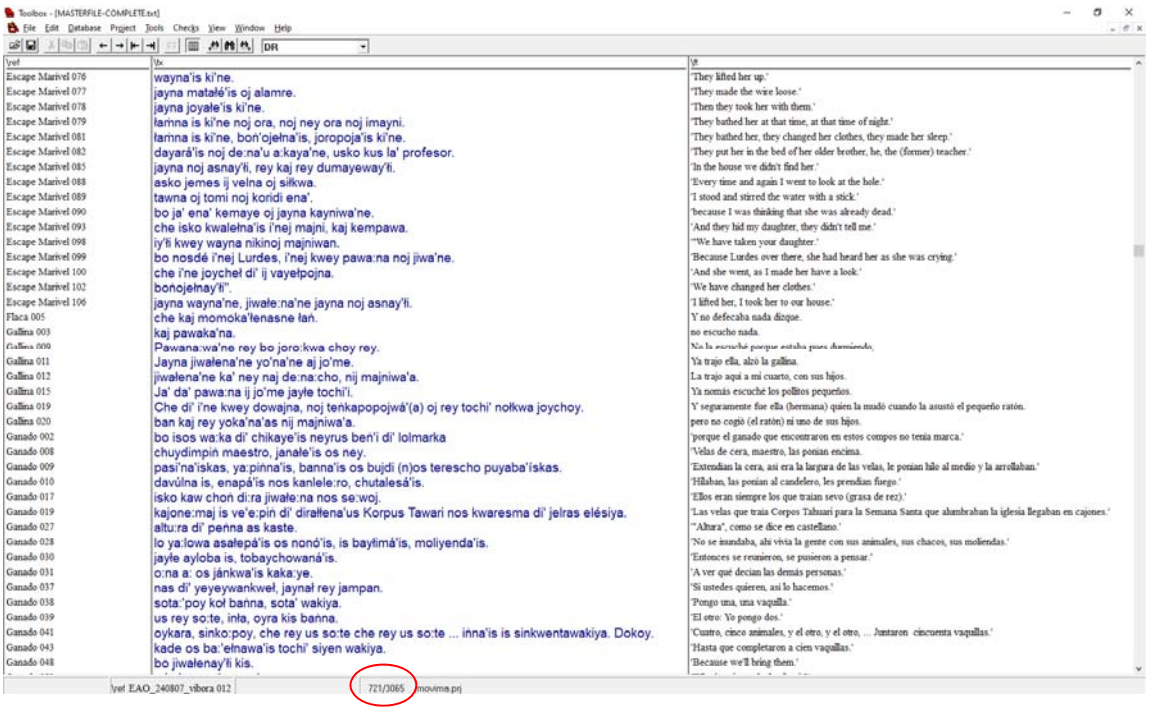


Figure 7. Excerpt of Toolbox "Browse View" containing records with annotation 'DR'

One obvious disadvantage of not glossing all the data is that this makes the corpus less accessible to outsiders. However, first of all, for the time being I am the principal user of the corpus, so my priority is to make it useful for myself. Furthermore, the most important step towards making data accessible is the translation: glosses can always be added later, even by a different analyst, and without the help of a native speaker. Therefore, transcription and translation are generally considered the minimal requirement for useful annotation (see

Schultze-Berndt 2006). Finally, the corpus does contain texts and individual segments that *are* glossed, so that there are selected pieces of data that allow external users deeper insights. Whenever required for a particular project, further glosses can be added without major problems by using the automatic glossing facilities in Toolbox or ELAN.

The approach I use for my data is not only time-saving in comparison to morpheme-by-morpheme glossing, since I do not have to gloss every single morpheme; it is also more useful for me than glossing, since it can capture zero morphemes. For instance, inverse transitive verbs (discussed in detail in 4.1.3 below) with a particular morphological structure are not overtly marked as inverse when nominalized. An example is given in (1). The nominalized verb *kay-poj-wa* 'feeding/being fed' is transitive, as is evident from the presence of two cliticized core arguments (see 4.1.1), *=y'ɬi* '1PL' and *--kisne* 'OBV:3F.AB'. However, the verb does not have the direct or inverse marker that, in Movima, indicates the semantic roles of the two arguments. For a subordinate transitive verb, the absence of a marker encodes the inverse, meaning that the internal argument (here, *=y'ɬi*) is the undergoer, while the external argument (here, *--kisne*) is the actor.

(1)	*bo*	*os*	***kay-poj-wa*=*y'ɬi*--*kisne***	*jayna*
	REAS	ART.N.PST	eat-CAUS-NMZ.EVT=1PL--OBV:3F.AB	DSC
	'… so that she would feed us (i.e., give us to eat).'			[EAO In between 206]

By analyzing my data on the basis of glosses alone, I would miss examples of zero coding like (1). With explicit tagging for inverse voice as identified from the larger structure, by contrast, these examples are captured. The Toolbox record containing example (1) is shown in Figure 8; the \ind(ex) fields indicate specific features of the example that can be filtered and searched for: It contains nominalization, a causative marker, and (zero) inverse marking. (The line "\3Pant", created as part of the "3Pant" project described in Section 2.1.3, indicates that the construction describes a three-participant event; here, the event is of "type A" and encoded with "strategy 2b".)

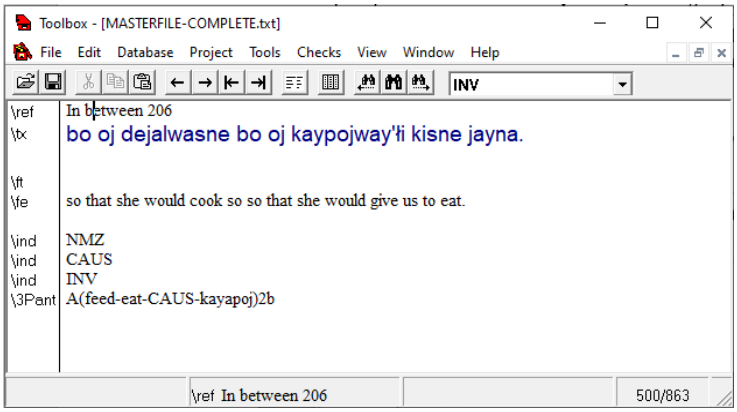


Figure 8. Toolbox record with a zero-marked inverse verb (example (1))

For some studies, I used more sophisticated tagging systems. When I investigated the direct-inverse alternation in the third-person domain (see Section 5.1 below), I wanted to identify all the basic transitive clauses in my corpus with two third-person arguments, be they overt or unexpressed. I created a special annotation tier for tagging transitive constructions, called \trans, in which I inserted the following information: the argument constellation (here: "3>3" for two third persons), direct or inverse marking, overt or zero expression and, if overt, grammatical relation of the arguments, and the animacy categories of the arguments' referents: human, animate non-human, and inanimate. In this way I tagged all the clauses with two third-person arguments in my corpus. On the one hand, this tagging facilitated the search for particular example types, e.g. examples of transitive clauses with two full noun phrases, or of clauses with zero arguments (which would, again, not be captured by glossing); it also helped me to look systematically for the correlation of inverse marking with animacy or with lexical vs. prononominal vs. zero expressions. On the other hand, the annotation method allowed for quantitative analyses, as described in Section 5.1. Again, not all of these features would have been captured by morphological glosses.

At present, I am particularly interested in intransitive clauses (see Section 6.1.2, Section 6.2). As is probably the case for any language, intransitive clauses outnumber transitive clauses by far, and intransitive clauses are represented by more different clause types. In my annotation of Movima intransitive clauses (in a tier "\clause") I distinguish between different predicate types (e.g. verb, adjective, possessed/unpossessed/proper noun) and between different argument expressions (e.g. free/bound pronoun, referential phrase, zero), which I identified on the basis of the data. The catgories and abbreviations I used for tagging are shown in Figure 9.

Predicate type:	
INTR	= intransitive verb
PREDN	= predicative noun
ADJ	= predicative adjective
POSSPRED	= possessive predicate (reduplication)
PREDNPOSS	= predicative possessed noun
PREDNPROP	= predicative proper noun
LOC	= locative noun (subtype of PREDNPOSS, but deverbal)
ADV	= adverbial predicate (e.g. *ney* '(be) here')
DEM	= demonstrative predicate
Argument expression type:	
#RP	= referential phrase
#FPRO	= free personal pronoun
#PRO	= bound personal pronoun
#0	= zero argument
#SAP	= speech-act-participant
#DEM	= demonstrative pronoun

Figure 9. Tagging categories for Movima intransitive clauses

With this schema, I have so far annotated about 4000 annotation units, which is only a small part of all the intransitive clauses in the corpus. Once the annotations are in place, I will dispose of valuable data that will help me for a number of research questions, like for instance the proportion of lexical expressions of intransitive subjects with respect to lexical expressions of transitive arguments (see 8.3).

My aim is to use the entirety of my corpus to approach my research questions. The corpus may not be large in comparison of standard language corpora, but it is large in comparison with those corpora that usually form the basis of comparative studies. For instance, the MultiCast corpus (https://lac2.uni-koeln.de/en/multicast/) allows text corpora with 1,000 clauses, as these can already provide significant results. However, my reasoning is that if I have a corpus of over 20,000 clauses, I should do all I can in order to exploit it in its entirety.

3.4 Engagement for an endangered language

Figure 10. Movima speakers
(from left to right: Jovina Amblo Ovales, her daughter Modesta Amblo Ovales, her niece Etelvina Gualusna Amblo, and Elías Ovales Rodríguez, a guest)

Linguists working on an endangered language cannot just collect data in the field and then leave. If they cannot work in a team with members of the speaker community, they have the minimal responsability to share the collected data with the community and, as far as possible, to share also their knowledge.

At each of my initial field trips I left copies of my written notes, the updated word list, and tapes with copies of the recorded texts with at least one representative of the speech community. With the DobeS documentation project (see 2.1.1), however, the scale of recorded data became too large to allow proceeding in this way. Giving the community access to the collected data was a central goal of the DobeS documentation project, however. The idea behind the DobeS programme was that the data be archived on central servers, where they could be accessed via the internet. However, as in many places where minority languages are spoken, only very few people in Santa Ana del Yacuma had a computer, there was no connection to the internet, and there was little literacy among the speakers I worked with. The DobeS archive is not very user-friendly anyway, and even academics have problems accessing and using it. Access via the internet was, therefore, not a solution that could be proposed to the speakers, and so, me and the other team members of the project tried to find other ways.

One of the things we did was to create what we called a “DVD library”: a collection of 15 DVDs with one DVD for each of our main collaborator. Most of the speakers had access to DVD players via their relatives, so in this way they could directly appreciate their contribution. One of the student assistents created a way to select, for each little movie, subtitles in either Spanish or Movima. This was very time-consuming: it involved selecting the right video segments including the time codes, adjusting the transcriptions so that they could serve as subtitles (i.e., “cleaning” them of errors and hesitation marks, making them fit in length etc.), and putting together the information for the student assistant who implemented the subtitles and burned the samples on DVDs together with a lead text featuring the metadata (sources, dates, participants, authors etc.). Inlets were created with photos of the respective speaker(s) featuring on the DVDs. We then distributed these DVDs to the individual speakers, since there was no common place where they could be stored that would be accepted by everybody.

The speakers liked the DVDs, but for many, it was difficult to watch them because they depended on other people for that. When we came back for fieldwork a year later, many of the DVDs were lost or broken and the people asked for replacements, which unfortunately we had not thought of. So in the end, our endeavor was well-intended, but the result did not match the effort.

For my last field trip, in 2012, I bought several external harddisks, on which I had saved a copy of the Movima DobeS archive. I distributed these DVDs among the indigenous institutions, such as the *Cabildo Indigenal*, the *Subcentral Indígena Movima*, and the *Consejo de Hablantes*. At every occasion, we took a photo in order to document that the data were handed over, and to make it clear that the data we collected were not “taken away”, but remained in the community. It remains a dilemma, however, that the harddisks are probably locked away at some place, that a computer is needed in order to see them, and that the organisation of the archive is not very accessible despite my efforts to give it a clearly understandable structure.

Figure 11. Handing over a harddisk with the Movima DoBeS archive to representatives of the *Cabildo Indigenal* in 2012

In any case, the speaker community, especially the younger language activists who are not themselves fluent in Movima anymore, were much more interested in materials that would help them more directly in their revitalization efforts. So I also tried to create materials that could help them to teach the language at school (as requested also by the Bolivian government). Already during my thesis, I had never left without distributing copies of the word list I was elaborating in Toolbox (see Figure 12). This word list, into which also a student assistant in Cologne had invested considerable time, met a surprising fate: a member of the Subcentral Indígena hand-copied it and gave it to the Bolivian Education Ministry for publication. When I came back to the village in the following year, I was at first rather shocked at this plagiarism. But then I told myself that, first of all, "my" word list had also been based on work by others (which I had explicitly mentioned in the introduction): by Eligardo Chirimani, one of my first contacts in the village, who had created a very nice illustrated dictionary but never intended to publish and distribute it; and by the SIL linguists, whose word list (Judy & Judy 1962) had also been included Chirimani's list. Furthermore, I consider that a word list, especially at this rather low level of sophistication, cannot as easily be claimed to be the collector's intellectual property as, for instance, a grammatical analysis. Therefore, now I do not regret the "unauthorized" publication of my unfinished work anymore. Rather, it gave me the liberty to state that members of the speaker community had taken in charge the dictionary and that I could fully concentrate on the grammar. I transmitted Toolbox, the necessary fonts, and all other helpful computer tools to the person who had published the dictionary, hoping that he will continue to improve it.

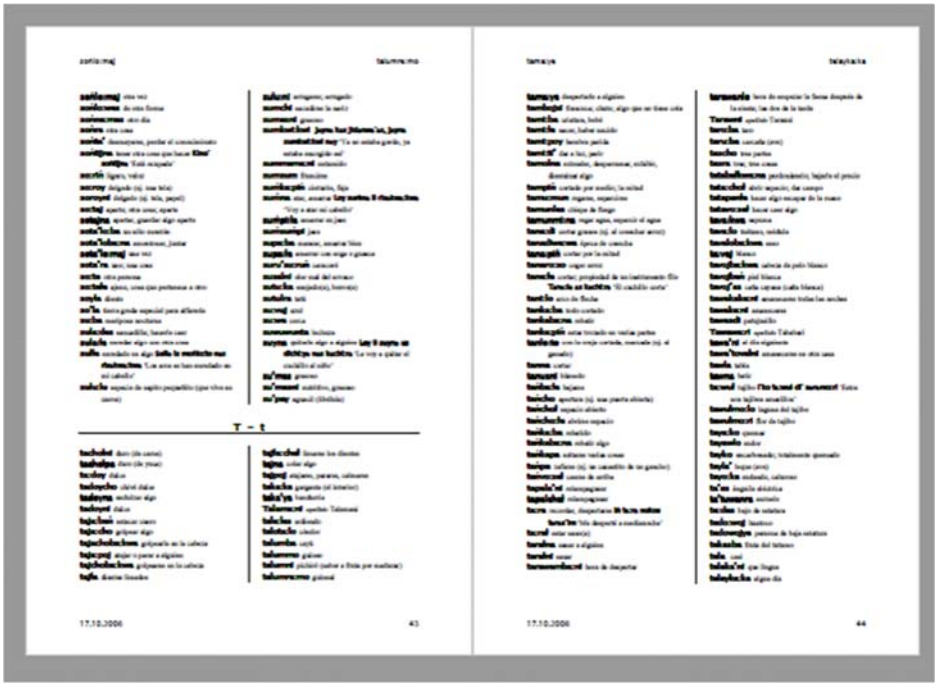


Figure 12. Excerpt of the draft word list

I also tried to engage in local capacity building and the creation of material that could serve as a basis for teaching. My proposals regarding the Movima alphabet and orthography were officially accepted by the speaker community, and my Spanish adaptations of the grammar, finally published as Haude (2011a) (see Figure 13), are actually being used by local teachers. I carried out several workshops with Movima speakers, including an orthography workshop, dictionary workshops, and classes on specific grammatical phenomena. shows two of the pictures with which I tried to illustrate my classes.

a. b.

Figure 13. The *Esbozo gramatical* in its first print version (a.) and as a revised publication (including the phonology) by Eibamaz/UNICEF (b.)

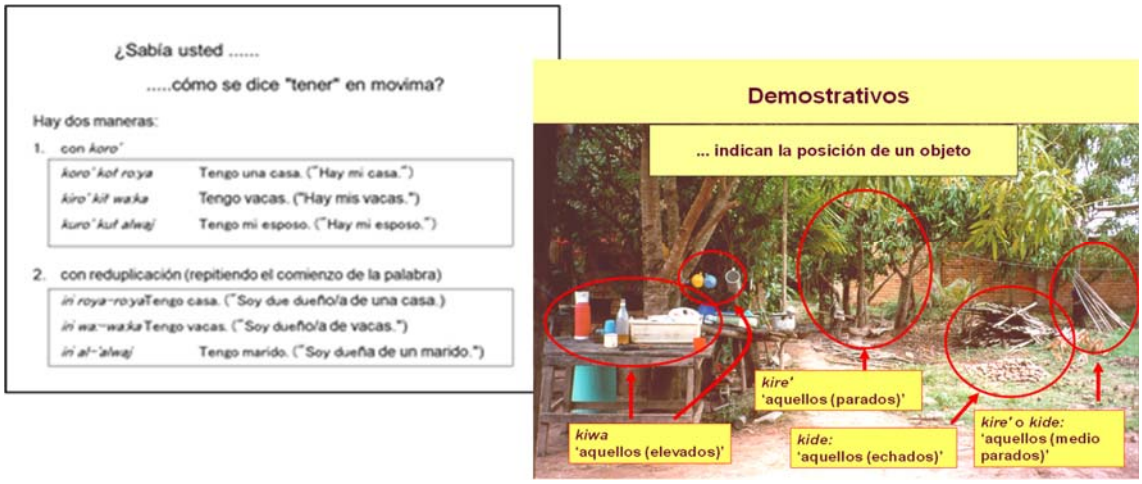


Figure 14. Mini-posters illustrating grammatical peculiarities of Movima

Many of my efforts seemed rather futile, and it will never be possible to give back to the speakers what I owe them. Still, I realized that in the course of the twelve years during which I regularly went to the field, my continuous presence in the village and my efforts to learn the language boosted the community's interest in their own language. During my last trip, when I brought the books to the village, I felt that the community was finding its ways to keep their language from being forgotten.

4. The Movima challenge to alignment typology

Haude, Katharina. 2008. Clause structure and alignment in Movima. *Amerindia* 32: 89-107. (Special issue on Grammatical Relations in languages of Amazonia, edited by Ana Carla Bruno, Frantomé Pacheco, Francesc Queixalós, and Leo Wetzels.)

Haude, Katharina. 2009a. Hierarchical alignment in Movima. *International Journal of American Linguistics* 75(4): 513-532.

Haude, Katharina. 2011b. Argument encoding in Movima: the local domain. *International Journal of American Linguistics* 77(4): 559-571. (Special issue edited by Antoine Guillaume and Françoise Rose on Argument Encoding Systems in the Languages of Lowland Bolivia.)

Haude, Katharina. 2019d. Grammatical relations in Movima: alignment beyond semantic roles, in Witzlack-Makarevich, Alena and Balthasar Bickel (eds.). *Handbook of Grammatical Relations*, 213-256. Amsterdam/Philadelphia: Benjamins.

This chapter presents my work on alignment patterns in Movima basic affirmative clauses. Having graduated from the Linguistics department of Cologne University with its typological-functional orientation, I was well acquainted with the typology of grammatical relations and the fundamentally different ways in which languages encode "who acts on whom". Therefore, one of the most challenging discoveries I made when writing my grammar was the difficulty to define grammatical relations in Movima.

In Section 4.1, I outline argument encoding in Movima direct transitive clauses and show that Movima shows features of three alignment types. In 4.2, I present my proposal that all these features can in principle be traced back to one single source, namely an intransitive, nominal clause.

Like most of my publications on this topic, the discussion in this and the subsequent sections focuses mostly on the expression of third-person argument encoding. First- and second-person argument encoding (discussed in detail in Haude 2011b) does not contradict anything of the findings presented here, but it is somewhat more complex. Rather than going into the intricacies that a description of the encoding of first- and second-person arguments would necessarily involve, I prefer to restrict the domain of description to make it easier to follow the argumentation.

4.1 A hybrid of different alignment types? Ergative, inverse, and symmetrical voice

Grammatical relations indicate the function an argument has in the clause it appears in. According to a useful definition formulated by Bickel (2011: 402), grammatical relations are "equivalence sets of arguments". The best-known and most-discussed grammatical relation is the "subject" of Indo-European languages, which shows up through the grammatical equivalence between the actor argument (A) in a transitive clause (typically an agent or experiencer) and the single argument (S) of the intransitive clause. They both have the same formal properties, e.g. they bear the nominative case (e.g. English *he* vs. *him/his*) and trigger verbal agreement (as in English *She runs, She calls him*). Apart from these formal properties, grammatical relations also show up through equivalent syntactic behaviour (see Keenan 1976). For instance, the Indo-European subject is zero-expressed in a coordinated clause: Both in *She calls him and Ø runs* and in *She runs and Ø calls him*, only the subject may (or must) be left unexpressed in the conjoined clause. This property has led to the proposal to call an argument that can be omitted in particular constructions "Privileged Syntactic Argument" (PSA; Van Valin & LaPolla 1997). This term is extremely useful since it does not imply that in a transitive clause, this argument is automatically the actor, which is typically associated with the notion "subject".

To identify grammatical relations, it is first of all necessary to identify the ways in which the arguments are encoded (e.g. in a particular position of the clause, or case-marked), and if and how their presence in a clause affects other elements of the clause (e.g. by triggering agreement on the verb). In Section 4.1.1, I show that in Movima, there are clear formal differences between the two arguments of a transitive clause. However, they are quite subtle, sometimes marked only by suprasegmental features. Since most of my publications deal with topics of morphosyntax, I have to explain these patterns in almost every single paper I write to enable the reader to understand the data.

As a next step, it must be investigated if there are formal equivalences between the arguments in the transitive and those in the intransitive clause, and on what principles they are based. Typically, grammatical relations are associated with semantic roles. In an "accusative" alignment pattern, the argument of a transitive clause that is equivalent to S of a canonical intransitive clause is the actor argument. In an "ergative" alignment pattern, by contrast, the argument of a transitive clause that is equivalent to S of a canonical intransitive clause is the undergoer argument (which may represent different non-actor roles like patient, theme,

recipient, stimulus). In Section 4.1.2, I show that this alignment pattern can be found in Movima, and this fact by itself is already very interesting.

However, Movima is more complicated. There are two transitive constructions: besides the one with the ergative pattern, there is another transitive construction, which patterns accusatively. This is the so-called "inverse" construction, which is used to describe events in which the actor ranks lower in terms of e.g. animacy than the undergoer (as in, for instance, the English sentence *A car hit my sister*). Direct-inverse systems have been presented in the typological literature as representing an alignment type of their own, and I discuss this possibility for the analysis of Movima in Section 4.1.3.

Another way to look at the Movima alignment pattern is from the perspective of so-called "symmetrical voice"-systems, as known from Western Austronesian languages. This approach is argued for in Section 4.1.4. While I find this perspective the most promising one, in Section 4.1.5 I report on a study which shows that despite the symmetry between the two transitive constructions, ergativity may be more basic after all in Movima.

In sum, from the perspective of morphosyntactic typology, the alignment patterns of Movima can be analyzed as any one of the following syntactic systems, which are often presented as being independent of each other (see e.g. Dixon & Aikhenvald 1997):

- ergative
- inverse
- symmetrical voice

Before I illustrate each of these possibilities below, Section 4.1.1 provides, first of all, the preliminaries by explaining the argument encoding patterns in the most basic transitive clause type, the direct clause. Section 4.1.2 adds information on argument encoding in intransitive clauses and illustrates the ergative pattern of the direct transitive clause. Section 4.1.3 goes on to describe the other transitive construction, the inverse, with its accusative alignment properties. Section 4.1.4 presents the system as a symmetrical voice system. Section 4.1.5., building up on the previous sections, again gives some evidence for a possibly underlying ergative bias of bivalent verbs. In Section 4.2 I present my proposal of how this unusual mixture of typological patterns might be explained.

4.1.1 Argument encoding in direct basic transitive clauses

A Movima basic clause is maximally monotransitive, i.e. it may have no more than two core arguments. The arguments are distinguished by their position in the clause and by the morpho-phonological effect their expression has on the preceding element. Furthermore, their semantic roles (A, i.e. the most actor-like role, and P, i.e. the most undergoer-like role) are indicated by 'direct' vs. 'inverse' morphology on the verb.

In what follows, I will focus on 'direct' transitive clauses (see Section 4.1.3 for the 'inverse' transitive counterpart). I do so because the direct construction is the most frequent one (see Section 5.1), because it is possibly more basic (see Section 4.1.5), and because the direct construction is the basis of the ergative pattern described in Section 4.1.2.

Movima basic clauses are predicate initial. In the direct transitive clause, the order of the core constituents (predicate and core arguments) is V-A-P. Furthermore, the two arguments of a transitive clause are distinguished by the ways in which they are linked to the predicate. I have labelled the different types of attachment "internal cliticization" and "external cliticization"; the former is represented by the clitic symbol " = ", while the latter is represented by the symbol " -- ", which I invented especially for this purpose (Haude 2006). As we will see, regarding Movima it is not possible to speak uncontroversially of case marking or of a subjects or object relation. Distinguishing two types of cliticization may seem strange, but it has proved to be the best solution for describing these matters.

Internally cliticized elements are suffix-like in that they participate in the stress pattern of the word. Consider first of all the direct transitive verb without person marking in (2a). The absence of an overt person marker on a transitive verb implies the first person singular. The verb without an overt internal enclitic follows the typical prosodic pattern of Movima independent words, with stress on the penultimate syllable and lengthening of this syllable when it is open. When a non-zero person marker is attached through internal cliticization, as in (2b), the stress shifts to the right, which lends the unit the appearance of a new prosodic word, and the lengthening of the vowel is lost (it does not shift to the new penultimate syllable, which would be the case if this were suffixation).

(2) a. *aya:-na=Ø*
wait_for-DR=1SG
[ʔaˈ**ja**:na]
'I waited for (you/him/her/it/them).'

b. *aya-na=us*
wait_for-DR=3M.AB
[ʔajaˈ**naʔ**us]
'He waited for (him/her/it/them).'

Internal clitics cannot be attached to a consonant, but require a preceding vowel: on consonant-final hosts, the linking vowel *-a* is inserted, from which a vowel-initial clitic is separted by a glottal stop; see (3). There is no suffix in Movima that triggers a comparable process.

(3) *kay-a-poj-a=us*
eat-DR-CAUS-LV=3M.AB
[kajapoˈh**aʔ**us]
'He fed (him/her/it/them).'

Internal cliticization is also distinguished from suffixation in that it involves determiners, i.e. elements belonging to a different syntactic phrase. This can be seen in (4): also before the cliticized article, the linking vowel is inserted, just as in (3). Determiners are thus ditropic clitics, which are phonologically attracted to the preceding element while belonging syntactically to the following one.

(4) *kay-a-poj-a=kus* *dichi:ye*
eat-DR-CAUS-LV=ART.M.AB child
[kajapoˈhakusʔitiˈlaːkwa]
'The/a boy fed (him/her/it/them).'

In direct transitive clauses, the argument which shows "internal cliticization" as described above, bears the A role. If the P argument is overtly expressed, it follows the A argument, i.e. it comes in second position after the verb. A pronoun expressing P in a direct transitive clause is phonologically attached to the preceding element through "external cliticization". External cliticization is weaker than internal cliticization and more like the juxtaposition of two elements. The external cliticization of a vowel-initial pronoun to a vowel-final host shows no sign of phonological attachment, such as shifting stress or lengthening. Consider (5), in which the zero-marked first-person form from (2a) is combined with an externally cliticized 3rd person pronoun. This example contrasts directly with (2b), repeated in (6), in which the pronoun *us* is

internally cliticized. The difference between ‘He hits him/her/it/them’ and ‘I hit him’ is only marked by prosody, as can be seen from the phonetic representations.

(5) *aya:-na=Ø--us*
wait_for-DR=1SG--3M.AB
[ʔaˈ**ja**:naʔus]
‘I waited for him.’

(6) *aya-na=us*
wait_for-DR=3M.AB
[ʔajaˈ**na**ʔus]
‘He waited for (him/her/it/them).’

The fact that an externally cliticized pronoun is phonologically attached to the host becomes apparent with consonant-final hosts. Here, the host-final consonant forms the onset of the new syllable. In contrast to internal cliticization, the phonological attachment only shows up with pronouns, not with determiners (which can be homophonous with pronouns, like the article *us* ‘ART.3M’ with the pronoun *us* ‘3M.AB’). An example with a full referential phrase is given in (7); as the phonetic representation shows, the glottal stop that precedes vowel-initial morphemes is retained on the article.

(7) *kay-a:-poj=Ø--us*
eat-DR-CAUS=1SG--3M.AB
[kaˈja:pohus]
‘I fed him.’

(8) *kay-a:-poj=Ø* *us* *itila:kwa*
eat-DR-CAUS=1SG ART.M man
[kaˈja:pohʔusʔitiˈla:kʷa]
‘I fed the/a man.’

Example (8) shows that the argument phrase that is external to the predicate phrase has no phonological connection to the predicate when expressed by an RP; the article is not cliticized. Furthermore, other elements can intervene between the predicate and this RP, e.g. particles, as

in (9); and, as is apparent from the fully grammatical sentences in (2)–(4), the externa; argument is not obligatorily realized.

(9) *kay-a:-poj=Ø* ***ja'*** *us* *itila:kwa*
eat-DR-CAUS=1SG just ART.M man
'I just fed the/a man.'

The difference between internal cliticization, on the one hand, and external (i.e. less tight) or no cliticization, on the other hand, is an important insight regarding Movima argument encoding. It shows that the difference between the arguments in a transitive clause is not only marked by linear order (as e.g. in English), but also by constituency. Following a suggestion by Francesc Queixalós (p.c.) I consider it most adequate to characterize the two arguments in terms of their being expressed in a position "internal" or "external" to the predicate phrase. The internal argument is inseparably combined with its head (i.e. the transitive verb), forming a syntactic phrase with it. The external argument, in contrast, shows a looser connection, can remain unexpressed, can be separated from the predicate phrase by other elements, and can be fronted (see below).

This use of the terms "internal/external argument" is not to be understood as in the original proposal by Williams (1981) and its followers (e.g. Grimshaw 1990). In the original definition, within the Generative framework, these argument positions are directly linked to semantic roles: the "external" argument bears the A role, while the "internal" argument, which is the object NP inside the VP, always bears the P role. In the Movima direct clause, however, the internal argument is the A argument. (The contrary is the case with inverse-marked verbs, to which I will turn in 4.1.3 below). The problem is due to the ergative character of the direct transitive clause, of which we will see more evidence in the following section.

4.1.2 Ergativity

4.1.2.1 Accusative and ergative alignment

Ergativity, in the way I use the term, is an alignment phenomenon. Alignment is when two different constructions share one element, i.e. when they both have an element that shows the same properties in both constructions: the elements are "aligned" over the two constructions. In morphosyntax, this term is applied, for instance, when comparing argument encoding in transitive and intransitive clauses. An intransitive clause has one argument (S), a transitive

clause has two, usually labelled A (for the more agent-like argument) and O or P (for the more patient-like argument). Usually, one argument of the transitive clause is expressed like the single argument of an intransitive clause. In Indo-European languages like English, French, German, or Latin, this argument is in the nominative case and governs verbal agreement. Consider the examples of a Latin intransitive and transitive clause in (10) and (11) (adapted from Comrie 1978: 331).

(10) *Puer* *veni-t.*
boy(NOM) come-3SG
'The boy came.

(11) *Puer* *puella-m* *ama-t*
boy(NOM) girl-ACC love-3SG
'The boy loves the girl.'

In the intransitive clause, (10), the single argument is in the nominative (unmarked) case. The transitive clause, (11), also has an argument in the nominative case, and this is the A argument. The P argument of the transitive clause, by contrast, is marked as accusative. Based on the case of the argument that is marked differently from the others, systems of this kind are called "accusative alignment systems".

The other major type of role-based alignment is that in which the argument of the transitive clause that shares the formal properties with the S of the intransitive clause is the P argument. This alignment type is much rarer in the languages of the world, and ergative patterns are usually not consistent over the entire system of a language. Ergative patterns are fascinating because of the "preference" that is given to the non-actor argument: in a canonical ergative system, the P argument lacks overt case marking, the verb agrees with P, and sometimes even syntactic privileges show up for the P argument.

In ergative languages with case-marking, the A argument is in the so-called ergative case, while both the P argument of the transitive and the single argument of the intransitive clause are in the absolutive case, which in many languages is morphologically unmarked. Consider the example from the Australian language Dyirbal below (from Comrie 1989: 112; I have completed the glosses on the basis of Comrie's comments and the explanations in Dixon 1994: 10). In the intransitive clause in (12), the single argument (S) is unmarked, i.e. in the absolutive.

In the transitive clause in (13), the unmarked argument is P. The A argument, by contrast, is marked with the ergative suffix *-ŋgu*.

(12) *bayi* *yaṛa* *bani-nʸu*
CLF man:ABS come_here-NONFUT
'The man came here.'

(13) *balan* *dʸugumbil* *baŋgul* *yaṛa-ŋgu* *balga-n*
CLF woman:ABS CLF man-ERG hit-NONFUT
'The man hit the woman.'

The possible syntactic effects of ergative or accusative alignment have been pointed out by Dixon (1972; 1994) on the basis of the Australian isolate Dyirbal. When in an accusative system a transitive and an intransitive clause are coordinated, the unexpressed argument of the intransitive clause is automatically interpreted as coreferential with the A argument of the preceding transitive clause. In the English sentence *The man hit the woman and came here* (Comrie 1989: 112), it is understood that it is the 'man' who came, not the woman.

In the Dyirbal example (14), by contrast, the unexpressed S argument of the coordinated intransitive clause is automatically interpreted as being coreferential with the P argument of the preceding transitive clause, the 'woman', with which it shares the absolutive case.

(14) *balan* *dʸugumbil* *baŋgul* *yaṛa-ŋgu* *balga-n,* *bani-nʸu*
CLF woman:ABS CLF man-ERG hit-NONFUT come_here-NONFUT
'The man hit the woman and (the woman) came here.'

Syntactic ergativity in coordination seems to be rare cross-linguistically, but I like to use this example because it is a very good illustration of P being the privileged syntactic argument of a particular construction.

Comrie (1978) has introduced useful schemas to represent role-based alignment patterns graphically. I have reproduced the schemas for accusative and ergative alignment in Figure 15. They show that in an accusative system, A and S have the same properties while P is distinct, and that in an ergative system, P and S have the same properties while A is distinct. Bickel (2011) proposes a simpler representation, representing accusative alignment as {A=S} and ergative alignment as {P=S}.

Figure 15. Accusative and ergative alignment schematically

For my research on Movima alignment, the above schemas were particularly helpful because they do not necessarily imply that the formal alignment must be manifested through case marking (even though case marking is the most uncontroversial criterion for defining alignment types): the shared properties can also manifest themselves in other ways.

4.1.2.2 Ergativity in Movima: {S=P_{DR}}

As was apparent in Section 4.1.1, in Movima there is no morphological case marking on the core arguments. Rather, the arguments of the transitive clause are differentiated by their constituency, which is evident from their linear order, their morpho-phonological connection to the predicate, and their obligatoriness. Furthermore, they show different behavioural properties (a criterion introduced by Keenan 1976), which become apparent in their potential to be relativized and "fronted" (see 5.3 below). Based on these criteria, I was able to show that in Movima, the single argument of the intransitive clause, S, shares the properties of the P argument of the direct transitive clause (henceforth: P_{DR}), while the A argument of the direct transitive clause (henceforth A_{DR}) is encoded differently.

When S of a Movima intransitive clause is represented by a pronoun, the pronoun is connected to the predicate by external cliticization. As shown by the phonetic representation in (15), the pronoun resyllabifies with a host-final consonant. When S is expressed by an RP, the article is not connected to the predicate at all, as in (16). An RP representing S can be separated from the predicate phrase by other elements, as in (17). Finally, S can remain unexpressed, as in (18). All these properties are the same for P_{DR}, as was shown in Section 4.1.1.

(15) *ajła:bał--us*
walk--3M.AB
[ʔahˈła:bałus]
‘He walked.’

(16) *ajła:bał us itila:kwa*
walk--3M.AB ART.M man
[ʔahˈła:bałʔusʔitiˈla:kʷa]
‘The/a man walked.’

(17) *ajła:bał **ja’** us itila:kwa*
walk--3M.AB just ART.M man
‘The/a man just walked.’

(18) *ajła:bał*
walk
‘(I/you/he/she/it/we/they) walked.’ (context-dependent)

There is only one coding difference between S and P_{DR}: a third-person pronoun representing P_{DR} is sometimes preceded by a *k-*, as in (19). I have analyzed this element as an ‘obviative’ prefix (Haude 2006: 279–280) because it only occurs when preceded by an internal argument referring to a third person (including the first person plural exclusive, ‘me and others’), as in (19). (‘Obviative’ is a category that differentiates between two third person referents with different discourse status; see 4.1.3.) Hence, I do not regard this *k-* as evidence for a formal difference between P_{DR} and S. Its occurrence is due to the presence of a further argument in the clause, which is only possible in transitive clauses.

(19) *vel-na=us--**k-**is*
watch-DR=3M.AB--OBV-3PL.AB
‘He watched them.’

Hence, despite this apparent difference, the S argument shares the property of P_{DR}; in the terminology of Bickel (2011), the direct transitive construction shows ergative alignment, with a grammatical relation {P=S}.

In addition to the coding properties of the arguments, alignment patterns can also show up in the arguments' syntactic behaviour, again with one argument of the transitive clause patterning with the single argument of the intransitive clause. Here, one can speak of a Privileged Syntactic Argument (PSA; Van Valin & LaPolla 1997). The syntactic processes that are relevant here (see, for instance, Dixon 1994; Van Valin & LaPolla 1997; Bickel 2011) include reduction under coordination (i.e., the privileged argument is unambiguously retrievable when omitted in a coordinated clause), the potential for raising, as well as relativization.

In Movima, the potential for reduction under coordination or raising shows no clear preference for one of the arguments of a transitive clause (see Haude 2009a; Haude 2019d). However, a clear grammatical asymmetry in the syntactic behaviour of the two arguments can be observed in relativization and fronting (see 5.3): Only the external argument can undergo these processes. Example (20) shows the relativization of P_{DR}. The relative clause is introduced by the particle *di'*. "Fronting" (a construction that is discussed in detail in Chapter 6) is shown in (21).

(20) *us* *dichi:ye* *di'* *kay-a-poj-a=sne*
ART.M child REL eat-DR-CAUS-LV=3F.AB
'the/a boy that she fed'

(21) *u'ko* *kay-a-poj-a=sne*
PRO.3M eat-DR-CAUS-LV=3F.AB
'She fed <u>him</u>.'

In order to relativize or front the internal argument, i.e. A_{DR}, the clause must first be detransitivized. As a result, the former A_{DR} becomes the S argument of the derived intransitive clause, and the former P_{DR} loses its core argument status and is demoted to oblique. This operation is typical of ergative systems and is known as "antipassive": just like the alignment patterns depicted in Figure 15, the antipassive is the mirror-image of the passive operation of accusative systems.

The Movima antipassive is illustrated in the relative clause in (22) and in the fronting construction in (23). It is marked by a valency-decreasing particle *kwey* (or its variant *kaw*). The predicate becomes syntactically intransitive, which means that it cannot take an internal enclitic. The former A_{DR} is now S of the derived intransitive clause. Therefore, this argument can now be relativized, while the former P_{DR} can only be expressed as an oblique, if expressed

at all. (As we will see in 4.2.2 below, the process can also be applied to nouns; this is why the particle *kwey/kaw* is glossed VALDECR, for 'valency decrease').

(22) *kinos* *kwe:ya* *di'* ***kwey*** *kay-a:-poj* *(n-u'ko)*
ART.F.AB woman REL VALDECR eat-DR-CAUS-LV=3F.AB OBL-PRO.3M
'the woman who fed (him).'

(23) *isne* ***kwey*** *kay-a:-poj* *(n-us* *dichi:ye)*
PRO.3F.AB VALDECR eat-DR-CAUS OBL-ART.M child
'She fed the boy.'

With the data seen so far, Movima presents itself as a "syntactically ergative" language (Dixon 1994). Syntactic ergativity is a cross-linguistically rare phenomenon, and Dyirbal (Dixon 1972) is its most famous representative. In most languages that show ergative features in terms of argument encoding (e.g. case marking) or agreement, the syntactic privileges are still with the A argument. So, the fact that Movima shows features of syntactic ergativity is an interesting finding. In Section 4.2 I will provide a possible explanation of this pattern. Before that, however, I present the "inverse" construction, which opens still other ways of classifying Movima morphosyntactically.

4.1.3 A direct-inverse system, and terminological consequences

4.1.3.1 Referential hierarchies, obviation, and inverse

As was already mentioned, the direct transitive construction illustrated above has a transitive counterpart, called "inverse". This is illustrated in (25), which is the inverse counterpart of the direct construction in (24) (repeated from (7) above). There is no change in the argument expressions between (24) and (25). The only formal difference between the two constructions is the marking of the predicate: the predicate in (24) is marked as 'direct' by the base-internal affix *-a* (on the base-internal vs. base-final allomorphs of the direct morpheme, see Haude 2006: 323–325), while the predicate in (25) is marked as inverse by the suffix *-kay*. Direct marking, which we have already seen in the examples above, indicates that the internal argument is A and the external argument is P, while inverse marking indicates that the internal argument is P and the external argument is A.

(24) *kay-**a:-**poj=Ø--us*
eat-DR-CAUS=1SG--3M.AB
'I fed him.'

(25) *kay-poj-**kay**=Ø--us*
eat-CAUS-INV=1SG--3M.AB
'He fed me.'

Thus, in Movima, the argument positions are not (or not exclusively) occupied according to the semantic roles of the arguments. This is in contrast to, for example, English, where the position of the arguments before or after the predicate indicates grammatical relations, as shown in (26).

(26) a. *the woman fed the child*
b. *the child fed the woman*

In Movima, the expression of an argument in the internal or external slot depends primarily on discourse status (such as givenness) or ontological properties (such as animacy) of the referents. These features can be seen as ordered in a hierarchy: whatever is higher in a hierarchy of person (1>2>3), of animacy (human>animate non-human>inanimate), and/or of discourse topicality (given>less given or new) is expressed as the internal argument, independently of its semantic role. This is depicted in Figure 16.

Predicate	**internal**	**external**
	1SG/PL	2PL, 3
	2SG/PL	3
	3 [> animate]	3 [< animate]
	3 [> topical]	3 [< topical]

Figure 16. Properties of the referents encoded as internal vs. external argument in a transitive clause (> = "more", < = "less")

First person and second person singular can only be encoded in the internal argument slot. When a first singular or plural and second person singular interact, only the first person is encoded, and the second person is usually inferred from the context, as in (27).

(27) *di' joy-sicha'kwa, jayłe jayna choy rey ił, eney,*
HYP go-DES then DSC really EPIST 1 (filler)
joy-a:-łe=Ø, jankwa=u
go-DR-CO=1SG said_thing=3M
'If (you) want to go, then of course I'll, er, take (you), he said.'
[EAO In between 084]

The second person plural and third persons can be encoded in either slot, but this is determined by the relative ranking of the two arguments (see Haude 2011b). A second person plural is encoded as the external argument when the other participant is a first person (singular or plural); this is shown in (28), which also illustrates that the external argument is identical in transitive and intransitive clauses also with in the domain of speech-act participants. When the other participant is a third person, the second person plural is encoded by a different form in the internal argument slot, as shown in (29), which features a direct and an inverse clause.

(28) *che dewaj-na=Ø**--y'bi**, jayna rey nokowa joy-chełł**--iy'bi***
and see-DR=1SG--2PL.EXTL DSC EPIST FUT go-R/R--2PL.EXTL
n-as ki'laj
OBL-ART.N far
'And I saw you (pl.), now you will go far away.' [CCT_120907_2 137]

(29) *kwaj sotak-lomaj dis daj<a>wa**=nkwełł**--us di'*
EMPH one-CLF.time OPT ask<DR>=2PL--3M.AB HYP
*bawra-kay-a**=nkwełł**--us di' ka:'i*
pay-INV-LV=2PL--3M.AB HYP no
'You (pl.) should ask him at once if he is going to pay you or not'.
[EAO Llega Estel 004]

The encoding of two interacting third persons is based on the animacy or the discourse status of their referents, and will be discussed in detail in Section 5.1. The event participant that ranks higher in terms of animacy (e.g. human > inanimate) and that is familiar from the context is encoded as the internal argument, and the verb is marked as direct or inverse according to its semantic role.

Direct-inverse systems have first been described for Algonquian languages. An example from Plains Cree is given in (30) (direct) and (31) (inverse; Dahlstrom 1991: 36, 38, respectively). As in Movima, the two arguments of the transitive verb are encoded by the same expressions (*ni-* '1st person' and *-w* '3rd person'), independently of their semantic roles. The semantic roles of the arguments are indicated by direct and inverse marking on the verb: Direct marking indicates that the first person is the actor and the third person the undergoer, inverse marking indicates the opposite direction of action.

(30) *ni-wapam-**a**-w*
1-see-DR-3
'I see him.'

(31) *ni-wapam-**ekw**-w*
1-see-INV-3
'He sees me.'

The terms 'direct' and 'inverse' allude to the way in which an event with two participants is expected to take place. The term 'inverse' was, to my knowledge, first used by Hockett (1966), who explained it with the 'obviation scale' in Figure 17. (The term "obviation was already introduced by (The term "obviation" was already introduced by Bloomfield 1946.)

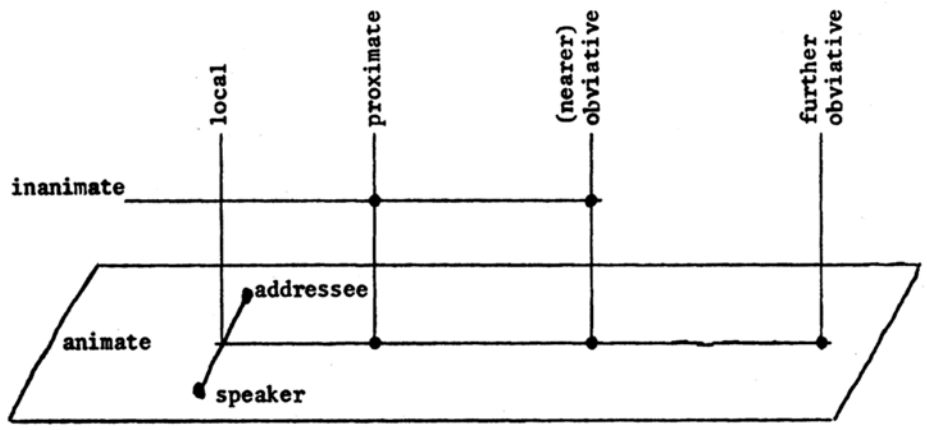


Figure 17. Hockett's (1966: 60) "obviation scale" for Algonquian

Hockett explains the inverse system as follows:

> Along the obviation scale given in [Figure 17], draw an arrow pointing away from local towards further obviative, and label it direct; draw an arrow in the opposite direction and label it inverse. When the actor of a transitive animate verb is closer (less obviated) than the goal, the form is direct; when the goal is closer than the actor, the form is inverse. (Hocket 1966:65)

The Movima inverse system can roughly be described along the same lines. Local or proximate event participants (i.e., speech-act participants and high-ranking third-person referents) are represented as the internal argument, while all other types of event participant are "obviated" and represented as the external argument (there is no "further obviative" category in Movima). When the proximate argument is the actor and the obviative argument is the undergoer, the direct form of the verb is used, and when the proximate argument is the undergoer and the obviative argument the actor, the inverse form of the verb is used.

Having introduced the inverse pattern, it is now possible to provide a more complete picture of Movima transitive clauses. Figure 18 is a schematic representation of Movima transitive clause structure. (I like to use it in powerpoint presentations, where I can build up the schema bit by bit.) We see that the predicate phrase, enclosed in square brackets, contains the verb and the internal argument ("ARG"). The verb is overtly marked as transitive, either by the direct or by the inverse marker. In the boxes on top of the schema, some formal properties of the argument expressions are given: The position of the internal argument – inside the same constituent as the verb, as indicated by the bracketing – is fixed, i.e. the element representing

it may not be omitted or fronted. This argument, furthermore, is obligatorily expressed: the absence of an overt element is a zero morpheme indexing the first person singular. By contrast, the external argument, enclosed in square brackets of its own, is not obligatorily overtly expressed. It can also be separated from the predicate phrase, as was shown in (9), and even be fronted, as will be shown below (Section 5.3 and Chapter 6).

The boxes below the nominal constituents ("ARG") indicate what kind of referent may be expressed by them, as explained above (Figure 16).

The two bottom lines, finally, indicate that both arguments can be either actor (A) or undergoer (P). Which one is A and which one is P is determined by the choice of the transitivity marker on the verb: direct or inverse.

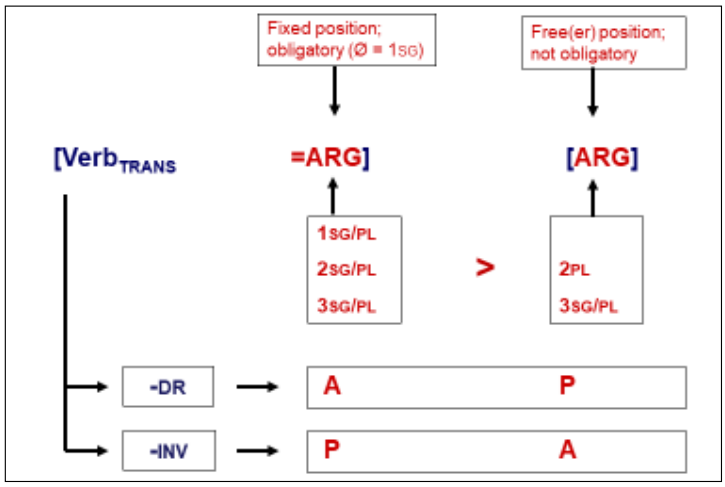


Figure 18. The assignment of syntactic argument positions and semantic roles in a Movima transitive clause

The Movima direct-inverse alternation thus functions similarly as in Algonquian. However, in Algonquian and most other known inverse systems, persons are indexed by bound affixes, whereas in Movima, the verb has no person affixes, and the direct-inverse alternation involves the nominal constituents. In Movima, furthermore, the direct-inverse alternation has immediate consequences on alignment: the external argument, which aligns with S, can either be P or A, depending on whether the verb is marked as direct or inverse. In Algonquian languages, by contrast, the question of alignment and grammaticalization is far more complex and less closely linked to the direct-inverse alternation (see Haude & Zúñiga 2016: 451).

Therefore, the ergative analysis presented in 4.1.2 must be adapted accordingly: There is an alignment split in the Movima system. When the referential hierarchy matches with the semantic-role hierarchy (actor>undergoer), transitive clauses pattern ergatively: the argument representing P aligns with S of the intransitive clause. When, however, there is a mismatch between the referential hierarchy and the semantic-role hierarchy, transitive clauses pattern accusatively: the argument representing A aligns with the single argument of the intransitive clause. This alignment split is shown schematically in Figure 19, which takes up the schematic representation in Figure 15 above, but additionally reflects the syntactic position of the arguments in the Movima clause.

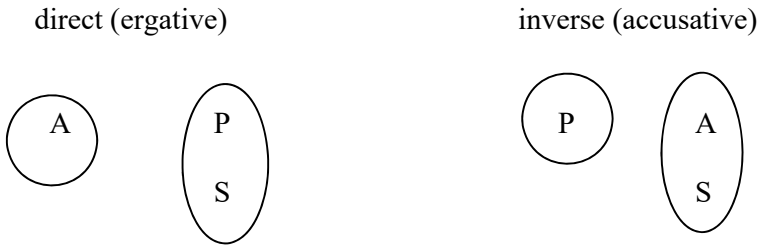


Figure 19. The direct/inverse split pattern

Alignment splits are a well-known typological phenomenon, and it is not uncommon that they are determined by referential hierarchies. For instance, Silverstein (1976) shows that in languages with split ergativity, elements higher in a "nominal hierarchy" (e.g. pronouns) tend to follow a nominative-accusative case-marking system, while those lower on the hierarchy (e.g. NPs) follow an ergative-absolutive system. However, an alignment split resulting from a direct-inverse system was so far not attested – at least not one in which the direct construction shows ergative alignment. Zúñiga (2006: 67) mentions the possibility of a "remapping inverse", in which the grammatical relations are swapped in the inverse construction, but in that case the direct construction is expected to pattern accusatively. Thus, the alignment pattern found in Movima is typologically remarkable, and to my knowledge nothing similar has been described for neighbouring languages.

4.1.3.2 How to term the grammatical relations?

The impact of the direct-inverse alternation on the identification of grammatical relations in Movima makes it extremely difficult to assign uncontroversial labels to the two arguments of the transitive clause. When taking the alignment split as a basis, one might label the internal

argument "A_{DR} and P_{INV}", and the external argument as "P_{DR} and A_{INV}". However, apart from being overly complex, this characterization does not reflect the principle underlying the grammatical relations in Movima. In my dissertation, I had labelled the arguments "ARG_1" and "ARG_2", according to their linear order. Here, however, the problem is that in some models, e.g. Relational Grammar, "ARG_1" is used to label the syntactically privileged argument (the subject), and this characterization is inadequate for the Movima internal argument.

Later on (e.g. Haude 2009a; Haude 2010a; Haude 2014a; Haude 2019a), I described the system in terms of obviation: the external argument encodes the "obviative" event participant. Therefore, inspired by Bickel (2011), I adapted the Algonquianist terminology and labelled the internal argument "proximate argument" ("PROX" or "ARG_{prox}") and the external argument "obviative argument" ("OBV" or "ARG_{obv}"). A potential problem here is that in the Algonquianist tradition, both terms only apply to third-person participants (see, for instance, Aissen's characterization: "obviation systems obligatorily rank *third person nominals* according to a complex function which includes grammatical function, inherent semantic properties, and discourse salience" (Aissen 1997: 705; my emphasis). However, in Movima, the "proximate" argument encodes any participant encoded in the internal argument slot, i.e. also speech-act participants. This is a problem for those who are used to the Algonquianist tradition. I personally do not see a problem in broadening the scope of the terms, but I do think it is inconvenient to label structural categories with terms that are based on semantic and discourse-pragmatic criteria. This can easily lead to confusion. For instance, Bickel (2011), where first uses the term "proximative" and "obviative" as labels for the Movima internal and external argument, respectively, equals the Movima "proximative" (i.e. internal) argument with the privileged argument in Tagalog, the so-called *ang*-phrase (see 4.1.4), which he also labels "proximate" without further explaining the choice. On this basis, he claims that grammatical relations in Movima and Tagalog function in exactly the opposite way. As I will show in 4.1.4, however, I see Movima and Tagalog as functioning pretty much in parallel ways, and so, it might be the case that Bickel's argument is partly based on an infelicitous choice of terminology.

Yet another possibility is to label the arguments according to the grammatical relations they represent. Since the external argument is syntactically privileged, it can be labelled "subject", while the nonprivileged argument can be labelled "non-subject". I did this in an article on voice (Haude 2012b; see 4.1.5), because voice alternations are usually characterized in terms of the role they assign the subject relation. This term is inconvenient, however, not only because in most transitive clauses (which have a direct-marked verb), the "subject" has the P role, and but

also because this argument – since it designates the "obviative" event participant – is never the discourse topic (see 5.1). It is thus counter-intuitive on two dimensions: that of semantic roles and that of discourse status.

With this difficulty of finding uncontroversial terms for labelling the Movima arguments, it seems that "internal/external" is still the most adequate choice. These terms were suggested to me by F. Queixalós in 2007, the reasoning being that they reflect formal properties alone and do not evoke on semantic or discourse-based connotations. Despite the objections from some Generative colleagues (for whom the "external argument" is automatically associated with the S/A role), I believe that these syntactically based terms are the most appropriate ones for characterizing the Movima arguments, and this is why I use them throughout this thesis.

4.1.4 Symmetrical voice

Haude, Katharina & Fernando Zúñiga. 2016. Inverse and symmetrical voice: on languages with two transitive constructions. *Linguistics* 54(3): 443-481.

When I talked to my former teacher at the Linguists department of the University of Cologne, Werner Drossard, about the difficulties of characterizing grammatical relations in Movima, he pointed me to literature on the "symmetrical voice" systems of Western Austronesian (or "Philippine-type") languages, especially Tagalog. Indeed, the structural similarities are striking, and the analyses of these systems in terms of ergativity, voice, or underlying intransitivity can almost identically be applied to Movima.

Philippine-type languages, with Tagalog as their most prominent representative, are known for possessing several transitive constructions, distinguished by verbal "voice" morphology that indicates the semantic role of the principal argument. The most significant alternation is that between "actor voice", shown in (32a), and "undergoer voice", (32b), which indicate the semantic role of the the *ang*-phrase, the central syntactic argument that has been described as "topic" (Schachter & Otanes 1972), "nominative argument" (Kroeger 1993), or "subject" (Himmelmann 2005a).

(32) Tagalog (Haude & Zúñiga 2016; adapted from Nagaya 2012: 50)

a. *K<um>ain* *ang=bata* *ng=tinapay.*
<AV.PFV>eat NOM=child GEN=bread
'The child ate bread.'

b. *K<in>ain* *ng=bata* *ang=tinapay.*
<UV.PFV>eat GEN=child NOM=bread
'The child ate the bread.'

The actor-voice (AV) infix <*um*> in (32a) indicates that the noun phrase with the 'nominative' proclitic *ang* (which also marks the single argument of the intransitive clause) is the actor. By contrast, the undergoer-voice (UV) infix <*in*> in (32b) indicates that the *ang*-phrase is the undergoer. (Note that constituent order is not fixed in Tagalog, so that the order of the noun phrases is interchangeable.)

The pragmatic functions or discourse conditions under which either of the voices is used in Tagalog are not entirely clear yet. In general, however, the *ang*-phrase has a definite interpretation, while the other noun phrase, marked by the 'genitive' proclitic *ng*=, may be indefinite. The syntactic properties of the different voices, by contrast, have been explored in considerable detail (see in particular Schachter 1976; Kroeger 1993). The voice alternations serve to maintain the *ang*-phrase as the syntactic "pivot", i.e. as a zero-expressed argument that is automatically retrieved in certain dependent constructions; for example, only the nominative argument can be relativized, and voice marking is chosen depending on the semantic role of the nominative argument. In (33a), the relativized nominative argument is the agent, therefore the actor voice is chosen in the relative clause. In (33b), the nominative argument is the patient, and therefore, the undergoer voice is chosen.

(33) Tagalog

a. *Matalino* *ang=lalaki=ng* *b<um>asa* *ng=diyaryo.*
intelligent NOM=man=LK <AV>read GEN=newspaper
'The man who read the/a newspaper is intelligent.'

b. *Interesante* *ang=diyaryo=ng* *b<in>asa* *ng=lalaki.*
interesting NOM=newspaper=LK <UV>read GEN=man
'The newspaper that the man read is interesting.'
(Haude & Zúñiga 2016, adapted from Schachter 1976: 500)

The Western Austronesian symmetrical voice systems, with Tagalog as a canonical representative, have been the subject of many typological discussions (see Riesberg 2014 for an overview) because they cannot be assigned to a role-based alignment type such as nominative-accusative or ergative-absolutive. Just like the Movima direct and inverse construction, the alternative voices of Tagalog are all transitive (unlike passive or antipassive voices in asymmetrical voice systems), and each construction is equally basic, i.e. none is derived from the other (unlike a passive, which can be derived from the active voice, or an antipassive, derived from the ergative). In both Tagalog and Movima, the constructions are largely chosen on the basis of discourse needs, and they are obligatory for certain syntactic processes, especially relativization.

The parallels between symmetrical voice systems and the Movima direct-inverse system are thus obvious: there are two (in the case of Tagalog and other Western Austronesian languages, even more) transitive constructions, none of which is derived from the other. There is a privileged syntactic argument, the *ang*-phrase in Tagalog and the external argument in Movima, which is the only argument that can be relativized. This is again illustrated below with examples that parallel the Tagalog relative clauses in (33). Example (34a) is a basic transitive clause with two RP arguments (see also (20) above). The direct marking on the verb indicates that the internal argument (*kinos kwe:ya*) is the actor and the external argument (*us dichi:ye*) is the undergoer. Example (34b) contains a relative clause. The external argument is relativized, and verbal direct marking indicates that it is the undergoer. A similar pair is given in (35a) and (35b) with inverse marking. Here, the external argument is the actor, and so, the actor can be relativized.

(34) a. *kay-**a**-poj-a=kinos* *kwe:ya* ***us*** ***dichi:ye***
eat-DR-CAUS-LV=ART.F.AB woman ART.M child
'The/a woman fed the/a child.'

b. *pa:kuk* ***us*** ***dichi:ye*** *di'* *kay-**a**-poj-a=kinos* *kwe:ya*
intelligent ART.M child REL eat-DR-CAUS-LV=ART.F.AB woman
'The boy that the/a woman fed is intelligent.'

(35) a. *kay-poj-**kay**-a=us* *dichi:ye* ***kinos*** ***kwe:ya***
eat-CAUS-INV-LV=ART.M child ART.F.AB woman
'The/a woman fed the/a child.'

b. *pa:kuk* ***kinos*** ***kwe:ya*** *di'* *kay-poj-**kay**-a=us* *dichi:ye*
intelligent ART.F.AB woman REL eat-CAUS-INV-LV=ART.M child
'The woman who fed the/a boy is intelligent.'

The Tagalog voice affixes, therefore, resemble the Movima direct and inverse affixes in that they allow relativization independently of semantic role. The Tagalog UV-infix parallels the Movima direct affix, which indicates that the external argument is the undergoer, and the AV-infix parallels the Movima inverse suffix, which indicates that the external argument is the actor.

The differences between the Tagalog and Movima system are gradual. One difference is that in Movima, definiteness does not seem to play a role: both the internal and the external argument can receive a definite or indefinite interpretation (see 6.2.4). Another difference is that in Movima, the choice of the direct or inverse voice is influenced by a person or animacy hierarchy, which is not the case in the Tagalog voice alternation. In Tagalog, a scenario involving the first and a third person can be described with either the undergoer voice or the actor voice, as shown by (36a) and (36b), respectively. In a Movima description of this kind of scenario, only the direct construction (corresponding to the Tagalog undergoer voice) can be used, with the first person encoded as the internal argument; see (37a). An inverse construction expressing the same state of affairs, tentatively exemplified in (37b), would be impossible because the first person cannot be expressed as the external argument of a basic transitive clause (recall Figure 18). When the internal argument of an inverse clause denotes an inanimate entity, the external argument (the actor) can only be another inanimate entity, but cannot outrank the internal argument in the referential hierarchy (see Chapter 5).

(36) Tagalog
a. *B<in>asag* *ko* *ang* *baso*
<UV>break 1SG.GEN NOM glass
'I broke the glass.'
b. *Nag-basag* *ako* *ng* *baso*
AV-break 1SG.NOM GEN glass
'I broke a glass.'
(Haude & Zúñiga 2016, adapted from Schachter & Otanes 1972: 299–300)

(37) Movima

a. *bay<a>cho=Ø* *as* *wa:so*
break<DR>=1SG ART.N glass
'I broke the/a glass.' [elicited]

b. *baycho-kay-a=as* *wa:so*
break-INV-LV=3N.AB glass
'The/a glass was broken by (it).' [made up]

In order to overcome the restrictions posed by the hierarchy, Movima has the valency-decreasing operation, which allows the treatment of a high-ranking actor as a privileged argument. In an event description with a human actor and an inanimate undergoer, the actor can only be relativized if this operation is applied, as in (38). (Technically, this valency-decreasing construction is also possible with the inverse, but I have never encountered this in the corpus; this is probably because undergoers that outrank actors in the hierarchy are rare, and because the direct construction can also be used in this case; see 5.1.)

(38) ***is*** ***buka'*** ***itila:kwa*** *di'* ***kaw*** *joy-**a:**-łe*
ART.PL DUR.MOV man REL VALDECR go-DR-CO
***n-**is* *buka'* *ke:so=is*
OBL-ART.PL DUR.MOV cheese=3PL.AB
'the men who were carrying their cheese'
(Haude & Zúñiga 2016: 470)

The valency-decreasing operation is also often used in relative clauses as an equivalent to the inverse when the two participants are equally ranked, as in (39) (see also (22)). Exploring the conditions for when a speaker chooses this operation – which functions as an antipassive – over the inverse is work in progress, on which I reported at a workshop on antipassives at the ALT conference in Naples (September 2016).

(39) *pa:kuk* ***kinos*** ***kwe:ya*** *di'* ***kwey*** *kay-a-poj* *n-us*
intelligent ART.F.AB woman REL VALDECR eat-DR-CAUS OBL-ART.M
dichi:ye
child
'The woman who fed the/a boy is intelligent.'

The differences between the Movima and the Tagalog systems, therefore, have to do with the discourse status (as reflected by the definiteness of the Tagalog *ang*-phrase) of the arguments and of their position in a referential hierarchy. In general terms, however, the systems are largely parallel. Acknowledging the parallels between Tagalog and Movima helps enormously for the analysis of Movima alignment. Leaning on accounts of Tagalog, especially those by Himmelmann (1991; 2005a; 2008) and Kaufman (2009), makes it possible to describe Movima as a system with one central grammatical relation which is assigned according to its referential rather than role properties. Moreover, the parallels do not stop at the existence of two transitive constructions; there are many more, which seem to be interestingly related, and to which I would like to get back in future research (see 8).

Nevertheless, language description means that one cannot stop when one has managed to fit a pattern nicely into a typologically established slot. One has to go further to see what else there is in the language, even if the findings disturb the argument one would like to defend. Therefore, in the following section I present some of my findings regarding the intransitive forms of Movima bivalent verbs, which suggest that this language shows a bias towards ergativity after all.

4.1.5 Undergoer orientation as a sign of underlying ergativity?

Haude, Katharina. 2012b. Undergoer orientation in Movima, in: Authier, Gilles and Katharina Haude (eds.). *Ergativity, Valency and Voice*, 259-288. Berlin/New York: Mouton de Gruyter.

Apart from the direct-inverse alternation of transitive predicates, there are other voice alternations in Movima, which more closely correspond to traditional definitions of voice (see Kulikov 2011) in that semantically bivalent verbs function as intransitive predicates. However, unlike passive or antipassive, there are reasons not to consider these voices as detransitivizing. As I will explain below, Movima bivalent verb bases can rather be considered inherently intransitive and undergoer oriented. Voice morphemes indicate the semantic role (actor/undergoer) of the single argument, and direct or inverse marking additionally increase the transitivity of the verb. In this way, Movima is an extreme example of a "chiefly transitivizing" language (Nichols, Peterson & Barnes 2004). The reasoning is as follows, and

shows once more that profound knowledge of a language is required in order to allow for its typological assessment.

Most semantically bivalent Movima verb roots, including those that denote typical two-participant events like 'kill', 'kick', or 'hit', belong to a set of 150 roots that cannot occur independently (Haude 2006: 555–557). These roots must take an affix, usually one of those listed in Table 10 (I consider reduplication a case of phonologically reduced affixation; see Haude (2014b).

Table 10. Voice affixes and the role of the external argument

affix	function	role of external argument
-chełˀ	reflexive/reciprocal (R/R)	actor + undergoer
-ełe	agentive (AGT)	actor
<CV~>	middle (MD)	potentially affected actor
-ˀi	resultative (RES)	undergoer of state resulting from externally induced event
-a-/-na	direct (DR)	undergoer of transitive verb
-kay	inverse (INV)	actor of transitive verb

The effect of these markers is illustrated in (40) with the bivalent root *jat-* 'hit' (recall that *=Ø* marks the first person singular internal argument of a transitive verb; for simplicity of presentation, elsewhere only third-person forms are given in the translations, but the unexpressed arguments can just as well be first or second persons.)

(40) reflexive: *jat-**chełˀ*** 'X hit themselves/each other'
agentive: *jat-**ełe*** 'X hits (continuously/habitually)'
resultative: *jat-**ˀi*** 'X has been hit.'
direct: *jat-**na**=Ø* 'I hit X.'
inverse: *jat-**kay**=Ø* 'X hits me.'

The element that most interested me here was the ending *-ˀi,* which marks the resultative voice, i.e. it derives a verb denoting a static situation that has been brought about by an external actor. The single argument of such a verb is the undergoer, as shown by (41).

(41) *ben-**’i*** *is* *chorimpa=sne*
paint-RES ART.PL fingernail=3F.AB
‘Her fingernails were painted.’ [Asilo 077]

In Haude (2012b) I have shown that verbs combined with with *-’i* can be considered less marked than verbs combined with one of the other morphemes. This is because the element /ʔi/ acts as a phonological dummy in other environments. For instance, an ending *-’i* occurs on some prosodically deficient nouns and pronouns, without adding any meaning. The monosyllabic noun root *nun-* is augmented with *-’i* to occur as an independent noun: *nun-’i* ‘bone’. When the noun forms part of a compound, however, it appears as *nun-* alone (see also Haude 2006: 207), as in (42).

(42) *punta:-**nun***
tip-bone
‘the bone from the tip (of the rib cage)’

Support for the idea that the resultative marker *-’i* originates from a phonological dummy that is also found in other environments comes from the fact that verb bases that can act as full prosodic words are unmarked when denoting a resultative state. An example is *rimłe* ‘sell’ in (43): When not combined with any other voice marker, this verb means ‘to be sold’; the suffixation with *-‘i* is ungrammatical.

(43) a. *ba:ra* *rimłe*
all sell
‘It is all sold.’/ ‘It has all been sold.’ [elicited]

b. * *ba:ra* *rimłe-**’i***
all sell-RES
(‘It has all been sold.’) [elicited]

If the resultative form of a bivalent verb is the least marked form, this has interesting consequences for the analysis of bivalent verbal bases. First of all, it means that bivalent verb bases have only one argument in their argument structure, i.e. that they are basically

intransitive. Secondly, it means that an unmarked bivalent verb is undergoer-oriented, i.e. that the single argument of the verb is the undergoer.

The addition of one of the other monovalent voice suffixes (reflexive, middle, agentive), then, changes the semantic role of the single argument: the reflexive/reciprocal suffix indicates that this argument is actor and undergoer at the same time; and the middle reduplication indicates that the single argument is the actor that is at the same time affected by the event; the agentive marker indicates that the single argument is the actor in an event affecting an unknown undergoer (see (34).

To derive a transitive verb, a direct or inverse marker must be added. These elements license an internal argument that is absent from the basic argument structure of the verb, and specify the semantic roles of the two arguments.

Now comes the interesting part, which links these observations to the discussion of aligment above: if the unmarked verb base has a resultative meaning, then the direct marker is a basic transitive derivation than the inverse. The direct marker does not change the undergoer orientation of the verb: it only adds a slot for an internally cliticized element expressing the actor. Compare (44) with the intransitive, resultative clause in (41) above. Both clauses express that someone's fingernails have been painted, and in both, the external argument phrase denotes the undergoer. The only difference is that in (44), the verb is marked as direct and an internal argument is added, encoding the actor.

(44) *ben-na=is* *kis* *chorimpa=is*
paint-DR=3PL.AB ART.PL.AB fingernail=3PL.AB
'They painted their fingernails (of the old ladies).' [Asilo 071]

The inverse marker is more complex. Like the direct marker it adds an internal argument. However, here the external argument encodes the actor and the added, internal argument encodes the undergoer. The inverse, therefore, can be considered as not being one in a row of several voice affixes, as it is presented in Table 10, but as being based on the transitive, direct form of the verb, whose role distributions it turns around. The inverse is only needed when the argument distribution of the direct voice does not match the referential hierarchy.

The effect of the direct and the inverse markers on the argument structure of semantically bivalent verbs is schematized in Figure 20, with the additions contributed by these markers highlighted in red.

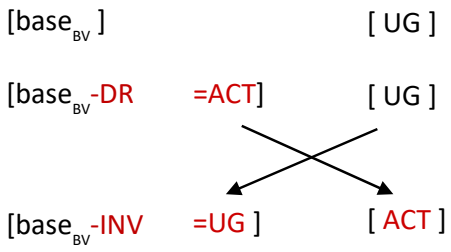


Figure 20. The effect of direct and inverse marking on the argument structure of semantically bivalent verbs

Thus, despite the parallel pattern of direct and inverse voice that favours an analysis as a symmetrical voice system, this is a sign that the ergative (direct) pattern is more basic than the inverse (accusative) one in Movima. In Chapter 5 I will show that also on the discourse level, the direct construction can be considered the more basic transitive construction.

4.1.6 Summary: the three alignment types in Movima

To sum up, Movima combines properties of three morphosyntactic systems that are often treated as separate in linguistic typology: an ergative system, an inverse systems, and a symmetrical-voice system. Table 11 lists the three types, the languages that most famously represent them, and the properties of Movima that favour and disfavour its classification as one of these types.

Table 11. Three alignment types present in Movima

	ERGATIVE	**INVERSE**	**SYMMETRICAL VOICE**
representative language	Dyirbal	Plains Cree	Tagalog
fitting property	DR as more basic	obviation + inverse	two transitive constructions
deviating property in Movima	additional INV transitive construction not morphologically derived from DR	no person affixes; direct=ergative	person hierarchy, valency-decreasing operation

What might be the underlying factor that accounts for the coexistence of three alignment types, all of which are relatively uncommon in the languages of the world? In Section 4.2 I present my hypothesis, inspired by the typological literature on ergative and symmetrical-voice systems, that the transitive constructions in Movima are underlyingly intransitive. This hypothesis, while requiring further research, seems to solve many of the typological puzzles found in this language.

4.2 Understanding the puzzle: basic intransitivity

Haude, Katharina. 2009b. Reference and predication in Movima, in: Epps, Patience and Alexandre Arkhipov (eds.). *New Challenges in Typology: Transcending the Borders and Refining the Distinctions*, 323-342. Berlin: Mouton de Gruyter.

Haude, Katharina. 2010a. The intransitive basis of Movima clause structure, in: Gildea, Spike and Francesc Queixalós (eds.). *Ergativity in Amazonia*, 285-315. Amsterdam/Philadelphia: Benjamins.

Ergative and symmetrical-voice systems pose related challenges to alignment typology (see e.g. Dryer 1997). In both systems, one argument of the transitive construction shares properties with the single argument of an intransitive construction, and these shared properties identify the argument as a single grammatical relation, both in the transitive and in the intransitive clause. However, this grammatical relation cannot be easily defined as a "subject" in these systems, since they do not share the semantic (and possibly also discourse-pragmatic) properties of a canonical subject: in an ergative system, the privileged grammatical relation bears the undergoer role, as it does in the undergoer voice of a symmetrical-voice system.

Because of the ergative pattern of the undergoer voice in symmetrical-voice systems, which even shows traits of being more basic than the actor voice (Nagaya 2012), symmetrical-voice systems have often been analyzed as ergative (see e.g. Payne 1982; Aldridge 2012). The relationship between symmetrical voice and at least one type of ergative system (in which, for instance, the ergative argument is encoded like a possessor) is sometimes explained as resulting from similar historical developments (see e.g. Sasse 2009).

One attempt to account for symmetrical voice systems can be summed up with the term "nominalist hypothesis" (Kaufman 2009). Under this hypothesis it is assumed that in symmetrical-voice systems, sentences expressing two-participant events are basically

intransitive, headed by participle-like predicate nominal. In Tagalogist terminology, the actor-voice affix would thus derive a noun denoting the actor (like *-er* in English *employ-er*) and the undergoer voice would derive a noun denoting the undergoer (like *-ee* in English *employ-ee*). Under this analysis, the Tagalog sentence in (36a) above would be paraphrased as 'The glass is my broken one', and the one in (36b) as 'I was the breaker of a glass.' As can be seen from these English paraphrases, one central component of this analysis is the possessor-like encoding of the non-argument. This is indeed the case in the languages concerned, as indicated by the glosses of the Tagalog examples, where the non-nominative argument is glossed as 'genitive'.

When I first got acquainted with studies proposing this analysis, notably Himmelmann (1991) and Sasse (1991), I was thrilled. Considering a Movima basic clause as an intransitive sentence with a nominal predicate made everything fall in place. In what follows, I will describe the different building blocks that allow for such an analysis: possessor-like encoding of one of the arguments (4.2.1) and the syntactic flexibility of nouns and verbs (4.2.2). Section 4.2.3 sums up how these phenomena permit a potential "all-intransitive" analysis of Movima basic clauses.

4.2.1 The internal argument as possessor

In Movima, an adnominal possessor is encoded by internal cliticization, exactly like the internal argument of a transitive clause (see 4.1.1). This is illustrated in the examples of possessed noun phrases below: in (45), note the appearance of the linking vowel, which also receives stress shift due to its penultimate position, exactly like the transitive verb in (46) (repeated from (3) above). Also the possessive pronoun is the same as that of the argument paradigm, as can be seen from the masculine pronoun *=us*, which also occurs in (46).

(45) *kinos* *alwaj-á=us*
ART.F.AB spouse-LV=3M.AB
'his wife'

(46) *kay-a-poj-á=us*
eat-DR-CAUS-LV=3M.AB
[kajapoˈh**aʔ**us]
'He feeds him/her/it/them.'

4.2.2 Lexical flexibility: nominal predicates and "verbal RPs"

Haude, Katharina. 2019c. From predication to reference: on "verbal DPs" in Movima. In *Individuation et référence nominale à travers les langues*. Bonnot, Christine, Outi Duvallon et Hélène de Penanros (eds), 53-77. Paris: Editions Lambert-Lucas. <https://halshs.archives-ouvertes.fr/hal-01971969v1>

Another component that a language needs to permit an all-intransitive analysis is the syntactic flexibility of content words. Nouns must be able to function as predicates, and verbs must be able to occur in referring expressions so that the encoding of an argument and a possessor (or vice versa) is indistinct. The predicative use of nouns is relatively common crosslinguistically, especially in languages that can form equational clauses without a copula. Movima is one of them. The argument of a nominal predicate is encoded in the same way as in an intransitive verbal clause. Example (47) illustrates a nominal predicate with a pronominal argument, (48) a nominal predicate whose argument is expressed by an RP, and (49) shows a nominal predicate whose argument is not overtly expressed.

(47) ***tolkosya****--'ne*
girl--3F
'She (is a) girl.' [Dial. EA&AH 012]

(48) *pe'łełe* ***itila:kwa*** *kis* *majniwa=sne*
all man ART.PL.AB offspring_of=3F.AB
'Her children are all men.' [EAO Ay'ku II 027]

(49) ***rulrul*** *jayna*
jaguar DSC
'It already (was a) jaguar (after transformation).' [LYO_250808_2 192]

The occurrence of verbs in a referential expression is less common cross-linguistically. It can be argued that English allows many words to occur equally well as a predicate and inside an NP (e.g. *to kiss – the kiss*). However, this is not freely possible for just any word, and in many

cases the syntactic position goes along with a different meaning (e.g. *the chair – to chair*). Therefore, these pairs are often analyzed in terms of zero derivation, conversion, or separate lexical entries rather than as single syntactically flexible lexical items.

In Movima, this is different. Any verb can occur in a noun phrase (which, therefore, I label "referential phrase", short RP), and the verb's meaning in this position is predictable: the RP containing a verb corresponds to the external argument of the verb in predicate function. (Exceptions are verbs with incorporated arguments; see Haude 2009b.) This is to say, when the verb inside the RP is intransitive, as in (50), the RP refers to the single participant of the state or event denoted by the verb, shown by (50b); when the verb inside the RP is transitive and marked as direct, as in (51), the referent is the undergoer, shown by (51b); and when the verb inside the RP is marked as inverse, as in (52), the referent of the RP is the actor, shown by (52b). Each of the b examples compares the "verbal RP" with the corresponding verbal clause in the a examples.

(50) Main clause vs. RP with intransitive verb

a. ***ji<wa:~>wa***--*us*
come<MD~>--3M.AB
'He came.'

b. *us* ***ji<wa:~>wa***
ART.M come<MD~>
'the/a (man/boy) who came; the/a (male) comer'

(51) Main clause vs. RP with transitive direct verb

a. ***vel-na=us***--*k-is*
watch-DR=3M.AB--OBV-3PL.AB
'He watched them.'

b. *kis* ***vel-na=us***
ART.PL.AB watch-DR=3M.AB
'the (ones) he watched; his watched (ones)'

(52) Main clause and RP with transitive inverse verb

a. ***vel-kay-a=us***--*k-is*
watch-INV-LV=3M.AB--OBV-3PL.AB
'They watched him.'

b. *kis* ***vel-kay-a=us***
ART.PL.AB watch-INV-LV=3M.AB
'the (ones) who watched him; his watchers'

Thus, in Movima, a predicate becomes part of a referring expression simply by having a determiner placed in front of it. This is true both for verbal and for nominal predicates, since, as was shown in (47)–(49), nouns not preceded by a determiner function as main-clause predicates. Both nouns and verbs acquire their referring potential only when preceded by a determiner; this is schematized in Figure 21. The figure shows that a content word (N/V) can occur in a predicate phrase (labelled this way because 'verb phrase/VP' would be misleading) on the clause level, but also in a predicate phrase inside an RP. In both positions, a bivalent content word (i.e. a transitive verb or a possessed noun) is followed by an internally encliticized element, which is interpreted as the possessor of a nominal form, as the A argument of a direct-marked verb, or as the P argument of an inverse-marked verb.

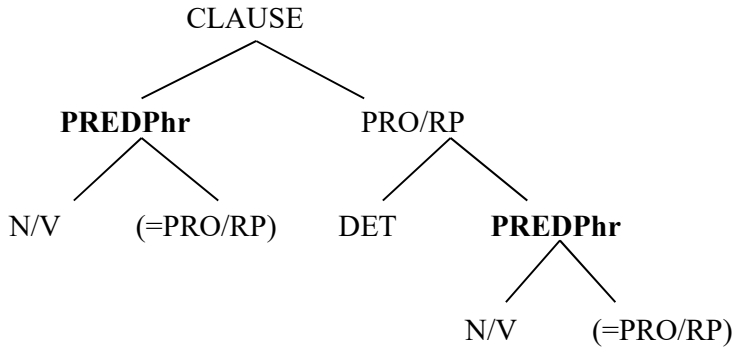


Figure 21. Hierarchical representation of the Movima clause (abbreviations: DET = determiner; N/V = noun or verb; PREDPhr = predicate phrase; PRO = bound pronoun)

The syntactic similarity between nouns and verbs inside RPs in Movima goes even further than the flexible exchange between the two syntactic positions (on whose pragmatic function I will report in Section 6.2.4). Also in RPs, the valency-decreasing operation (4.1.2, example (22)) can apply, and nouns and verbs can likewise undergo this operation. A verb preceded by the valency-decreasing particle *kwey/kaw*, as in (53b), becomes intransitive, and its former internal argument (A, i.e. *=us* in (53a) becomes the single argument (S). In this way, the actor

becomes the referent of an RP. The former external argument (P) is demoted to adjunct status, i.e., the undergoer is expressed by an oblique-marked RP if expressed at all.

(53) a. *vel-na=us* *is* *waki:ya*
watch-DR=3M.AB ART.PL calf
'He watched (the) calves; (the) calves were his watched ones.'

b. *us* ***kaw*** *vel-na* ***n-****is* *waki:ya*
ART.M VALDECR watch-DR OBL-ART.PL calf
'the (one who) looked after (the) calves; the watcher of (the) calves'
[Ganado 061]

The effect of the valency-decreasing operation on a noun is shown in (54). Example (54a) shows a "normal" RP, i.e. an RP containing a noun with an enclitiziced possessive pronoun. Inserting the valency-decreasing particle before the noun, as in (54b), makes it impossible to encliticize a possessor to it. In the same way in which the participant encoded as the internal argument of a transitive verb becomes the referent of a valency-decreased verbal RP (see (53)), the entity normally encoded as a possessor of a noun is now the referent of the nominal RP, while the possessee is encoded as oblique (if at all).

(54) a. *is* *nono=us*
ART.PL domestic_animal=3M.AB
'his animals' [GBM Ganado 040]

b. *os* ***kaw*** *rey* *no:no* ***n-****isko*
ART.N.PST VALDECR EPIST domestic_animal OBL-PRO.3PL.AB
an owner of the animals (lit.: "an animal owner of them")
[GBM Ganado 004]

The syntactic flexibility of nouns and verbs is another reminiscence of Tagalog, alongside with the symmetrical transitive constructions. In (55a), we see an *ang*-phrase with a verb in the actor voice, which refers to the actor of the 'buying' event denoted by the verb root, and in (55b) we see an *ang*-phrase with a verb in the patient voice (analyzed as a zero suffix by the author), which refers to the patient of the 'buying' event. (The Tagalog voice affixes also

convey mood and aspect values, which leads to controversial analyses as to the distribution of the different coding features.)

(55) Tagalog (Kaufman 2009: 5)

a. *ang=b<um>ili*
NOM=<AV:BEG>buy
'the buyer/one who bought'

b. *ang=b<in>ili-Ø*
NOM=<BEG>buy-PV
'the (thing) bought'

The lexical flexibility is even stronger in Movima than in Tagalog. While in Movima, the referent of the verbal RP always corresponds to the external argument of the verb, in Tagalog, the interpretation is not always that straightforward: the referent can also be the event denoted by the verb (Himmelmann 2008). Furthermore, while the valency-decreasing operation provides strong evidence for the noun-verb analogy in Movima, there is no such operation in Tagalog. Finally, in Tagalog, the predicative vs. referential use of a lexical item is sometimes distinguished by its stress pattern (Himmelmann 1991), but this is not the case in Movima.

Some would argue that the placement of a content word inside an RP is a nominalization (see e.g. the account of Tagalog by Comrie & Thompson 2007: 337), and that a verb in a nominal position is "functioning" as a noun, "i.e. as an argument to another V" (Hopper & Thompson 1984: 737). Shibatani (2019), too, sees this function as a nominalization, following Vendler (1967):

> The device of nominalization transforms a sentence into a noun phrase, which can then be inserted into another sentence; it is a means of packing a sentence into a bundle that fits into other sentences. (Vendler 1967: 125)

This makes sense, but I have problems with this account. If a verb is considered "nominalized" just because it occurs inside an RP, why should a noun, which also functions as predicate in clause-initial position, not be considered "nominalized" as well when occurring inside an RP? Conversely, shouldn't a noun be considered "verbalized" when it occurs as a predicate, as in (47)–(49), since predicates are prototypically verbs?

The question becomes especially tricky when words cannot straightforwardly be attributed to a lexical class. In Movima, many content words can be identified as either nouns or verbs. For instance, a word that can receive an internal clitic without containing a direct or inverse marker is a possessed noun. However, this is not as straightforward as in languages where nouns are marked for case, number, or gender, and verbs are marked for tense, aspect, mood, evidentiality and the like. In Movima, these categories are not systematically marked, and some morphemes are even shared by nouns and verbs (see Haude 2006: 106–111). Also, intransitive verbs – which do not contain a direct or inverse marker – are often difficult to distinguish from common, unpossessed nouns. In short, it is not always easy to class a Movima content word as either noun or verb. Therefore, if one speaks of a syntactic nominalization involving the transformation of a predicate (or "sentence", in Vendler's terms) into a noun phrase, this notion should be applicable without having to make complicated tests regarding the lexical category of the predicate.

Therefore, I prefer to consider nominalization a lexical process, separate from what might better be called "referentialization" (Alvarez González 2012). Any Movima content word is a predicate when occurring independently (i.e., without a preceding referential element), and becomes referential by being placed inside an RP. At the same time, in line with Verndler's analysis, the position behind a determiner can be considered an embedded position, as is shown in Figure 21 above.

4.2.3 Basic intransitivity in a typological perspective

If it had not been for Werner Drossard, my former Linguistics teacher in Cologne, I do not know when (and if) I would ever have been put on the track of the works by Hans-Jürgen Sasse and Nikolaus Himmelmann that makes so much sense for analyzing the syntactic system of Movima.

Sasse's paper "Predication and Sentence constitution in Universal Perspective" (Sasse 1991) was an eye-opener for me. In this article, Sasse sets up a typology of languages according to the "sentence-constituting operations" they display and according to the way in which "actant expressions" are syntactically related to a "state of affairs expression". In one of these types, which corresponds to the bipartite Aristotelian sentence, an actant expression, which represents the "predication base" (i.e., the subject), is combined with a "*characterizing* state of affairs expression" (Sasse 1991: 85; emphasis mine). This type is represented by a "nominal sentence, with or without a copula, where ... the non-referential predicate noun [is] the state-of-affairs expression characterizing the subject" (Sasse 1991: 85). Sasse illustrates this idea with

examples from Tagalog and Mayan, arguing that the Tagalog sentence meaning “the woman bought the rice” has the structure “buyer of the rice (is) the woman” and that the Mayan-type version of the English sentence “I am seeing you” is “You are my seen one”.

In a correspondence that followed after I read this article, Sasse told me that afterwards, he was not so happy with the paper because he found he had too much simplified matters in order to transmit his idea. Be this as it may, the paper got me thrilled, because all of a sudden it made me understand what is going on in Movima. What if Movima content words, whether nouns or verbs, are *characterizing* expressions, attributing a characteristic to an actant (i.e. a participant in a state or event) rather than expressing the state or event in which the actant is involved? This would explain the exchangeability of verbs and nouns in either predicate function (“be a watcher/watched one”) and the function in a referential phrase (“the/a watcher/watched one”, “the one who watches/is watched”). The predicate, whether noun or verb, ascribes a property to the referent of the argument, and the argument refers to an entity having this property.

The status of the internal argument, which does not have syntactic privileges, can also be explained in this way. Similar to the Mayan languages Sasse describes, possessed nouns are marked for the possessor in the same way as transitive verbs are marked for their actor (if direct) or undergoer (if inverse). Consequently, when combined with a verb, this enclitic forms a phrase together with the verb; in Sasse’s words regarding Mayan, also basic Movima verbal clauses (like (2)–(9)) can be described as follows: “The entire verbal phrase with the possessively bound actor serves as the predicate of the ascriptive sentence” (Sasse 1991: 89).

The problem with this analysis, and the reason why it is still often met with scepticism, is that the linguistic facts are usually not quite that neat. With respect to Movima, for instance, possessed nouns cannot function as predicates with a pronominal enclitic, which means that the syntactic flexibility is not complete (see 6.1.2). This is probably because they are referential. In any case, the gist of the idea is as follows: If verbs and nouns behave syntactically alike, and one verbal argument is encoded like a possessor – i.e. not like a syntactic argument at all – then there is no reason *not* to assume an equational reading everywhere, i.e. not only with nominal, but also with verbal predicates.

The syntactic flexibility of Movima verbs and nouns and its relationship with the alignment system is such a central factor of Movima grammar that I have discussed it in several publications. (Not all of them have been published; for instance, a paper with Spike Gildea that will take a more historical perspective has never come to conclusion.) It is also one of the topics that I intend to deal with in the future (see 8.2), since I believe that it should become more central in morphosyntactic typology. I would like to propagate a linguistic type that is based

on intransitivity, even if its synchronic surface patterns appear as transitive clauses with a symmetrical-voice or split-ergative alignment pattern.

5. Referential hierarchies and prominence

The following sections describe my efforts in determining the relative impact of semantic and (discourse-)referential factors on Movima third-person argument encoding (5.1), which are very well captured by the notion of "salience" or "prominence" (5.2). In Movima, there is a very interesting mismatch between the factors that govern argument encoding and the syntactic properties of the arguments (5.3).

5.1 Agentivity, animacy, topicality

Haude, Katharina. 2014a. Animacy and inverse voice in Movima: a corpus study. *Anthropological Linguistics* 56(3-4): 294-314.

Aside from the study of alignment, which involves the comparison of transitive and intransitive clauses, one of my central research interests since 2009 has involved the ways in which the referential properties of an argument expression influence the choice of either the direct or the inverse transitive construction in Movima. I did so, for instance, in the context of the project *Referential Hierarchies in Morphosyntax* (see 2.1.2).

The hypothesis that formed the starting point of the RHIM project was that languages reflect, in one way or another, a hierarchical order of the entities that are referred to during communication. "Referential hierarchy" is one of the labels for this scale, on which speech-act participants rank higher than third persons, animate entities higher than inanimate ones, and familiar entities higher than newly introduced ones. A rough hierarchy of this type is given in (56).

(56) SAP > human/animate/familiar > non-human/inanimate/newly introduced

This hierarchy may find its reflex, for example, in the use of a pronoun instead of a full noun phrase for a higher-ranking referent, or by the choice of a passive over an active construction

when a higher-ranking referent is the patient in a two-participant event. It is assumed that referential hierarchies play a key role in language processing and so they might reveal possibly universal aspects of human cognition. Movima is a particularly interesting case, since here, the referential properties of event participants cannot be ignored when investigating grammatical relations and alignment. My investigation within this project, therefore, concentrated on the role of the different referential properties of argument expressions in shaping a Movima transitive clause.

Under "referential properties" I understand the properties a discourse referent has in the minds of speaker and hearer. They include the inherent properties of the referent, such as its animacy or sex, which can be termed "ontological properties". Other referential properties are discourse-internal, including, for instance, the degree of givenness, i.e. the familiarity from the previous discourse. These are frequently referred to as "discourse status". Since in discourse, a referent is usually presented as partaking in some situation (event or state), the role it has in this situation – e.g. that of being an actor – is also part of its properties at the moment at which it is being referred to. These inherent, discourse-based, and semantic referential properties are independent from each other, but together they contribute to the choice of a particular syntactic construction: The choice of a syntactic construction is influenced by the identity of the referent, its familiarity from the preceding discourse, and the role it has in the situation that is described by the construction. For instance, in German or English, the passive is used more frequently when the patient is animate (see van Nice & Dietrich 2003: 828 for German) and/or discourse-given in opposition to a discourse-new agent. I subsume the influence of the different properties of a referent on a synactic construction under the term "reference effects".

The grammatical system of Movima with its direct-inverse opposition in transitive clauses represents a particularly interesting case for the investigation of reference effects on clause structure. It is comparable to the active-passive alternation in English or German. The most important functional difference between the Movima inverse voice, on the one hand, and the passive in more familiar languages, on the other hand, is that in Movima, the choice is obligatory according to person values: When a person higher in the 1> 2> 3 hierarchy is the undergoer, the direct construction, i.e. the equivalent of an active phrase like *he struck me*, is excluded by grammar; to describe such a situation, one must use the inverse, which is functionally comparable to the passive: *I was struck by him*. While the use of the inverse in the domain of the hierarchy of persons (1> 2> 3) is prescribed by the grammar, this is not the case with two third-person arguments: Here, the direct construction ('the girl hit the boy') can just as well be used as the inverse equivalent ('the boy was hit by the girl'). In the same way as

with the active-passive alternation, the question that arises here is: Which construction is used when?

In Haude (2014a) I approached this question by postulating several categories that I hypothesized are relevant for the discourse-semantic prominence of a referent, and by tagging my corpus data for these categories. On the agentivity scale, these categories are actor and undergoer. On the animacy scale, I distinguished between human (coded as "hum"), non-human animate (coded as "$anim" in order to exclude hits for "inanim"), and inanimate ("inanim"). On the discourse scale, I distinguished between arguments expressed by pronouns, RPs, and (in the case of the external argument, where this is possible; see 4.1.1) zero expression. In order to measure the discourse status of the arguments, I simply assumed that the referent of a pronoun ranks higher than the referent of a noun phrase: In a binary opposition, a pronoun is used to encode an entity that is known from the context, while an NP is appropriate for referring to new, or less familiar, discourse participants.

Before starting the annotation, I formed a subcorpus that only contained transitive clauses describing an interaction between two third persons. This resulted in approximately 1,260 Toolbox records of this type (the number has increased since the publication, but the overall results of the study are the same), which I then annotated for the different categories, as depicted in Table 12. After this, I re-inserted the records into my "Toolbox Masterfile" that combines all Movima texts I transcribed and that I use for most of my analyses.

Table 12. Annotation categories for basic transitive clauses with 3>3 scenarios

	category	annotation
a.	person scenario	3>3
b.	construction	DR or INV
c.	internal argument expression	=NP or =PRO
d.	external argument expression	#NP or #PRO or #0
e.	animacy scenario	hum/$anim/inanim > hum/$anim/inanim

(Symbols and abbreviations: > 'acts on'; = 'internal argument'; # 'external argument'; NP 'noun phrase'; PRO 'bound pronoun')

A Toolbox record annotated according to this schema is exemplified in Figure 22. The example contains two clauses, an intransitive (*joychełˑ--isne* 'she went') and a transitive one (*way-na=sne os toˑmi* 'she fetched water'), which is followed by an oblique phrase (*n-os siłkwa*

'of/from the waterhole'). The field \trans contains the annotation for transitive clauses according to the categories in Table 12. In this particular example, the field \trans indicates that the record contains a transitive clause with a direct-marked predicate and two third-person arguments, the internal argument being expressed by a bound pronoun and the external argument by an RP (labelled "NP" in the annotations). Furthermore, it is indicated that the transitive clause describes a scenario in which a human participant acts on an inanimate one.

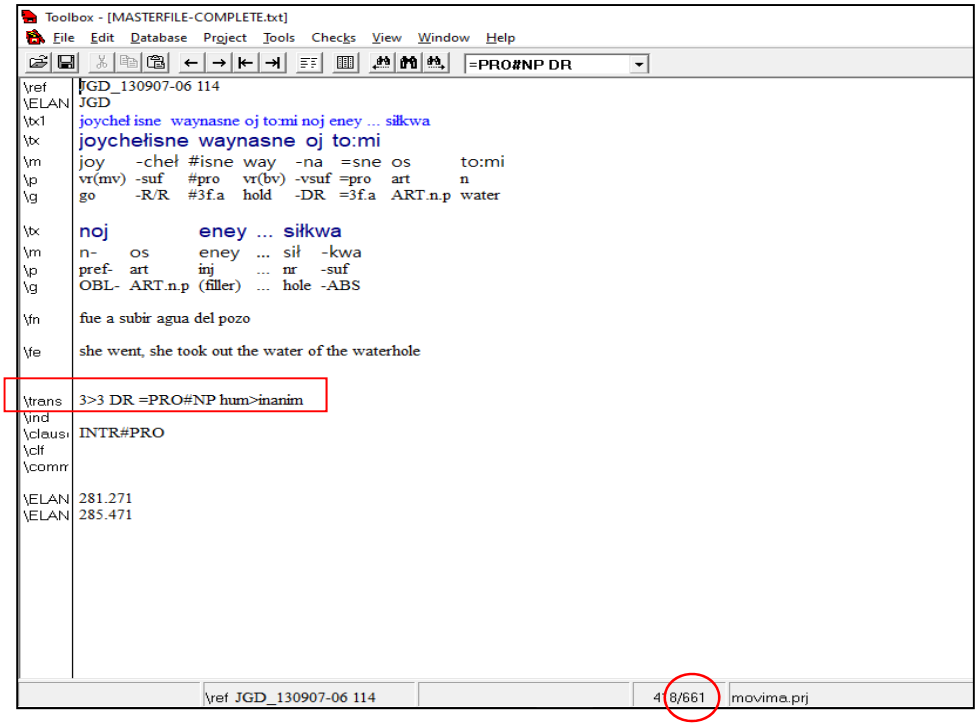


Figure 22. Annotation of basic transitive clauses according to participant scenario and argument expression

Based on the categories of this annotation schema, I built filters into Toolbox that allowed me to make counts for each category. The result of the filtering is shown in the grey line at the bottom of Figure 22: Here, it can be seen that this record is one out of 661 results for the filter "=PRO#NP DR"; i.e., there are 661 clauses with two third-person arguments whose predicate is marked as direct and in which a pronoun codes the internal and an RP the external argument.

With this system, it was easy to confirm, first of all, the impression that the direct construction is the default to express a two-participant event: 94% of the clauses with two third-person participant are direct, i.e. the actor is encoded as the internal argument, while only 6% are inverse, i.e. the undergoer is encoded as the internal argument. From here, the question

follows: Why is the inverse used at all? Which referential properties of the arguments contribute to its use?

First, I looked at the distribution of pronouns vs. RPs. The impression that the internal argument is usually a pronoun was confirmed: This is the case in 93% of the clauses in the database, while in only 7% of the clauses, the internal argument is an RP. If one counts the distribution of RPs and bound pronouns, it turns out that of the 842 RPs occurring in the database, the majority (748, i.e. 89%) occur as external arguments, while only 94 (i.e. 11%) occur as internal arguments. By contrast, only 92 bound pronouns, i.e. 7% of the bound pronouns in the database, represent an external argument. This is graphically represented in Figure 23, based on the numbers given in Haude (2014a). (Here, "PROX" stands for the internal argument and "OBV" for the external argument, the terms that I used in the study described here.)

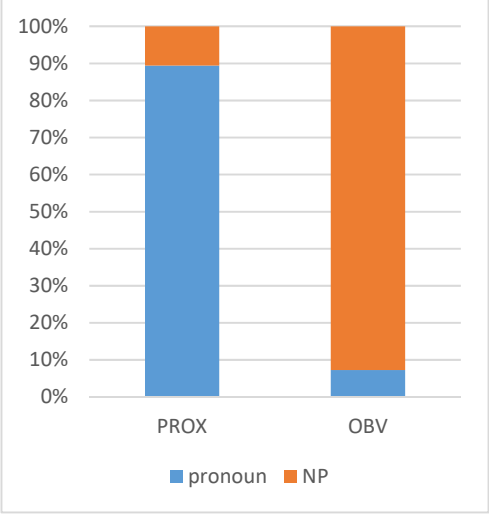


Figure 23. Distribution of pronouns and RPs over the internal (PROX) and external (OBV) argument functions

Discourse topicality, therefore, is highly relevant for argument encoding in Movima. When a discourse-prominent referent is the undergoer in the event, the verb is marked as inverse. Consider (57) for an illustration. Here, the discourse topic is a boy. He is referred to by an RP (*us oveniwankwa*) in the first, intransitive clause (57a), and then taken up by a pronoun (=*us*), which is the internal argument in the following transitive clause (57b); since the boy is the undergoer, the verb is marked as inverse.

(57) a. *jayna* *wele:łe* ***us*** ***oveniwankwa*** *[…]*
DSC climb ART.M young_man

b. *jayna* *julra-kay-a=**us*** *is* *so:te*
DSC win_over-INV-LV=3M.AB ART.PL other_person
'Then the boy climbed up […] (because) he had been outwon by others (who had already collected all the fruits that were on the ground).' [HRR_120808-tigregente 056ff.]

However, in over 30% of all transitive sentences in the database, the external argument is not expressed at all. If one assumes that the external argument encodes the less familiar discourse referent, this is surprising: Zero encoding is expected for a referent that does not require overt mentioning because it is discourse-given. So, could this mean that when a referent is highly discourse-given, it is represented in Movima by a zero-expressed external argument rather than as a pronominal internal argument?

It turns out that there seem to be two kinds of "givenness", which one might refer to as inferrability and protagonism. (I have so far not found explicit mentions of this distinction in the literature, and have not yet pursued this question further.) A closer look at the examples shows that the referent of the unexpressed external argument is usually inferrable from the (linguistic or extralinguistic) context, the meaning of the verb, and/or world knowledge. Consider (58) below. There are two transitive clauses in (58b), which are preceded by an intransitive adverbial clause, (58a). (The adverbial clause contains a nominalized predicate, see 7.2, which triggers possessive argument encoding.)

(58) a. *jayna* *n-os* *su<we>we=as,*
DSC OBL-ART.N.PST near-NMZ.ST=3N.AB

b. *way-**na**=us* *łat,* *man-**na**=us*
take-DR=3M.AB EV shoot-DR=3M.AB
'Then, when it (the jaguar) was near, he took (his gun) and shot (the jaguar).'
[PMP_HRR_etal_210908 077]

The argument of the intransitive adverbial clause (*=as* 'it') in (58a) can be interpreted as referring cataphorically to the unexpressed external argument of the second transitive predicate (*man-na=us* 'he shot (it)'). However, the referent of the unexpressed external argument of the first transitive predicate in (58b) (*way-na=us* 'He took (it)'), which must be a shooting weapon,

was not previously mentioned; it can only be inferred from the larger context – which is a hunting story –, the meaning of the verb, and (culture-specific) world knowledge. Similar examples confirm that when the external argument remains unexpressed, this is because the referent is given through the discourse environment, but that this is different from being a discourse topic or protagonist.

However, it is difficult to pin down what it means to be a "protagonist" without risking circularity. A protagonist could be described as the discourse referent whose perspective is taken, but here as well, it is difficult to find robust and independent criteria. Assessing the factors underlying argument expressions exceed the simplistic concepts I used in my study, such as givenness. (In further research I intend to tackle these issues, see 8.3.2.2.) Anyway, the hypothesis is that discourse topics, impressionistically defined, are encoded as the internal argument.

Apart from that, the internal argument tends to encode the actor, i.e. the direct construction is more common. Indeed, the inverse construction occurs more often in those cases in which the internal argument (i.e. the undergoer) is realized as a pronoun. When the internal argument is represented by an RP, however, the inverse is highly exceptional, occurring in less than 5% of the clauses. I conclude from this that the actor is encoded as the internal argument even if it is less discourse-given than the undergoer.

Referential hierarchies are generally assumed to include animacy, a factor that is easier to test for. Therefore, I also analyzed the role of animacy for the encoding of arguments and, accordingly, in the choice of the direct or inverse construction. I examined the nine possible interactions of the three major animacy categories I had annotated for in the corpus: human, non-human animate, and inanimate, counting how many of them are expressed by direct and how many by inverse constructions. The relative results of the distribution of direct and inverse of these scenarios are given in Figure 24. Below I will first discuss the leftmost section (with high-ranking actors and low-ranking undergoers) and then the rightmost one (with inanimate actors). Scenarios with interacting animates, schematized in the middle section of Figure 24, are discussed last.

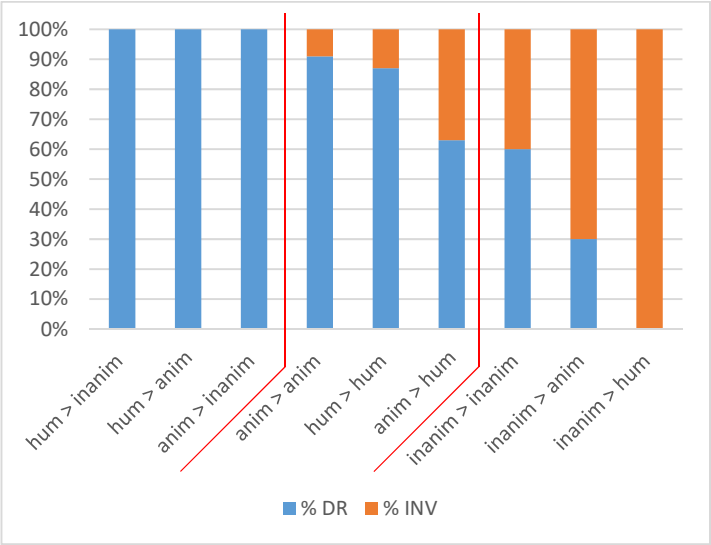


Figure 24. The distribution of direct and inverse constructions according to animacy

High-on-low scenarios, in which an entity higher on the animacy scale acts on a lower-ranking one, are the most common ones in the corpus (60%). Over 3/4 (i.e. 76%) of these contain human actors, mostly acting on inanimate entities, while the remaining 24% involve animate acting on inanimate entities. In this enormous segment of the corpus, the inverse construction is not found at all; in other words, we never find an actor lower in animacy that is encoded as the internal argument. This is independent of the expression of the arguments as either pronoun or RP: All constellations of argument expressions occur, in a ratio similar to that for the overall corpus and all scenarios (see Figure 23).

When individual examples are considered in context, it can be seen that animacy generally outranks topicality. There are many examples in which the discourse topic is an animal, but as soon as it is acted upon by a human, the direct construction privileging the human is used (in contrast to the passive I just used spontaneously in this sentence; in Movima one would probably have said: "the discourse topic is an animal, but as soon as a human acts upon it …"). For instance, (59) stems from a story about the cow herd of the Movima community; the cows, mentioned in (59a), might therefore be considered prominent in the discourse. Still, in the transitive clause in (59b), the human actors are encoded as the internal argument (=*i*) of the direct-marked predicate, while the cattle is encoded as the external argument (--*k*-*i'*).

(59) a. *asko* *jayna* *bijaw-ni-na=i,* ***kis*** ***ney*** ***wa:ka*** *jayna,*
PRO.3N.AB DSC old-PRC-NMZ.LOC=3PL ART.PL.AB here cow DSC

b. *wa:* *rey* *di'* *jayna* *botra-na****=i--k-i'***
even_if EPIST HYP DSC replace-DR=3PL--OBV-3PL
'This is where they$_i$ were raised, these cows$_i$, even if they$_j$ (i.e., its actual keepers) replaced it.' [ATL_230806 288f.]

Thus, animacy plays a crucial role in the choice of a transitive construction: When the actor outranks the undergoer on the animacy scale, one simply chooses the direct voice.

The other important result of this study, represented in the right-hand segment of Figure 24, was that scenarios involving inanimate actors are more likely to be expressed with an inverse construction. Naturally, the number of this kind of examples is extremely low: there are only 23 of them (i.e. 2% of the entire analyzed corpus), but still, the result is telling. In particular, when an inanimate entity acts on an animate one, this is almost always expressed with the inverse construction, as in (60).

(60) *bak-kay-a=is* *os* *sarampiyon*
pluck-INV-LV=3PL.AB ART.N.PST measles
'They got infected by the measles.' [JGD_160808-Fundacion-02 447]

In the few exceptions in the corpus, in which an inanimate > animate scenario is expressed by the direct construction, the actor is a natural force acting on animals, as in (61). This brings us to the problems with the notion "animacy", which I defined in an ad-hoc manner in my study. What does it mean to be animate – especially when viewed from different cultural perspectives? Is an animate entity defined by being able to control its movements? By being able to act on and affect other entities? By having cognitive capacities and being "sentient"? These questions cannot be answered easily, but studies of animacy effects on language have shown that natural forces are often classed together with animates, probably due to their agentive properties (see e.g. Lowder & Gordon 2015). By considering natural forces as animate, descriptions of natural forces acting on animals, as in (61), would be classed as scenarios with equally ranked participants. If these examples are excluded from the "low-on-high" scenarios, there are no exceptions to the inverse encoding of scenarios with "real" inanimate actors acting on animate undergoers being coded as inverse.

(61) *ban* *is* *loy* *rey* *sup-łe:-wa,*
but ART.PL NEG.SUB EPIST tie-CO-NMZ.EVT
jaa *rey* *kavuj-ka-łe-na=**a***
IJ EPIST blow-MLT-CO-DR=3N
'... but those who are not tied fast, ah!, it (i.e. the hurricane) takes (them) with it.'
[HRR_2009_tape1_A 486]

I will now turn to the middle segment of Figure 24, which evaluates the scenarios with interacting animates that are either equal in rank or in which the undergoer outranks the actor. The high frequency of direct constructions in (non-human) animate > human scenarios can be attributed to the overall preference of the direct construction. Still, in comparison to equal scenarios, the percentage of inverse constructions for the animate > human scenario is relatively high (37%). There are examples from texts with animals as main protagonists, which are encoded as the external argument of an inverse construction as soon as they act on a human. For instance, example (62) stems from a text about a dog, which is emcoded as the internal argument in all preceding direct constructions that describe it acting on objects or on other dogs. However, when a human enters the scene, the dog is encoded as external, and the human as internal argument in all subsequent clauses. (Here, expressions referring to the dog are given in boldface, and expressions referring to the landlord are underlined. To convey the pragmatic effect of the inverse construction, the inverse clauses are translated with the English passive.)

(62) *Jayłe* *jayna* *potmo* *<u>us</u>* *pa:toron-a=y'łi* *di'* *Alejandro*. *[…]*
then DSC get_up ART.M boss-LV=1PL REL Alejandro
Jayłe ***os*** ***pa:ko***, *kajłe-kay-a=<u>us</u>* ***os*** ***pa:ko***.
then ART.N.PST dog meet-INV-LV=3M.AB ART.N.PST dog
Chaywa-kay-a=<u>us</u> *"mm",* *jankwa=**os*** ***pa:ko***.
answer-INV-LV=3M.AB *"mm"* say=ART.N.PST dog
Jayłe *chuspa-kay-a=<u>us</u>* *n-os* *charki.*
then show-INV-LV=3M.AB OBL-ART.N.PST *charque*
'Then our boss, Alejandro, got up. ["What are you doing?", he asked.] Then the dog, he was greeted by the dog. He was answered "mm" by the dog. Then he was shown the dried meat (by the dog).' [JMH Perro II 041-045]

Thus, in a human > animate scenario, the human is always the internal argument of the inverse predicate even if the animal is a major protagonist of the story. In the animate > human scenario, in contrast, there is a slight preference for the animal to be encoded as the internal argument, which points towards the overall preference of the direct construction. Still, the inverse construction is found more often here than in equal scenarios. Therefore, despite the apparently vague numbers, there is a general preference for humans to be represented as the internal argument. It can therefore be maintained that low-on-high scenarios are particularly likely to be described with the inverse construction, in correspondence with the animacy hierarchy.

So, based on my corpus I was able to show that even though animacy and discourse topicality generally go together, animacy plays a central role in the choice of the construction: for example, the inverse is excluded to describe a situation in which a human acts on an object. In addition, the study shows that the central opposition in Movima is that between human and non-human, while in other languages, the major opposition is that between animate and inanimate (Aissen 1997).

Obviously, this study was not able to explain each single instance of the construction choice, especially since some of the criteria applied are somewhat opportunistic. For instance, as mentioned above, the concept of "animacy" I used is probably not quite in line with the factors that play a role in the system; however, it may be impossible to tackle the question of why a non-human entity might be considered "animate", independently from agentivity. Also the concept of "agentivity" is problematic (see Dowty 1991 for an early discussion). Here, a clearer picture might be arrived at by considering the lexical content of both the verbs and the argument RPs occurring in the corpus. My corpus of spontaneous discourse data is not large enough to see indications that particular verbs favour either the direct or the inverse forms. For instance, in 45 descriptions of animates acting on humans, 18 different verbs are involved. Of these, the verb *lap-* 'bite' is represented relatively often (10 times), and it occurs more often in the inverse than in the direct form; however, most of these examples stem from the same speaker, and the actor of the inverse form (i.e. the biter) is either a mosquito or a reptile, i.e. an animal that must be considered low in the referential hierarchy. Therefore, analyzing verb meaning would not be helpful given the limited corpus.

Besides animacy and agentivity, discourse topicality is a category that merits further investigation. Unlike semantic factors, this is a topic that can be investigated on the basis of a corpus like mine. Here, of course, problems arise from the speaker's rhetorical freedom, e.g. the impact the speaker wants to achieve, or his/her assessment of the attention paid by the

hearer, which are difficult to account for. However, in the future I intend to concentrate on the impact of "protagonism", a concept that has, to my knowledge, not yet received full attention in the linguistics literature. I intend to tackle this question in the context of several collaborative activities in which I am involved (see 8.3.2.2).

5.2 Salience/Prominence

Haude, Katharina. 2012c. Saillance inhérente et saillance discursive en movima. Faits de Langues 39: 169-180. (Special issue *La Saillance* edited by Katharina Haude and Annie Montaut.)

Haude, Katharina. 2019a. On discourse-semantic prominence, syntactic prominence, and prominence of expression: The case of Movima. *Journal of Pragmatics* 154: 22–38. (Special Issue on *Prominence in Language* guest-edited by Petra Schumacher and Klaus von Heusinger.)

Animacy, agentivity, and discourse topicality contribute to what is often referred to as the "salience" or "prominence" of a discourse referent. Animate entities, especially humans, catch the interest of the human mind more easily than inanimate entities. Agents do likewise, as has been shown in many studies. And referents known from the context are more easily retrievable than new ones. The result of the study I described above (Haude 2014a) can be summed up as "salient/prominent referents in Movima are encoded as the internal argument of a transitive clause".

When I started my collaborations with the CELIA in 2007, this lab had a research strand entitled "La notion sémantique de saillance et ses corrélats morphosyntaxiques", which understood under "salience" hierachies of animacy and similar phenomena. At the time at which I joined the CNRS in 2010, the CELIA had become part of a new research lab, the SeDyL, together with members of the *Cercle Linguistique de l'INALCO* (CLI). The members of the CLI had also worked on the notion of salience, so this was a clear connecting point between the otherwise rather independent teams. However, the CLI had taken a different perspective on the notion of salience. Their investigation focused on the means a language uses in order to make a phrase or sentence stick out from the surrounding discourse, in particular thetic (or sentence-focus) constructions. Thus, there were two interpretations of the same term: Salience$_1$, as studied by the CELIA team, involves to properties of a discourse

referent and the effects these can have on linguistic expressions (the term here corresponds to its use by Klaiman 1991). Salience$_2$, the phenomenon investigated by the CLI, refers to linguistic strategies, e.g. changes in an expression, used by speakers to make information "stick out" for the hearer. These two interpretations seem contradictory, or at least unrelated: Salient$_1$ entities tend to be referred to by precisely those expressions that are *not* linguistically salient$_2$, such as unstressed pronouns. Salient$_1$ referents are assumed by the speaker to be already present in the mind of the hearer, so the speaker does not need to spend extra energy on reminding the hearer of their identity. Salience$_2$, by contrast, is a linguistic means that is used to convey unexpected information, i.e. information that is assumed to be not yet accessible in the hearer's mind.

In 2016, I was again confronted with the problem in the context of the Collaborative Research Center (CRC 1252) "Prominence in Language" at the University of Cologne, in whose activities I was invited to participate. This large research collaboration has the goal of providing a comprehensive characterization of prominence in language. Several research teams investigate how prominence structures language regarding prosody, morphosyntax, semantics, and discourse in a large variety of languages world-wide.

The notion "prominence" receives the same two-sided interpretation as "salience", and the definition in linguistics is usually quite the same; for instance, Matthews (1997: 299) characterizes prominent entities as "standing out for whatever reason". In prosodic analysis, prominence is a well-established concept, used to describe e.g. a syllable that stands out among others due to its acoustic properties (pitch, duration, etc.; see Crystal 2008). The term is also sometimes used by linguists to explain why, on the discourse level, a particular referent is encoded e.g. by a pronoun rather than by a noun phrase: A "discourse-prominent" referent is identifiable because the hearer's attention is already directed at it, and therefore it is sufficient to refer to it with a pronoun. The term prominence is also used for syntactic phenomena (a usage that I have never seen for the term "salience"): here, "prominence" is usually associated with the privileged status of one syntactic argument, usually the subject (e.g. Jasinskaja, Chiriacescu & Hinterwimmer 2015), but it is sometimes also used to refer to clause-initial or left-dislocated constituents (e.g. Moravcsik & Healy 1998; Falk 2014). As with "salience", therefore, some of the different applications of the term "prominence" contradict each other.

In fact, the different employments of the notion of prominence (or salience) only seldom pose problems, since authors generally do not use the term on more than one level of description. For instance, Chafe (1994) only employs the term when discussing prosody, but never to characterize an, in his terminology, "accessible" or "identifiable" referent. However,

if one starts to think about this concept beyond a limited domain, one is immediately confronted with this issue. This was the case in my collaboration with the two teams working with the concept of *Saillance* in Paris, as well as with the CRC *Prominence* in Cologne. Therefore, both in my paper for the special issue of *Faits de Langues* on salience (Haude 2012b) and in a contribution to a special issue of the *Journal of Pragmatics* on prominence (Haude 2019a), I tried to point out the systematic nature of the discrepancy between the two notions salience/prominence, based on my Movima data. In order to bring across my point, I explictily stretched the notions of "prosodic prominence" and "discourse prominence".

I interpreted prosodic prominence (or salience) as involving all factors that make a linguistic expression "stick out". In the narrow sense, these can be prosodic factors such as pitch accent, lexical stress, length, and intensity. In a broad sense, an expression can be made prominent through increased length or complexity or through placement in a non-canonical position in the sentence. This is summed up in (63).

(63) Prominence of expression

- syllable level:
  - pitch
  - stress
  - length
  - intensity
- lexical level:
  - number of syllables (e.g. long vs. short pronoun)
  - choice of expression (e.g. NP vs. pronoun vs. zero)
  - simple vs. complex (e.g. modified) expression
- sentence level:
  - noncanonical construction
  - fronting

What I treated under discourse prominence (or salience) is also broader than the term suggests. A discourse referent is prominent when the speaker assumes that the hearer's attention is already directed towards this referent, so that the hearer is able to identify the referent without major effort. This corresponds quite closely to the referential hierarchy

(speech-act participants > human > non-human animate > inanimate): The higher a discourse referent is in the referential hierarchy, the more likely it is to be discourse-prominent, since it is easily identifiable. Speech-act participants are inherently prominent because they participate in the speech situation; animate, especially human, referents are more prominent than inanimate ones because, as many studies have shown, this is where humans tend to direct their attention (a point that is discussed in Gildea's contribution to the *Faits de Langues* issue (Gildea 2012). In addition, moving entities attract more attention than static ones (Talmy 2007), and the agent in a two-participant event attracts more attention than the patient (Himmelmann & Primus 2015). All these are the features of a prototypical discourse topic, i.e. of the protagonist of a text or paragraph. This type of prominence is quite well captured by Comrie's (1989: 198) notion of "topicworthiness", which links animacy to discourse topicality. However, a discourse topic can obviously also lack one or all of these ingredients: A speaker is free to establish as the discourse topic an entity that is, for instance, inanimate and acted upon. In that case, the referent is also discourse-prominent, albeit not in the prototypical way.

The factors that contribute to the discourse prominence of a referent can be conceived of as in Figure 25. The figure shows how different features feed into discourse topicality, and how each feature can independently contribute to the prominence of a referent as well. (The term used here is "discourse-semantic prominence", which was suggested by a reviewer of Haude (2019b) rather than "inherent prominence", the term with I prefer when thinking about this type of prominence.)

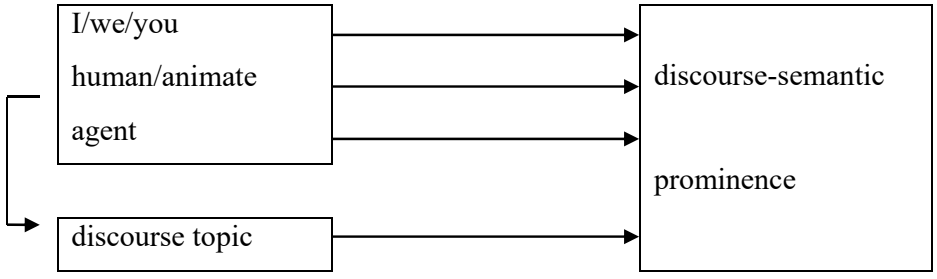


Figure 25. Deictic, semantic, and discourse features contributing to discourse-semantic prominence (Haude 2019a: 24)

The prominence of a referential expression is, on the one hand, in opposition to the prominence of the referent: Discourse-prominent referents are universally expressed with lighter linguistic material (see Chafe 1976; Prince 1981; Givón 1983b: 18; Ariel 1990; Gundel,

Hedberg & Zacharski 1993; Chafe 1994: 75–76; Arnold 2008: 495; Kibrik 2011; Haig & Schnell 2016). On the other hand, the inherently prominent referent is usually encoded as the privileged syntactic argument (or subject), if there is one. This is quite explicitly stated by Kuno's (1987) "Syntactic Prominence Principle": "Give syntactic prominence to a person/object you are empathizing with" (Kuno 1987: 232). Also Talmy (2007: 275) observes that "greater attention tends to be focused on the entity mentioned as subject". Due to the universal tendency to express prominent referents with lighter linguistic material, there is also a cross-linguistic preference to express the subject of a transitive clause by a pronoun rather than a noun phrase (see the "avoid lexical A" principle of DuBois 1987). These coss-linguistically common positive and negative correlations between the different types of prominence are shown in Figure 26.

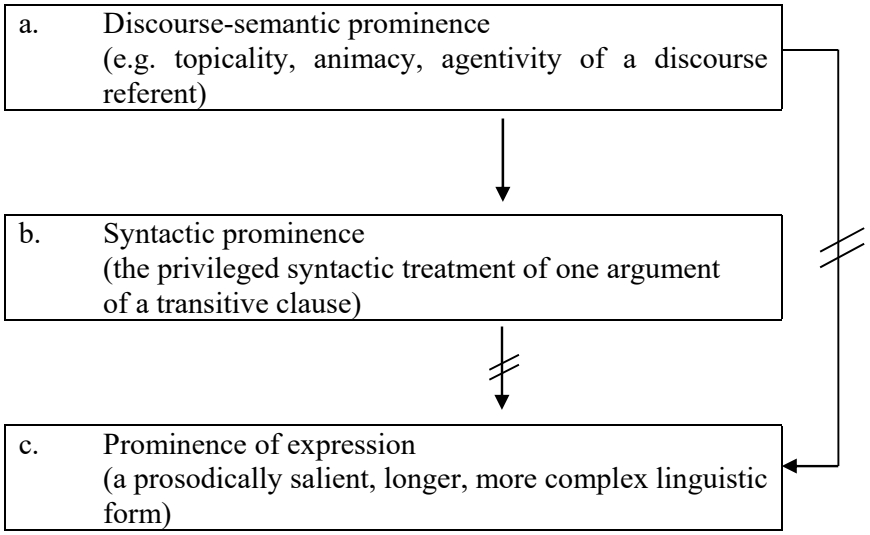


Figure 26. The correlation between discourse-semantic prominence, syntactic prominence, and prominent expressions (barred arrow = negative correlation) (Haude 2019a: 23).

5.3 The mismatch of discourse-semantic and syntactic prominence in Movima

Summing up what was shown in 4.1.3 and 5.1, argument encoding in a Movima transitive clause is governed by the discourse(-semantic) prominence (or salience) of the discourse referents. The more prominent/salient referent is encoded as the internal argument, and the less prominent referent is encoded as the external argument. In line with the universal opposition between discourse prominence and prominence of expression, the internal argument is

typically expressed by a pronominal enclitic, i.e., a nonprominent linguistic expression, whereas the external argument is more frequently encoded by a referential phrase (see Figure 23 above).

In addition, the external argument – both of the intransitive and the transitive clause – can be represented by a free pronoun, which is only a very restricted option for internal arguments. Free pronouns differ from pronominal enclitics in that they are disyllabic and receive stress of their own. Table 13 shows this for the third-person forms.

Table 13. Third-person pronouns

	presential		absential	
	free	enclitic (=/--)	free	enclitic (=/--)
3M	*u'ko*	*u'*	*usko*	*us*
3F	*i'ne*	*(i)'ne*	*isne*	*(i)sne*
3N	*a'ko*	*a'*	*asko*	*as*
3PL	*i'ko*	*i'*	*isko*	*is*

In Movima, free pronouns typically occur clause-initially. The possibility to be expressed by a fronted free pronoun (just like relativization) is a syntactic privilege of the external argument, which is not shared by the internal argument (at least not as productively, and not with the same function; see Haude 2019d). It is illustrated in (64)–(66), where the English counterparts of the fronted pronouns are underlined in the translation. The fronting construction, which I later termed "pronominal construction", will be further discussed in Chapter 6.

(64) Intransitive basic clause vs. fronting

a. *ji<wa:~>wa--us*
come<MD~>--3M.AB
'He came.'

b. *usko* *ji<wa:~>wa*
PRO.3M.AB come<MD~>
'<u>He</u> came.'

(65) Direct transitive basic clause vs. fronting

a. *vel-na=us--k-is*
watch-DR=3M.AB--OBV-3PL.AB
'He watched them.'

b. *isko* *vel-na=us*
PRO.3PL.AB watch-DR=3M.AB
'He watched them.'

(66) Inverse transitive basic clause vs. fronting

a. *vel-kay-a=us--k-is*
watch-INV-LV=3M.AB--OBV-3PL.AB
'They watched him.'

b. *isko* *vel-kay-a=us*
PRO.3PL.AB watch-INV-LV=3M.AB
'They watched him.'

When the internal argument of a transitive clause is to be topicalized through fronting, the valency-decreasing operation (introduced in 4.1.2.2) is necessary. (If the undergoer is overtly expressed in this construction, this is done with an oblique-marked RP or free pronoun.)

(67) Fronting with valency decrease

usko ***kwey*** *vel-na* *n-isko*
PRO.3M.AB VALDECR watch-DR OBL-PRO.3M.AB
'He watched them.'

If the argument with syntactic privileges is seen as syntactically "prominent", then Movima shows a mismatch between discourse prominence and syntactic prominence, and is different from other languages in this respect. My two publications on this topic, one in the context of the working group on "salience" and the other in the context of the collaborative research center on "Prominence", are summarized below.

In Haude (2012b), published in *Faits de Langues*, I presented a first description of the fact that only the argument with the non-prominent (here called non-salient) discourse referent has syntactic privileges. This paper gives details on the different, also less common forms of pronoun fronting. It includes data that I have never published elsewhere, which show that when a first-or second person pronoun is involved, fronting can allow the expression of a lower-ranking discourse referent as the internal argument. However, examples of this are this are scarce, and therefore I will not discuss them further here.

In Haude (2019a), by contrast, published in a special issue of the *Journal of Pragmatics*, I concentrated on constructions that are very well represented in the corpus. As requested by the reviewers, I demonstrated again the role of referential factors for argument encoding in transitive clauses, and showed that the less prominent referent has more syntactic possibilities than the argument with the more prominent referent. As mentioned above, the former can be relativized, and it can be fronted. Moreover, the external argument can also be questioned without any further syntactic operation. This is because question formation involves a "verbal RP" (see 4.2.2) containing the questioned element, why the question itself is formed by an interrogative predicate. This is illustrated in (68) with a direct-marked predicate in a question asking for the undergoer, and in (69) with an inverse-marked predicate in a question asking for the actor in a two-participant event.

(68) *łéła* *kos* *dewaj-na=nkweł*
be_what ART.N.AB see-DR=2PL
'What did you see (lit.: What was [the entity] you saw)?'
[HRR_120808-tigregente 519]

(69) *łéła* *kos* *tarał-kay-a=nkweł*
be_what ART.N.AB heal-INV-LV=2PL
'What (was the entity that) you (pl.) were healed by?'
[ERM_140806_1 0938]

The discrepancy between discourse-semantic and syntactic prominence in Movima is schematized in Table 14.

Table 14. Discourse-semantic vs. syntactic prominence of internal and external argument

		internal	external
referent	SAP (> 3rd)	+	–
is:	discourse topic	+	–
	human/animate	+	–
	Agent	+	–
has	relativization	–	+
access	pronoun fronting	–	+
to:	wh-question	–	+

The main point of this paper, however, was that if one looks more closely at the functions of the constructions that privilege the external argument, the typologically unusual discrepancy between discourse-semantic prominence and syntactic prominence in Movima makes sense. Relativization (especially restrictive), pronoun fronting, and wh-question formation have one fundamental property in common, which distinguishes them from argument deletion rules: They establish or increase the identifiability of a discourse referent.

A restrictive relative clause is a means to render a referent identifiable and available for the subsequent discourse, as formulated by Fox (1987: 861): "[R]elative clauses serve to situate the referent that is being introduced as a relevant part of the ongoing discourse; in a sense, they justify the introduction of the referent in the first place". Undergoer arguments, with their *a priori* nonprominent status, are a common target for relativization also in nominative-accusative languages (see Van Valin & LaPolla 1997: 306–307; Gordon & Hendrick 2005; Bickel 2011: 428–429; Ganenkov 2016). That in Movima transitive clauses only the external argument has access to relativization is coherent with this tendency: This argument encodes the discourse-semantically less prominent referent, and relativization provides this referent with a prominence feature that it does not originally possess, namely better identifiability. SAPs and other discourse-semantically prominent referents that are referred to with pronouns are identifiable, which is why (restrictive) relativization is usually not an option for them.

Pronoun fronting, whose function is discussed in more detail in Chapter 6, shares with relativization the pragmatic property that "the clause in which the displaced NP functions is always *about* the referent of the NP" (Van Valin & LaPolla 1997: 627). In Movima, the fronted free pronoun explicitly singles out the last-mentioned referent as the sentence topic, about

which the predicate then provides the comment (see 6.2.1 below). In this way, the free pronoun has a reference-tracking function, which helps the hearer to identify the referent about which the comment is made. Importantly, as will be shown below, the referent of the free pronoun is usually *not* the discourse topic, and this may explain why a fronting construction is needed to render this referent identifiable.

Questions (and answers) differ from both relativization and fronting in that they are focus constructions. However, asking a question regarding the identity of an event participant is a way to render a referent identifiable, too: The question guides the hearer's attention to an entity whose identity the speaker does not know or is not sure about. Ideally, the hearer will be able to provide this information, thereby rendering the referent identifiable; both attention and identifiability are features of discourse-semantic prominence. In sum, the constructions that single out the OBV argument in Movima render a referent prominent that is not prominent in the first place.

Apart from their pragmatic prominence-lending function, the constructions that privilege the external argument are prominent expressions, in the sense of (63) above. They are complex, being the result of an extraction process. They also involve more physical, i.e. acoustic, material than basic clauses. A relative clause creates a long and complex NP. Fronting, while simple on the surface, involves a pronoun that occurs in non-canonical, initial position and that is prosodically heavier (disyllabic, stress-bearing) than its bound counterpart; furthermore, the initial position of the pronoun is inherently prominent due to its "edge placement" (Himmelmann & Primus 2015: 50). A question, finally, requests an answer, thereby interrupting the ongoing flow of discourse and potentially triggering additional linguistic material.

The fact that the argument with the nonprominent discourse referent has access to these constructions is in line with the universal negative correlation between discourse-semantic prominence of a referent and a linguistic form of high prominence of expression (see Figure 26): Not only is a discourse-semantically prominent referent expressed by a linguistic form of low prominence of expression, but also vice versa: a nonprominent referent can have privileged access to a linguistic form of high prominence of expression. If, as (Talmy 2007: 282) puts it, "a longer form attracts more attention to the concept, while a shorter form attracts less attention," then longer forms are more adequately applied to concepts that are not prominent in the first place.

Thus, it is not just a puzzling idiosyncratic property of Movima that the discourse-semantic prominence of a referent is opposed to the syntactic prominence of the argument by which it is

encoded. On the contrary: In a grammar that pays much attention to the relative discourse-semantic prominence of the participants in a two-participant event, it makes sense that the participant with less prominence features is more likely to figure in constructions whose function is to describe an entity in more detail (like a relative clause), to single it out as a sentence topic (by pronoun fronting), or to identify its referent (as is asked for in a question). In other words, the Movima findings support and reinforce the idea that prominent referents are likely to be expressed in a non-prominent way, while nonprominent referents require a more explicit description that stands out among the surrounding discourse.

Whether syntactic prominence correlates with the discourse-semantic prominence of a referent or not is a different matter. This depends on the syntactic domains in which syntactic prominence shows up. The fact that in many languages, subjects allow deletion in coordination (see 4.1.2) most probably goes hand in hand with the typical topic-encoding property of subjects. Syntactic prominence showing up in the possibility to be relativized or otherwise 'extracted', however, is functionally quite different and more likely to be dissociated from discourse-semantic prominence. So, this study calls for a reassessment and differentiation of the properties that privilege a particular grammatical relation, which were introduced by Keenan (1976) and applied many times since.

6. Nonverbal predication and information structure

The idea that in Movima, all content words are primarily predicates which only acquire a referential function when combined with a referential element (see 4.2 above) has kept me busy ever since I finished my PhD thesis. In the context of several collaborative activities since then (see 2.2.1–2.2.3), I took a closer look at this issue from different perspectives. The event that prompted me to investigate nonverbal predication in its own right was the workshop "Nonverbal Predication in Amazonian languages" at the *Amazónicas* conference in Belém in 2014 and the ensuing collective volume (Overall, Vallejos & Gildea 2018a). For this purpose I systematically investigated all the elements that can occur as predicates in Movima (6.1). These include not only verbs and common nouns (including adjectives, which form a subclass of common nouns; see 6.1.1), but also function words. For instance, I could identify the negator as a copula, which solved the problem of why a negated predicate is nominalized.

A discovery with far-reaching consequences in this domain was that "fronted" free personal pronouns, already illustrated in (64)–(66), are actually predicates of a complex sentence. Since this "pronominal construction", which looks very simple on the surface, will come up again and again in this chapter, it is represented schematically in Figure 27.

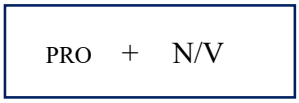


Figure 27. The pronominal construction

For instance, I realized that nonverbal clauses with a fronted pronoun are needed to express identification (6.1.2), and that a fronted pronoun marks the difference between a predicate-nominal focus construction and a cleft (6.2.4). The fact that this construction is discussed below from different perspectives (and perhaps in more length than necessary) reflects my gradual progress in understanding it. It also shows that this is ongoing work, and that the publication of a more concise account that does full justice to this construction is still pending.

When investigating information structure in the context of the LABEX project, I found out in which ways nonverbal clauses form pragmatically marked sentences (6.2), even though not necessarily with the expected effect. The pronominal construction shows similarities with a cleft, but it does not mark argument focus (6.2.1) and is probably better analyzed as a simple equational clause (6.2.2), in which the pronoun assumes the function of a copula (6.2.3). Focus is expressed with other, though related types of nonverbal sentences (6.2.4): with a simple clause consisting of a nominal predicate and a "verbal RP" (6.2.4.1), or with a syntactically complex construction (6.2.4.2) that I discovered rather late and which is based on the pronominal construction depicted above.

6.1 Types and functions of nonverbal predicates

Haude, Katharina. 2018a. Nonverbal predication in Movima. *Nonverbal predication in Amazonian languages*, 217-244. Overall, Simon, Rosa Vallejos, and Spike Gildea (eds.). Amsterdam/Philadelphia: Benjamins.

6.1.1 Identifying nonverbal predicates: nouns, pronouns, and a negative copula

In Haude (2018a) I asked the following two questions in order to a) distinguish a verbal from a nonverbal predicate in Movima and b) identify the predicate among the several elements in a clause:

a) How is a predicate derived when embedded?
b) Which of the elements in an embedded clause is nominalized?

In Movima, the predicate occurs in clause-initial position. The predicate is typically a verb, as in (70), but it can also be a noun, as in (71), or an adjective, as in (72). Adjectives are near-identical with common nouns, which is why I usually just treat them as nouns.

(70) *jo'yaj--us* *neyru*
arrive--3M.AB here
'He arrived here.' [EAO_120906_3 007]

(71) *tolkosya--'ne*
girl--3F
'She is a girl.' [Dial. EA&AH 012]

(72) *tochik--as*
small--3N.AB
'It was small.' [EGA Cazando 082]

Predicates can form a clause of their own, i.e. without an accompanying argument expression (in practice, such clauses usually also contain at least a discourse particle). This is illustrated in (73) with an intransitive verb (already shown in 4.1.2.2, example (18)), in (74) with a noun, and in (75) with an adjective.

(73) *nokowa* *jayna* ***jo'yaj***
FUT DSC arrive
'(He) will arrive.' [ERM_140806_1 0269]

(74) *di:ra* ***tolkosya***
still girl
'(They were) still girls.' [ERM_150806 194]

(75) *ban* ***tochik*** *ney*
but small here
'but (they were) this small.' [AMY_180806 177]

Most nouns are not formally distinct from intransitive verbs: There is nothing about the form of the word *tolkosya* '(be a) girl' that identifies it as a noun, and also the verb *jo'yaj* 'arrive' shows no formal sign of being a verb. (This is different with verbs that contain a voice marker, as shown in Table 10 in Section 4.1.5, which are usually verbs.) A good diagnostic for distinguishing verbal from nominal predicates is their form in embedding. Embedded predicates are morphologically nominalized, and verbal and nonverbal predicates are nominalized differently. Verbs take the suffix *-wa*, as illustrated in (76). Nouns, in contrast, undergo reduplication, as shown in (77). (Both are marked as possessed when embedded; see 7.2.)

(76) *bele:ka* *n-os* *joyaj-**wa**=Ø*
happy OBL-ART.N arrive-NMZ.EVT=1SG
'(She) was happy when I arrived (lit.: "… at my arriving").' [GCM Bacho 029]

(77) *dottoł--isne* *n-os* *tolkos**<ya~>**ya=sne*
bad_person--3F.AB OBL-ART.N.PST girl<NMZ.ST~>=3F.AB
'She was a bad person when she was (lit.: "at her being") a girl.' [EAO Mala 002]

Adjectival predicates are like nouns in that they are reduplicated when embedded, as in (78). In negation, which also involves embedding (see below), they can alternatively receive a suffix *-łe*, as in (79), which is the only hint that there might be a formal distinction between nouns and adjectives.

(78) *jayna* *n-os* *ja**<ya~>**yaw-a=is*
DSC OBL-ART.N.PST good<NMZ.ST~>-LV=3PL.AB
'Then, when they were good …' [MCA_280806_1 421]

(79) *ka=s* *jayaw-**łe**=kos* *łakapba*
COP.NEG=DET good-NMZ.ADJ=ART.N.AB earth
'The earth is not good (for pottery).' [CCT_120907_2 004]

Also some types of function words can function as predicates. When embedded, these predicates take the suffix *-niwa*, which is probably a fossilized combination of a verbalizer *-ni* and the verbal embedding marker *-wa*. Example (80) illustrates this with an embedded locative adverb.

(80) *jayna* *pakuk-na=Ø* *os* *nosde-**niwa**='ne*
DSC understand-DR=1SG ART.N.PST over_there-VBZ:NMZ=3F
'I already knew that she (was) over there (lit.: "I already knew her being over there")." [EAO In between 023]

Predicative demonstrative pronouns are illustrated in (81). Like other function words, pronouns take the suffix *-niwa* when embedded, as shown in (81b).

(81) a. *n-as* *Kachwela* ***koro'*** *kos* *ra:diyo*
OBL-ART.N Cachuela DEM.N.AB ART.N.AB radio
'In Cachuela there is a radio.' [EAO_120906_3 112]

b. *ja:yaw* *as* ***koro'-niwa*** *kos* *dajwanas*
nice ART.N DEM.N.AB-VBZ:NMZ ART.N.AB question
'It's nice when there is a question (lit.: the there-being a question).'
[ERM_140806_1 0621]

The criterion of nominalization in embedding has also made me realize that "fronted" personal pronouns are predicates as well. A personal pronoun can function as a clause on its own, as shown in (82):

(82) ***u'ko***
PRO.3M
'It's him.' [GCM_290806_4 149]

A fronted pronoun in a main and an embedded clause is illustrated in (83a) and (83b), respectively.

(83) a. ***usko*** *ji<wa:~>wa*
PRO.3M.AB come<MD~>
'He came.' [ATL_230806 173]

b. *kem<a:>ye=Ø* *os* ***a'ko-niwa*** *ja'* *ji<wa:~>wa*
believe<DR>=1SG ART.N.PST PRO.3N-VBZ:NMZ just come<MD~>
'I thought it (the hen) had just come (on its own) (lit.: I believed the it-being [who] came).' [EAO Gallina 018]

The investigation of nonverbal predicates, furthermore, has allowed me to find a more adequate analysis of negation, already seen in (79) above. Earlier, I had analyzed the initial element *kas* as a negative particle and was unable to explain why it was followed by a nominalized form (Haude 2006: 543–544). However, when I took as criteria for identifying a predicate the ability to form a clause and the nominalization in embedding, I was able to identify the element *ka* as a negative copula, to which a determining element *=s* is attached and that is followed by a nominalized, possessed lexical predicate: "X's Y-ing is not". Support for this analysis comes from the fact that all determiners, i.e. articles and demonstrative determiners, end in an *s*, which allows the same lenition to [h] in these environments. In an environment involving a first person, furthermore, the *=s* can be replaced by the optional first-person clitic *=ɬ* both on determiners, as in (84), and on the negative copula, as in (85).

(84) ***aɬ*** *dokwe=Ø*
ART.N:1 clothes=1SG
'my clothes' [LYO_250808_2 058]

(85) *jayna* *ka=**ɬ*** *joy-wa=Ø*
DSC COP.NEG=1 go-NMZ.EVT=1SG
'I don't go (there) anymore (lit.: "My going is already not").'
[ERM_140806_1 0846]

Furthermore, like other main-clause predicates, the negative copula can form a clause of its own, where it can be followed by a bound pronoun, as in (86). Its vowel is then lengthened and combined with the dummy element *-'i* (see 4.1.5), resulting in a full prosodic word.

(86) *jayna* ***ka:-'i--is*** *jayna*
DSC COP.NEG-D--3PL.AB DSC
'They aren't (there) anymore.' [MCA_280806_2 355]

This long form of the negative copula can also occur in embedded clauses, as illustrated in (87) where it is negated with *-niwa* like all function words. (Note, however, that there are only two such occurrences in the corpus, and in both, the meaning seems to be conventionalized as "not to be in one's normal state".)

(87) *jayna* *n-os* *da'* ***ka:-'i-niwa*** *jayna*
DSC OBL-ART.N.PST DUR.NSTD NEG-D-VBZ:NMZ DSC
'(when she was ill), when (she) couldn't do anything anymore (lit.: "in the not-existing anymore") …' [EAO Ay'ku II 009]

The element *ka*, therefore, can be analyzed as a negative copula. It is a special kind of predicate, which is prosodically defective and only rarely occurs independently. Analyzing it as a copula followed by a nominalized phrase makes sense, since this the only way to explain the nominalization of the lexical predicate.

6.1.2 Nominal vs. pronominal clauses: Categorization and identification

Nominal predicates form equational clauses of the type "X is Y". This clause type can have several functions. The main distinction is usually made between identification, on the one hand, and categorization, on the other hand. Identificational predication establishes the identity of the referent expressed by the subject with the referent denoted by the predicate (as in *Emmanuel*

Macron is the president of France). Categorizing predication indicates a class membership (as in *Brigitte Macron is a teacher*). In their introduction to the volume on nonverbal predication in Amazonian languages, Overall, Vallejos & Gildea (2018b) write that crosslinguistically, Amazonian languages make either no difference between identification and categorization, or only a difference in the form of the predicate NP: An identificational predicate is a definite/referential noun (as indicated by definite article *the* in the first example above), whereas a categorizational predicate is indefinite/non-referential (as indicated by the indefinite article in the second example). They state that they have not found statements on "any South or Central American language that utilizes different constructions for these two functions" (Overall, Vallejos & Gildea 2018b: 7). By the time I was finishing my contribution to the volume, however, I realized that Movima does make a difference between the two clause types: The pronominal construction can establish identificational predication, while a nominal clause without a preceding free pronoun is always categorizational (or attributive, if the predicate is an adjective).

Only unpossessed, common nouns (including adjectives) can freely function as predicates in Movima. In (88), we see three clauses in a row, each containing an intransitive predicate with its argument expressed by an encliticized bound pronoun. The first two, *polkababa* 'roll around' and *potmo* 'get up' are verbs, while the last one, *rulrul* 'jaguar', is a noun.

(88) *jayna* *pol<ka>ba:ba--as* *łat,* *potmo--as,* *jayna* ***rulrul****--as*
DSC roll_around<MLT>-- 3N.AB EV get_up--3N.AB DSC jaguar--3N.AB
'Then it rolled around, it got up, then it (was a) jaguar.' [LYO_250808_2 231]

By contrast, possessed nouns and proper names, which are generally nouns with specific referents cannot be combined with an encliticized bound pronoun. This is shown by (89) (which is the only negative example I have of this kind).

(89) **pa:ko=us--k-as*
dog=3M.AB--OBV-3N.AB
(intended: 'It is his dog.')

Instead, the argument must be expressed by a fronted free pronoun (see Figure 27), as in (90). Also with proper names, the free pronoun must be placed before the noun in order to form an equational clause, as illustrated in (91).

(90) *asko* *pa:ko=us*
PRO.3N.AB dog=3M.AB
'It is his dog.'

(91) *jina:nak* *u'ko* *Ernan*
perhaps PRO.3M Ernan
'Perhaps it's Ernan.' [EAO Cbba 171]

In general, the construction with the fronted pronoun is the most productive way to form equational clauses, also non-identificational ones. For instance, (92) is clearly categorizational.

(92) *usko* *chiman*
PRO.3M.AB Chimane
'He was a Chimane (Indian).' [EAO_120906_3 010]

An embedded equational clause with a fronted pronoun is shown in (93). As was shown above for the same construction with a verb (see (83b)), also here, it is the pronoun that is nominalized; further parallels with the verbal pronominal construction are pointed out in Section 6.2.2.

(93) *n-os* *usko-niwa* *pa:pa=is*
OBL-ART.N.PST PRO.3M.AB-VBZ:NMZ father_of=3PL.AB
'that he was their father (lit.: "in the he-being their father")'
[JGD_160808-Fundacion_1 453-454]

To sum up, in Movima there is a difference between categorizational and identificational clauses. Categorizational clauses can be formed by simply placing a noun in predicate position. Identificational clauses are obligatorily formed with a pronominal predicate. Thus, common nouns behave syntactically like verbs, while nouns with a specific referent are more restricted in their distribution.

6.1.3 The content word in the pronominal construction: a relative clause

The pronominal construction, in which the lexical predicate is preceded by a free personal pronoun (see Figure 27), has occupied me for several years and is a matter of ongoing research (see 6.2.3). When analyzing the construction in terms of fronting, i.e. as a placement of a clause-final element in clause-initial position for pragmatic purposes, I struggled with the observation that when a clause with a fronted pronoun is embedded, it is the pronoun that is nominalized, not the verb. The examples from above are repeated again in (94), with the pronoun in the main clause in a and in the embedded clause in b.

(94) a. ***usko*** *ji<wa:~>wa*
PRO.3M.AB come<MD~>
'He came.' [ATL_230806 173]

b. *kem<a:>ye=Ø* *os* ***a'ko-niwa*** *ja'* *ji<wa:~>wa*
believe<DR>=1SG ART.N.PST PRO.3N-VBZ:NMZ just come<MD~>
'I thought it (the hen) had just come (on its own) (lit.: I believed the it-being [who] came).' [EAO Gallina 018]

Obviously, if the criterion of nominalization in embedding for identifying a predicate is taken seriously, then also in the main clause, the pronoun must be the predicate. For me, however, this was a problem: How can a referential expression be a predicate at the same time? In March 2014 I presented this puzzle to my colleagues of the SeDyL programme *L'énoncé et ses composantes* (see 2.2.2), most of whom have a background in Culiolian grammar. They did not consider this analysis as problematic at all and pointed out that a predicate is in the first place a relational expression, independently of its semantic content. From then on, I analyzed the free pronoun as a predicate, and found some support in the literature on Salish languages (Shank 2003) and Tagalog (Schachter & Otanes 1972: 64), where pronominal predicates have been described as well.

While the problem of the pronoun was now solved, the status of the content word remained problematic. It was ruled out as a main predicate, and it could not be analyzed as an argument, since in that case, it should be preceded by a determiner, as in basic clauses or clauses with a demonstrative predicate (see (81)). I pondered about this for a long time, and, as so often, the decisive step towards the solution came in an unexpected situation. During a meeting of the

CorTypo project (see 2.2.4) in June 2014, Zygmunt Frajzingier asked me what I was working on, and when I briefly described this problem, he said: “Maybe it is a relative clause?”

At first, I discarded this suggestion since I did not see why the content word in this construction should be a relative clause. But then I realized that if the pronoun is a predicate, it would mean “She/he/it etc. is ...”; and when such a predicate, in English or German, is followed by a verbal expression, this would be in the form of a relative clause. I then took a closer look at the predicate that follows the free pronoun and compared it to the predicate of headed relative clauses, illustrated once more in (95) (see (34b) for a similar example).

(95) *us* *dichi:ye* *di’* *kay-a-poj-a=sne*
ART.M child REL eat-DR-CAUS-LV=3F.AB
‘the/a boy that she fed’

And indeed: While on the surface, the predicate of a relative clause is indistinguishable from a main-clause predicate, there are several syntactic features that distinguish the two. Unlike a main-clause predicate, the predicate of a relative clause (and, as we will see, the content word of the pronominal construction as well as the content word in an RP) has the following properties:

- It may not be followed by an externally encliticized pronoun.
- It can undergo the valency-decreasing operation.
- It is negated with the particle *loy*.

Firstly, the relative clause, and the pronominal construction, can be analyzed in terms of “extraction”: the relativized/fronted element is not expressed again with a coreferential encliticized pronoun following the predicate. Secondly, as was shown in Section 4.1.4, only the external argument can be relativized. Likewise, only the external argument can be “fronted”, i.e. function as the predicate of the pronominal construction, as was shown in 5.3, (64)–(66). Example (96) illustrates it again with a transitive predicate.

(96) *usko* *kay-a-poj-a=sne*
PRO.3M.AB eat-DR-CAUS-LV=3F.AB
‘She fed <u>him</u>.’

As is the case with relativization, the internal argument can only be fonted after a valency-decreasing operation, which promotes this argument to external argument status (see (67) above). Example (97a) illustrates, once again, a basic transitive direct clause, in which the actor (*us Ernan*) is expressed as the internal argument. Example (97b), which contains the same direct-marked verb and has a similar propositional content, is a pronominal construction, in which the actor is expressed by an initial free pronoun. For this to be possible, the verb is preceded by the valency-decreasing particle *kwey*, which derives an intransitive clause, and the undergoer can only be expressed by an oblique-marked element.

(97) a. *jayna* *jay<a>moł-a=us* *Ernan* *us* *pa:toron-a=y'łi*
DSC call<DR>-LV=ART.M Ernan ART.M landlord-LV=1PL
'Then Ernan called our landlord.' [EAO Cbba 196]

b. *jayna* ***usko*** ***kwey*** *jay<a:>moł* *n-os* *aviyone:ta*
DSC PRO.3M.AB VALDECR call<DR> OBL-ART.N.PST plane
'He called the plane.' [EAO_240807_vibora 144]

Thirdly, the predicate of a headed relative clause and the content word in the pronominal construction are negated in the same way, which is different from main-clause negation. A main clause is negated with the negative copula *ka* followed by a nominalized predicate (see also 6.1.1), as shown in (98).

(98) *ban* *ka=s* *iwani-wa=us*
but COP.NEG=DET speak-NMZ=3M.AB
'But he doesn't speak.' [ERM_140806_2 259]

In contrast, the predicate of a relative clause and the content word in the pronominal construction are negated with a particle *loy*, as shown in (99) and in (100), respectively. (The nominalized intransitive predicates are not possessed in this construction.)

(99) *kis* *tolkosya* *di'* ***loy*** *iwani:-wa*
ART.PL.AB girl REL NEG.SUB speak-NMZ.EVT
'the girls who don't speak' [CCT_120907_2 124]

(100) *u'ko* ***loy*** *iwani:-wa*
PRO.3M NEG.SUB speak-NMZ.EVT
'He doesn't speak.' [CCT_120907_2 102-104]

Negation with *loy* is found elsewhere only with complement and adverbial clauses, and is illustrated with a complement clause in (101).

(101) *jayna* *kaw-yemes* *as* ***loy*** *joy-wa=y'łi*
DSC much-day ART.N NEG.SUB go-NMZ.EVT=1PL
'It's been many days already that we haven't been going (there).' [EAO Asilo 079]

Hence, a verb that follows a referring expression can be analyzed as a relative clause that specifies semantically the preceding referring expression. The pronominal construction can be paraphrased with "(It is) X, the (one who) Vs".

Furthermore, the pronominal construction and the relative clause share their properties with RPs (see 4.2.2). An RP refers to the entity that constitutes the external argument of the content word in predicate function. In order to refer to an internal argument, the valency-decreasing construction must be used, as shown in (102) (repeated from (53b) above). In addition, also the content word inside an RP can be negated, and this is likewise done with the particle *loy*. This is illustrated in (103).

(102) ***us*** ***kaw*** *vel-na* *n-is* *waki:ya*
ART.M VALDECR watch-DR OBL-ART.PL calf
'the (one who) looked after the calves; the watcher of the calves' [GBM Ganado 061]

(103) *ban* ***is*** ***loy*** *rey* *sup-łe:-wa,*
but ART.PL NEG.SUB EPIST tie-CO-NMZ
jaa *rey* *kavuj-ka-łe-na=a*
IJ MOD blow-MLT-CO-DR=3N
'... but those who are not tied (onto something), ah!, it (i.e. the hurricane) takes (them) with it.' [HRR_2009_tape1_A 486]

From all this I conclude that content words that are preceded by a referential expression (a pronoun, an article, or a referential phrase) are subordinate predicates, which correspond to the different types of relative clauses listed in (104).

(104) Types of relative clauses depending on the preceding referential element/phrase

a) RP + *di'* → headed RC
b) Article → light-headed RC
c) Free pronoun → headless RC

In all three cases, the content word specifies the referent of the preceding referential expression. Regarding their analysis, more research is still required. The possibility to analyze RPs and headed relative clauses as nominals seems straightforward, since both the placement in a referential phrase and relativization can often be traced back to nominalization crosslinguistically (see Shibatani 2019), but is this also possible for the content word following the free pronominal predicate? Some ongoing research on the pronominal construction is described in the following section.

6.2 The pragmatic markedness of nonverbal predicates

New insights on the pronominal construction as well as on other nonverbal predicates came through my participation in collaborative projects that dealt with the investigation of information structure. The programme *The Typology and Corpus Annotation of Information Structure and Grammatical Relations* (LABEX-EFL, see 2.2.3) helped me to advance in this domain, especially through the conferences *Information Structure in Spoken Language Corpora 2* I organized with Evangelia Adamou and Martine Vanhove (December 2015) and *Clefts and related focus constructions* I organized with Enrique Palancar and Martine Vanhove (March 2018), and through the ensuing publications. These activities forced me to take a closer look at the pragmatic effects of noncanonical constructions, and in doing so, I gained new insights into their morphosyntactic and prosodic structure.

6.2.1 Is the pronominal construction a cleft?

Haude, Katharina. 2018b. A topic-marking cleft? Analyzing clause-initial pronouns in Movima. *Information structure in lesser-described languages: Studies in prosody and syntax*. Adamou, Evangelia, Katharina Haude, and Martine Vanhove (eds), 217-244. Amsterdam/Philadelphia: Benjamins.

The pronominal construction is exemplified once more in (105)–(107) (repeated from (64)–(66)) with an intransitive, a transitive direct, and a transitive inverse verb, respectively.

(105) *usko* *ji<wa:~>wa*
PRO.3M.AB come<MD~>
'He came.'

(106) *isko* *vel-na=us*
PRO.3PL.AB watch-DR=3M.AB
'He watched them.'

(107) *isko* *vel-kay-a=us*
PRO.3PL.AB watch-INV-LV=3M.AB
'They watched him.'

If, as was established above, the pronoun is the predicate of the pronominal construction and the content word in this construction is a headless relative clause, this means that this construction is a syntactically complex, bipartite sentence. It can be characterized as a cleft, defined by Payne (1997: 278) as follows: "A cleft construction is a type of predicate nominal consisting of a noun phrase (NP_i) and a relative clause whose relativized NP is coreferential with NP_i". In the Movima pronominal construction, Payne's NP_i is represented by the free pronoun, and the relativized NP, i.e. the gapped S/OBV argument of the verb, is coreferential with this pronoun. In addition, like a cleft, the pronominal construction can be used as an alternative to a basic simple clause with the same propositional content (see (64)–(66) above).

This was, for me, a spectacular finding: A construction that looked like the result of simple fronting turned out to have a "hidden complexity" (a term coined by Bisang 2015), which only

becomes apparent in derived structures (embedding, possibility of valency decrease, negation with *loy*; see 6.1.3).

I then realized that the phenomenon is not as unique as I had thought, however. In personal communication with Judith Tonhauser (Ohio State University), for instance, I learned that a similar proposal had been made for Yucatec Maya. For this language, some scholars (Bricker 1979; Tonhauser 2003; Bohnemeyer 2004) have argued that the so-called "agent focus" construction with a preverbal actor argument is, in fact, a cleft. Others, however (Verhoeven & Skopeteas 2015), argue in favor of a monoclausal analysis, whose focus-marking effect is the result of fronting. This is an attractive solution because it means that no syntactic complexity needs to be postulated. However, in Movima, the evidence for the predicate status of the fronted pronoun and the "nonfinite" status of the content word is clear, and therefore, the cleft analysis of the pronominal construction seems more adequate.

For the purpose of my publication (Haude 2018b) I then started to look at prosodic properties and discourse functions of the pronominal construction. Here, it turned out that the cleft analysis for the Movima pronominal construction is difficult to keep up. First of all, the free pronoun is almost never prosodically prominent. A prosodic analysis I carried out together with Candide Simard (then SOAS, London) confirmed that the free pronoun does not receive any prominence; on the contrary: it is not the pronoun, but the content word that is prosodically prominent, marked by higher pitch, wider excursion, more intensity and longer duration. For instance, in the graphic representation of a pronominal construction (108) in Figure 28, the pronoun *usko* 'him' is part of the same intonation unit as the verb phrase *yey-na=is* 'they want', but is clearly not prosodically salient in comparison the verb.

(108) *usko* *yey-na=is* *ja'a*
PRO.3M.AB want-DR=3PL.AB just
'They want only <u>him</u>.' [MCA_280806_1 044]

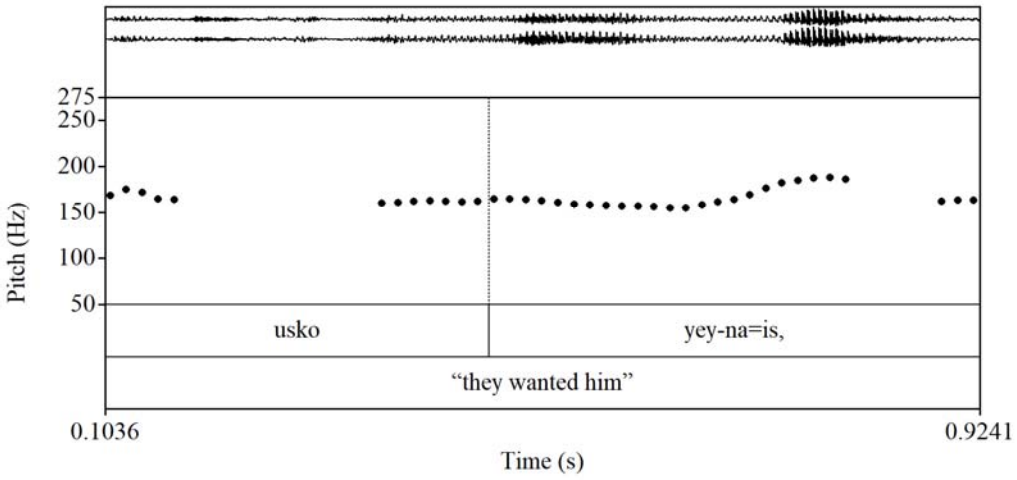


Figure 28. Wave form and pitch representation of a pronominal construction

This prosodic pattern does not correspond to the information structure of a cleft like the English sentence *It is her who cooked the dinner*, where we would expect the pronoun (*It is HER*) to be prominent and the relative clause (*who cooked the dinner*) to be nonprominent. Of course, focus is not necessarily associated with prosodic pronominence (see Rialland & Robert 2001). However, in Movima basic clauses, the predicate (in initial position) receives the main sentence stress. Therefore, if the pronoun is the predicate in the pronominal construction and bears the focus, we would expect the pronoun to be prosodically prominent, which it is not.

The cleft analysis is not supported by the discourse use of the construction, either. Most definitions of clefts (except purely structural definitions like the one by Payne above) state that the clefted constituent marks argument focus:

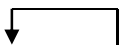
> Cleft constructions are focus-marking devices used to prevent unintended predicate-focus construal of a proposition. Clefts serve to mark as focal an argument that might otherwise be construed as nonfocal, or as nonfocal a predicate that might otherwise be construed as focal, or both. (Lambrecht 2001: 489)

In the Movima pronominal construction, however, the free pronoun has the function of establishing as a sentence topic a discourse participant that was introduced immediately before and that does not persist in the subsequent discourse. (For this reason, and on a purely intuitive basis, I had labelled the pronominal construction as "marked-topic construction" e.g. in Haude 2009a).

For example, in (109), the referent is introduced into the discourse by an RP in the clause preceding the pronominal construction. The translation contains the continuation of the text, to show that the referent of the pronoun does not become the topic of the subsequent text passage. The point that the speaker makes is not so much that something happens to the maize balls, but rather, that sweet food could be prepared without sugar.

(109) *jisa-na=is*$_{i}$ *is* *deretto*$_{j}$, *che* ***isko***$_{\mathbf{j}}$ ***dan-na=is***$_{\mathbf{i}}$
make-DR=3PL.AB ART.PL maize_ball and PRO.3PL.AB chew-DR=3PL.AB
'(They toasted the maize.) They$_{i}$ made maize balls$_{j}$ and **they$_{i}$ chewed them$_{j}$**. (… They chewed them to make them sweet; there was no sugar at that time.)'
[Erlan Rojas 281]

The function of the pronominal construction, then, is to turn a nontopical discourse referent into a sentence topic. It can be depicted as in (110), where "X" stands for the main protagonist in the text passage. At some point, another referent, "Y", is introduced, which only persists during a short sequence of the text. This referent is then taken up anaphorically by a free pronoun in the pronominal construction, after which it disappears again from the story. (Needless to say, this oversimplistic representation does not imply that a new referent is obligatorily expressed in the pronominal construction, that there can be no intervening referents, or that the new referent cannot be taken up at all in the subsequent context.)



(110) X ... X ... X ... Y ... **PRO** ... X ... X ... X …

Thus, what I had analyzed structurally as a cleft does not have the focus-marking function that is part of the typical definitions of clefts (e.g. Lambrecht 2001).

In my paper (Haude 2018b) I maintained the cleft analysis, arguing that in order to define a syntactic construction, one does not need to add pragmatics to the definition. Why should a cleft, i.e. a split sentence, *have* to be a focus-marking device? I rejected the idea to mix a syntactic definition – a cleft as a split sentence consisting of a nonverbal predicate followed by a relative(-like) clause – with a functional one. My reasoning was that by the same token, a passive can be defined syntactically for different languages – e.g. morphological marking of a formerly transitive predicate, resulting in an intransitive clause whose P becomes S and whose A is demoted to oblique status –, while the discourse function of the construction can vary from

one language to another. Therefore, I used the term "topic-marking cleft" in the title of my paper, to propose a separation between syntactic structure and pragmatic function.

6.2.2 Rather: an equational clause

However, while working on this paper (Haude 2018b), I realized that there were several other things that were disturbing about the Movima "cleft". I had too much focused on the pronominal construction with verbs, and neglected the fact that with a noun, the construction is nothing else than an equational clause.

As was outlined in 6.1.2, the construction with a fronted free pronoun is a common way to form an equational clause. It is, moreover, the *only* way to form an identificational clause with a pronominal argument. So, the pronominal construction can host any kind of content word, even those that cannot easily occur as main-clause predicates, such as proper and possessed nouns.

The pronominal construction shares this property with the other relative clause-constructions listed in (104) – the headed RC and the RP. None of them makes a structural difference between a verbal and a nonverbal content word, which can be both mono- and divalent (i.e. intransitive/unpossessed and transitive/possessed).

Moreover, there is no categorial internal difference between nominal and verbal relative constructions. As was already shown with an RP in (54) above, repeated in (111), nouns can be combined with the valency-decreasing particle *kwey/kaw* in these constructions just like verbs. In that case, they refer not to the noun's denotee, but to the possessor.

(111) a. *is* *nono=us*
ART.PL domestic_animal=3M.AB
'his animals' [GBM Ganado 040]

b. *os* ***kaw*** *rey* *no:no* ***n-****isko*
ART.N.PST VALDECR EPIST domestic_animal OBL-PRO.3PL.AB
an owner of the animals (lit.: "an animal owner of them")
[GBM Ganado 004]

Consider now the equational clause with a possessed noun in (112a) and the construction with *kwey* in (112b).

(112) a. ***a'ko*** *asna='ne*
PRO.3N home=3F
'This is her home.' [CVM_020906_1 400]

b. ***i'ne*** ***kwey*** *asna* *ney*
PRO.3F DETR home here
'She is the owner of this house/the one who lives here.' [EAO Agua sucia 020]

Also the negation with *loy* is occasionally found with nouns occurring in relative constructions, as shown in (113) and (114). (All examples of this kind in the corpus are with adjectives, which can be nominalized with the suffix *-łe*; however, there is no evidence that this kind of negation would not be allowed for nouns.)

(113) *is* *motloto-wanra:-ni* ***di'*** ***loy*** *rey* *oro:-łe*
ART.PL earring-INSTR:CLF.NTR-PRC REL NEG.SUB EPIST gold-NMZ.ADJ
'earrings that are not (of) gold' [EAO Aros II 055]

(114) ***asko*** ***loy*** *jayaw-łe*
PRO.3N.AB NEG.SUB nice-NMZ.ADJ
'That's not nice.' [ERM_140806_1 0994]

(115) *ka=s* *rey* *ja'* *sal-na-wa=i*
COP.NEG=DET EPIST just look_for-DR-NMZ.EVT=3PL
kos ***loy*** *rey* *mowi:maj-łe*
ART.N.AB NEG.SUB EPIST Movima-NMZ.ADJ
'The don't just look for someone who is not Movima.' [EAO Tolkosya II 014]

Thus, in the three constructions characterized here as "relative constructions", there is no categorical difference between verbs and nouns. Therefore, the pronominal construction can be simply analyzed as an equational clause, independently of the lexical category of its content word: the pronoun refers to an entity which is specified or catergorized by the content word.

The equational-clause analysis can also account for the information structure of the pronominal construction, which is not that of a canonical cleft. The information structure of an

equational clause with a pronominal subject is neatly described by Himmelmann (2005b: 156) with the example *She is an actress*:

> Such [equational, KH] clauses are clearly bipartite, consisting of a referential expression denoting the entity about which an assertion is made – the subject – and an expression conveying the assertion – the predicate. (Himmelmann 2005b: 156)

It fits well with the equational-clause analysis that in the pronominal construction, the pronoun represents the entity about which an assertion is made, i.e. the topic or theme, and that the content word expresses the assertion, i.e. it is the comment/rheme. If the construction were a cleft (which would correspond to an English *It*-cleft), by contrast, the pronoun would be expected to express the assertion – focusing on a definite entity – and the relative clause would be expected to express a presupposition.

Thus, in the end, it seemed that the whole cleft story was just a huge deviation from a far simpler solution! The construction is a simple equational clause. However, the problem of the pronominal construction is not yet solved in this way. The possible solution is a matter of ongoing research, which I have not yet published on in its own right, and which is discussed in the following section.

6.2.3 An emergent copula? (Ongoing research)

After realizing that I was dealing with an equational clause rather than with a cleft, I wondered, once again, how to deal with the fact that the pronoun is syntactically the predicate. Can a main-clause predicate be the non-assertive part of a sentence at all? It became clear to me that the only way in which the pronoun can be regarded as a predicate is from the perspective that it establishes a relation (see 6.1.3), i.e., that it functions as a copula.

After a literature search on predicatively used pronouns cross-linguistically, I found that indeed, analyzing the free pronoun as a copula is not an implausible solution at all. In the typological literature on nonverbal predicates, pronouns are usually not included, even though they are occasionally analyzed as predicates in descriptions of individual languages (see Schachter & Otanes 1972: 64 on Tagalog). What suddenly caught my attention, however, was that pronouns are known as possible sources of copulas (see Li & Thompson 1977; Hengeveld 1992: 249f.; Katz 1996; Stassen 1997: 90ff.; Korn 2011; Overall, Vallejos & Gildea 2018b). What if the clause-initial free pronoun in Movima is on the way of becoming a copula?

Up to then, I had always claimed that there is no copula in Movima affirmative clauses. When thinking about a possible copular origin of the free pronouns, I had seen it as counterevidence that the last syllables of the different free pronouns look very different from each other: /ne/ on the feminine, /ko/ on all other third-person forms, still different endings in 1st and 2nd person forms. However, maybe these formal variations are not an obstacle to the idea that the free pronoun *functions* as a copula.

A copula is a very particular type of predicate. Not only does it link two syntactic expressions, it assigns one of these expressions the status of predicate. In this way, the copula is only syntactically a predicate, but not in terms of the information structure of the sentence: It does not, by itself, bear the rheme function or convey an assertion, but assigns this function to an adjacent element. Therefore, analyzing the Movima free pronoun as a copula makes it possible to analyze the following content word as the predicate of the construction in terms of information structure.

I still have not presented this idea to the public because I was not sure whether observations about information structure can serve as strong enough evidence for proposing a function to something that is not physically there. I would also reject the idea of assuming a zero copula, since embedding shows that the pronoun itself is the predicate. When I presented my analysis to colleagues of the LABEX project, they suggested that I have a look at identificational clauses containing two nouns with specific referents, such as *My husband is the mayor*. If the pronoun is used between the two nouns, this would be a hint that it is functioning as a copula.

Examples containing a free pronoun in between two nouns are difficult to search for in the corpus: The number of examples that are found in this way is huge, and in most of them maximally one of the two content words has a specific referent, but not both. I know segments of my corpus by heart from having worked with the data for so long, but I realized that I could not think of having come across any such example.

In fact, the sentences with the kind of meaning I was looking for all contain a left-dislocated RP followed by the pronominal construction with an intonation break in between, of the kind *My husband, he is the mayor.* An example is (116). Here, the man referred to by *us itila:kwa* 'the/a man' is a specific man (who was standing in the neigbouring yard at the moment the sentence was uttered). Therefore, the sentence identifies the referent of the left-dislocated RP with the referent of the nominal predicate (*majniwa=kinos Modesta*) in the following pronominal clause.

(116) *us* *itila:kwa,* *u'ko* *majniwa=kinos* *Modesta*
ART.M man PRO.3M offspring_of=ART.F.AB Modesta
'The man, he is Modesta's son.' [EAO Neighbours 015]

The prosodic realization is represented in Figure 29. The intonation break between the left-dislocated NP and the pronominal construction, represented by a comma in (116), is visible from the rising pitch on *itila:kwa* 'man', after which the pitch line is lower and more monotonous. The intonation break is a sign that the utterance is not (yet) a single clause, in which two nominal referents are linked by the copular pronoun.

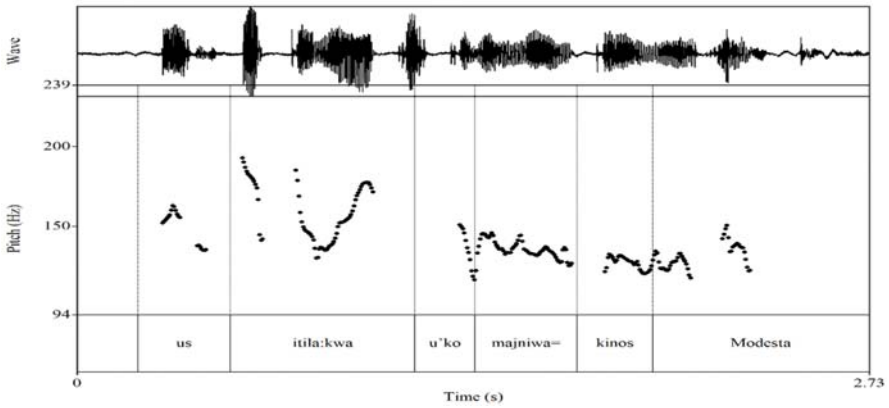


Figure 29. Prosodic contour of (116).

Example (117) is one more illustration of the phenomenon. Here, the intonation break is made even more prominent by a breath intake between the left-dislocated phrase and the pronominal sentence, separating the two intonation units; see Figure 30, where the pause caused by the breath intake is marked with a red circle. As in (116), this is an identificational sentence with a definite specific referent, a man who is visible at the moment of the speech situation. In (117), moreover, the specificity of the referent is indicated by the possessive RP.

(117) *che* *us* *alwaj-a=kine'e=s* *tolkosya,*
and ART.M spouse-LV=ART.F.AB=DET girl
u'ko *kweyninła* *n-as* *kamiyon* *di'* *amme=i*
PRO.3M owner OBL-ART.N truck REL vehicle=3PL
'And the husband of that girl, he is the owner of the truck which is their vehicle.'
[EAO Neighbours 017]

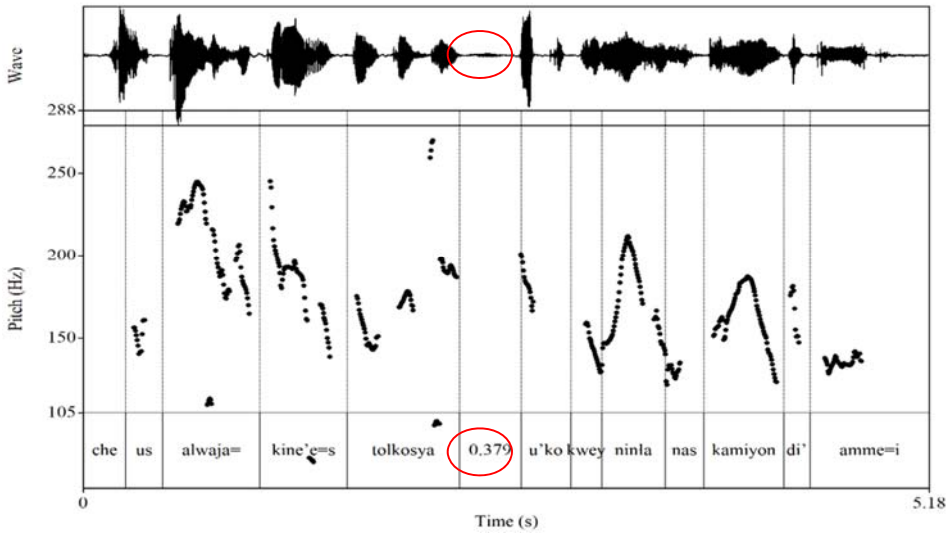


Figure 30. Prosodic contour of (117)

Further support for the hypothesis that the free pronoun is not (yet) a full-fledged copula came from elicitation by telephone. For a sculpture project by the artist Vera Röhm, who collected translations of the sentence "The night is the earth's shadow" in as many languages as possible (https://veraroehm.com/textkuben_en.html), I was asked how to say this sentence in Movima. I did not know, first because I didn't know how the concept "earth" in the sense of "planet" is expressed, but also because I had doubts as to how to express the identity between two terms. I therefore called one of my Movima teachers, Herlan Rojas. His answer confirmed my corpus findings: Such a sentence must be formed with a left-dislocated NP and a resumptive free pronoun.

(118)	*as*	*i:may,*	*a'ko*	*lavayasna=as*	*łakapba*
	ART.N	night	PRO.3N	shadow=ART.3N	earth

'The night, it is the shadow of the earth.'

Thus, Movima does not express identification between two RPs in a simple clause. An identificational sentence is formed with a left-dislocated RP, which is taken up by a pronominal predicate that relates its referent to the subsequent subordinate predicate. The free pronoun is not (yet) a copula, since it is not (yet) able to relate two NPs directly, but rather functions as a resumptive pronoun.

The problem of whether to analyze a construction as monoclausal with a pronominal copula or as a sentence with a left-dislocated RP is not new. A similar situation seems to hold with respect to the third-person pronoun in verbless sentences in Biblical Hebrew, where it has led to much discussion (see Holmstedt & Jones 2014). In any case, in Movima it is clear that the pronoun is formally a predicate, and that it links a topical referent to a lexical description that bears the assertive content. I must therefore argue that the pronoun itself is, just like a copula, a predicate that attributes the assertive function to the subsequent content word. This is work in progress.

Summing up, analyzing the initial free pronoun as an emergent copula seems to me the only analysis that does justice to the syntactic, pragmatic, and prosodic properties of the pronmoinal construction. My struggle with an apparently simple construction – a content word preceded by a free pronoun –has taken several detours over the years before I arrived at this final analysis, which are recapitulated in Figure 31.

FRONTING (but the pronoun is the predicate) → CLEFT PHRASE (but no argument focus) → EQUATION (but the pronominal predicate is not assertive) → COPULA (but does not link two specific nouns) → **EMERGENT COPULA**

Figure 31. The detours that led to the emergent-copula analysis

6.2.4 The nominal-predicate construction and the "real" cleft

Haude, Katharina. 2019c (in press). Clefting and nominal predication: two focus-marking constructions in Movima. *Linguistic Discovery* 17(1). 106-127. (Special issue on *Clefts and Related Focus Constructions*, guest-edited by E. Palancar and M. Vanhove.)

At the workshop on *Clefts and related focus constructions* that I co-organized with Enrique Palancar and Martine Vanhove, I presented my "topicalizing cleft" again, also in order to defend the view that clefting can be fruitfully regarded as a syntactic phenomenon that does not necessarily mark focus. Since I had already published on this, at first I felt unable to

contribute anything to the volume ensuing from the workshop (especially since its title was going to be *Clefts and related focus constructions*, and I did not share the view that a cleft should be defined in functional terms).

But then – on a railway station waiting for the train – I started to think about *how* one would express argument focus in Movima. I mentally went through the possible constructions in my corpus, and discovered that these constructions were formed through the combination of an equational clause with a pronoun (see 6.1.2) and a nominal predicate with a verbal RP (see 4.2.2 above), with the following result:

(119)	*asko*	*rulrul*	*os*	*man<a>ye=is*	*pa:ko*
	PRO.3N.AB	jaguar	ART.N.PST	encounter<DR>=ART.PL	dog
	'It was the/a JAGUAR (what) the dogs had encountered.'				[EAO Jaguar 085]

So, it turned out that I did have highly adequate material for a contribution to this volume. Before explaining this newly discovered cleft construction in 6.2.4.2, in 6.2.4.1 I will discuss the predicate-nominal construction with a verbal RP, which is a focus construction in its own right.

6.2.4.1 The predicate-nominal focus construction

Even though common nouns can freely function as predicates and verbs can occur in referential phrases (see 4.2.2), also in Movima the prototypical predicate is a verb, and a prototypical referential phrase contains a noun. Constructions that deviate from this pattern either represent a special construction type (e.g. an equational clause), or they are pragmatically marked. Pragmatically marked constructions convey the same propositional content as the corresponding "basic clause" (treated in Chapter 4), but are arranged in a way that orients the hearer more explicitly towards the information that the speaker considers important.

Actually, "verbal RPs" occur most frequently in clauses with a predicate nominal, and only rarely with a verbal predicate. There is a pragmatic effect of placing a noun in predicate and a verb in argument function: The construction places the focus on the participant in the situation described by the verb. Compare the basic clause in (120), which consists of a verbal predicate and a nominal argument, with (121), where the two lexical categories have "swapped" their syntactic status.

(120) *man<a>ye=is* *pa:ko* *os* *rulrul*
encounter<DR>=ART.PL dog ART.N.PST jaguar
'The dogs encountered the/a jaguar.' [EAO Tigre y perro 003]

(121) *rulrul* *is* *man<a>ye=as*
jaguar ART.PL encounter<DR>=3N.AB
'It (the dog) encountered JAGUARS.' [EAO_tigreyperro_150808 149]

The construction in (121) is the result of a syntactic rearrangement of the lexical categories of the basic clause: The noun (*rulrul*) is the predicate, and the verb with its internal argument is placed in the external argument phrase (*is man<a>ye=as*). In my paper I labelled this construction "nominal-predicate focus construction", which indicates that the predicate is a noun and that the construction as a whole is a focus construction. (The term "argument focus" would be misleading, since the element in predicate position is not a syntactic argument).

The pragmatic markedness of this construction results, in part, from the predicative use of the noun. Launey (2004: 50) describes a parallel pattern in Classical Nahuatl, a language he analyzes as an "omnipredicative" or "rheme-dominant", as follows: "the noun [in the nominal-predicate construction, KH] is the center of information, the focus or, in the European tradition, the rheme, …: that which is said about something or someone". The other crucial feature of this construction is the placement of the verb in the argument RP. It is, therefore, not the nominal predicate alone that triggers the focus effect, but the *inversion* of the prototypical association between lexical categories and syntactic-pragmatic functions, i.e. the association of verbs with predication and of nouns with reference (see Croft 2003: 185).

For characterizing the predicate-nominal focus construction, it is once again helpful to look at descriptions of Tagalog. According to Kaufman (2018: 233), "the predicate position in Austronesian languages functions as a kind of de facto focus position by virtue of Austronesian languages tending to package presuppositions as subjects". In Movima, the "subject" is the external argument, corresponding to the Tagalog *ang*-phrase (although unlike the latter, it is not necessarily definite), and the strategy is exactly the same: by placing the verb in subject position, the situation it denotes is presented as a presupposition.

Therefore, the focus construction in (121) is a simple clause, whose focus-marking effect results from the non-prototypical employment of the noun as predicate and of the verb as part

of the argument phrase. Figure 32 shows that the focus effect is also reflected prosodically: The nominal predicate receives high pitch, while the verbal RP is prosodically nonprominent.

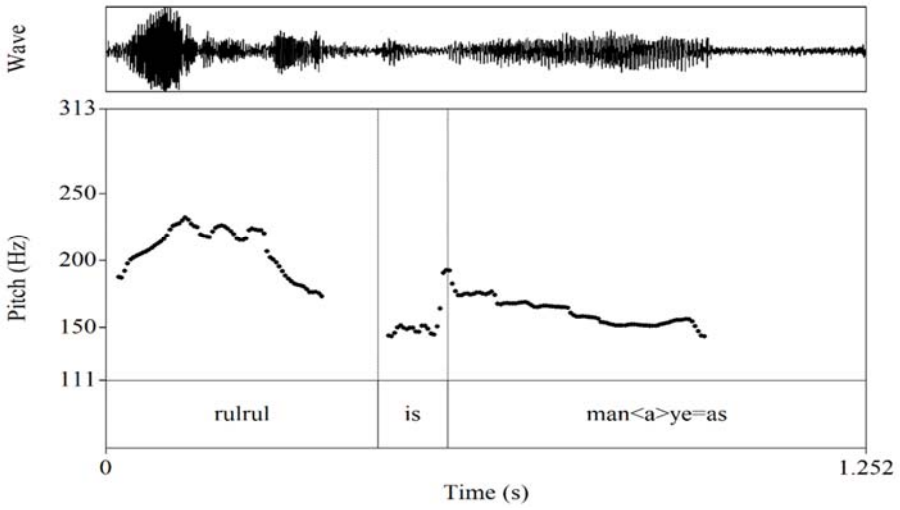


Figure 32. Prosodic contour of (121)

6.2.4.2 The "real" cleft

The construction which, unlike the pronominal construction discussed in 6.2.1, can undoubtedly be analyzed as a cleft is illustrated in (122). It looks very similar to the nominal-predicate focus construction (121), repeated in (123).

(122) *asko* *rulrul* *os* *man<a>ye=is* *pa:ko*
PRO.3N.AB jaguar ART.N.PST encounter<DR>=ART.PL dog
'It was the/a JAGUAR (what) the dogs had encountered.' [EAO Jaguar 085]

(123) *rulrul* *is* *man<a>ye=as*
jaguar ART.PL.PST encounter<DR>=3N.AB
'It (the dog) encountered JAGUARS.' [EAO_tigreyperro_150808 149]

The only formal difference between these two constructions is that the cleft contains an initial free pronoun. This pronoun forms an equational clause with the nominal predicate: *asko rulrul* 'It is/was the/a jaguar'. The equational clause is followed by a verbal RP: 'the dogs' found (thing)' or 'what the dogs (had) found'.

In the past, I had never found a good explanation for the presence of the free pronoun in constructions like (122) (see Haude 2006: 291–292). The construction looks like a hybrid of two sentences formed ad-hoc by an undecided speaker: a), there is an equational clause with a free pronoun (*asko rulrul*), b), there is a predicate-nominal focus construction (*rulrul os man<a>ye=is pa:ko*). The two constructions share the noun, which is a subordinate predicate in a) and the main-clause predicate in b). Furthermore, the RP can be analyzed in different ways: as the argument of the predicate nominal or as an attribute to the pronominal subject ('It, what they found …'). (This problem is also present in the analysis of the relative clause of clefts in other languages.)

While working on the paper for the "Cleft" volume (Haude 2019c), I realized that independently of how one analyzes the internal structure of the construction in (122), there is a slight pragmatic difference between predicate-nominal focus construction and the cleft construction. In the predicate-nominal focus construction, (121), the focused noun cannot be definite or even specific: it is not referential at all. What the speaker intended to say when uttering (121) was that among all the wild animals that can be encountered during a hunt, this dog was particularly apt at finding jaguars. This sentence, therefore, answers the hypothetical question "What (kind of) animals did the dogs encounter?" Its focus is on the class membership of the event participant specified by the RP (i.e. the undergoer of the direct-marked verb).

The sentence in (122), by contrast, is internally complex. Its matrix clause is composed of a pronoun and a nominal predicate, which together constitute a full-fledged equational clause: 'It is/was the/a jaguar.' The argument phrase containing the verb corresponds to the relative (or "relative-like", Lambrecht 2001: 467) clause that is included in most, if not all, definitions of clefts proposed in the literature.

The context of the cleft sentence in (122) is provided in (124) below. It is an excerpt from a personal anecdote in which the speaker, a young woman at that time, and her husband go hunting. At first they do not find anything, but all of a sudden, the dogs start to run and bark, which is a clear sign that they have spotted an animal. With the cleft sentence rendered in boldface in (124), which is the translation of (122), the speaker points out that it was a *jaguar* that the dogs had found, and not any other animal. The sentence is a kind of meta information to the hearer, anticipating what the protagonist herself does not yet know at that moment in the story: She believes that the animal is an anteater. Hence, this is a case of narrow focus: The cleft picks out one among several potential alternatives to the participant that might be the patient of the 'encounter' event, thereby answering a question like "What animal did the dogs encounter?" Like the relative clause of clefts in other languages, the RP containing a verb

phrase expresses a presupposition, since the fact that the dogs are barking indicates that they must have encountered a wild animal.

(124) ‘We hadn’t walked far when he saw the tracks of the/a jaguar. At that moment the dogs started running and barked, “*Wauwauwau!”*. And then it screamed as well, it made “*Yeyeye!*”. And then the jaguar fled. **It was the/a JAGUAR what the dogs had encountered, madam.** Then I said, “What may the dogs have encountered?”, I said to him. “It sounds like an anteater,” I said. Because it roared like an anteater.’
[EAO Jaguar 081-088]

The intonation contour of the cleft in (122) is illustrated in Figure 33. The focused noun receives high pitch, while the rest of the utterance is prosodically nonprominent.

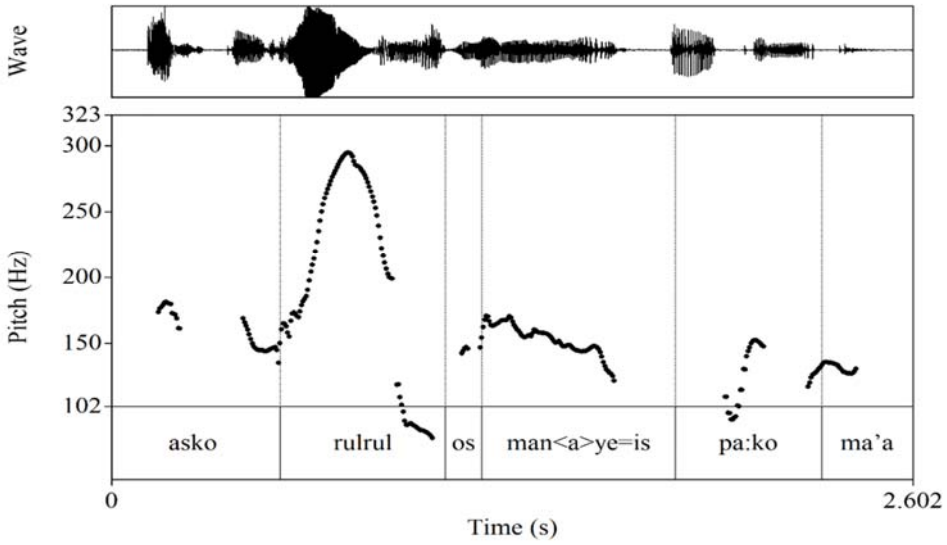


Figure 33. Prosodic contour of (122)

Thus, the cleft brings an event participant into focus, both in the sense of selecting from potential alternatives and in the sense of marking a contrast. The RP, by contrast, contains presupposed information: In the present example, for instance, it is clear from the behaviour of the dogs that they have encountered a wild animal.

6.2.4.3 Summary: the two focus constructions

To sum up, Movima has two syntactic focus constructions, which, on first sight, look very similar. They both contain a nominal predicate and a verbal argument RP, i.e. an RP that contains a subordinate predicate characterizing the referent. One of these constructions – the predicate-nominal focus construction (6.2.4.1) – is a simple intransitive clause, whose pragmatic markedness stems from the fact that the prototypical association of the information-structural categories comment (predicate) and topic (argument) with the lexical categories verb and noun, respectively, is inverted. As a result, the comment is a characterization of the participant involved in a given event, and not, as in a basic (verbal) clause, a description of the event in which a given participant is involved. The other construction is a cleft (6.2.4.2), made up of an equational clause with a pronominal predicate (*It was a jaguar*) representing the clefted constituent and an RP conveying the presupposed information (*what the dogs had encountered*). The cleft is represented schematically in Figure 34.

PRO	N	ART	V
'It's N		the (entity that) Vs'	

Figure 34. The cleft construction

Based on their use in discourse, the functional contrast between the predicate-nominal focus construction and the cleft can be understood as corresponding to categories that are well-established in the literature: the difference between predication and specification (Declerck 1988), or of predicate focus and argument focus (Lambrecht 1994). In Lambrecht's (2001: 485) terms, the predicate (or "unmarked") focus has the function of "predicating a property relative to a given topic (… also called 'topic-comment' or 'categorical' function)". The argument focus, in contrast, has the function of "identifying or specifying an argument in a presupposed open proposition (… also called 'focus-presupposition,' 'specificational,' 'identificational,' or 'contrastive' function)" (Lambrecht 2001: 485). It is the latter function that is (or can be) expressed by a cleft in many languages, while the former is expressed by a simple predication. As I showed in the paper summarized here, the Movima nominal-predicate focus construction is a predicate focus, while the construction with the free pronoun has an internally complex, biclausal structure; this construction can without any problem be characterized as a cleft, as it also shares with clefts in other languages the pragmatic function of marking argument focus.

7. The referential phrase: nominal tense and subordination

In the above sections 4–6 I summed up the central aspects of my research so far, which centered around the expression of "who acts on whom" in Movima. Occasionally, and more or less in parallel to this ongoing research, I also worked on the role of referential phrases (RPs) in the sentence and in the discourse context, as well as on subordination and complex sentences.

One of the first topics I investigated in this respect, already early in my post-doctoral career, was the tense-marking function of the determiner. The determiner is an obligatory and defining part of an RP. Usually it is an article, but it can also be a demonstrative (a category I did not specifically work on after my doctorate). The forms of the article and the categories marked by it are listed in Table 15.

Table 15. The Movima articles

	presential/generic	absential (AB)	past (PST)
human male (M)	*us*	*kus*	*us*
human female (F)	*i'nes*	*kinos*	*isnos*
non-human (N)	*as*	*kos*	*os*
plural/mass (PL)	*is*	*kis*	*is*

The article indicates +/- humanness, sex, number, and presence or absence of the referent, as well as actual vs. ceased existence of the referent. In this way, it is the most important tense-marking element in a Movima text.

The article is also one of the most frequent elements in a Movima text. One reason is that, unlike many other Amerindian languages with their polypersonal verbs, Movima is not hesitant to express arguments by full referential phrases (see 5.1). Another reason is that adverbial and complement clauses have the form of RPs. Therefore, in discourse, reference is made constantly to different kinds of referents and their spatio-temporal location, and this information also contributes to the temporal interpretation of the events narrated in a text. In 7.1, I report on my work on the tense-marking function of the article, where I discussed the "independent" vs. "propositional" tense-marking functions (7.1.1) as well as the way in which the article's tense-marking potential is influenced by the type of concept the RP refers to

(7.1.2). In 7.2, I explain that partly (though not only) because of the tense-marking function of the article, RPs representing complement and adverbial clauses can be regarded as containing more grammatical information than main-clause predicates.

7.1 "Nominal tense": tense as a referential category

Haude, Katharina. 2010b. 'She kisses her late husband' = 'she kissed her husband': nominal tense marking in Movima, in: Cysouw, Michael and Jan Wohlgemuth (eds.). *Rara & Rarissima: documenting the fringes of linguistic diversity*, 95-116. Berlin/New York: Mouton de Gruyter.

Haude, Katharina. 2011c. Tense marking on dependent nominals in Movima, in: Musan, Renate and Monika Rathert (eds.). *Tense across Languages*, 189-206. Berlin/Boston: De Gruyter.

Tense is not a verbal category in Movima. It is indicated by dedicated particles, listed in (125), which usually occur once in a text or text passage to signal the time of the narrated events.

(125)

kwil	a long time ago
la'	the previous occurrence of the event
kwey	on the same day, but before the time of speaking
loy	directly after the moment of speaking or a future reference time
nokowa	after the moment of speaking

Tense particles can occur basically anywhere in the sentence, even inside a referential phrase, as shown with the particle *la'* 'anterior' in (126). Example (127) illustrates the hodiernal-past tense-marking particle *kwey* ('earlier on the same day') preceding a main-clause predicate. (This particle is homophonous with the variant *kwey* of the valency-decreasing particle described above, which, however, does not occur in a main clause.)

(126)

jo'yaj--us	*n-os*	***la'***	*walaylo*
arrive--3M.AB	OBL-ART.N.PST	ANT	afternoon

'He arrived yesterday afternoon.' (lit.: "on the previous afternoon.")

(127)

kwey	*joy-cheł*	*us*	*pa:pa=kinos*	*majni=Ø*
HOD	go-R/R	ART.M	father_of=ART.F.AB	child=1SG

n-as *ra:diyo*
OBL-ART.N radio
'The father of my daughter went to the radio (station) (earlier today).'

Apart from that, tense is indicated by the articles. In my grammar I already illustrated the semantic, spatial and temporal properties indicated by the article in much detail. I reported on the tense-marking function also right after finishing my thesis at a workshop on *Rara and Rarissima* in Leipzig in 2006 (Haude 2010b), when I considered the tense-marking function of the article as the most striking one of all the typologically unusual features of Movima, and at a workshop *Tense across languages* that took place at the DGfS conference 2008 in Bamberg (Haude 2011c). My analyses are recapitulated in the following subsections. In 7.1.1, I illustrate that Movima nominal tense marking basically indicates the temporal properties of the referent, but that through implication, this usually has an impact on the temporal interpretation of the proposition as a whole. In 7.1.2 I show that the possibility of the article to make temporal distinctions also depends on the type of referent. In particular, complement and adverbial clauses are expressed as RPs in Movima, and these are highly sensitive for marking temporal categories.

7.1.1 Independent and propositional nominal tense marking

Already while working on my dissertation I had published a paper on nominal tense marking in Movima (Haude 2004) and showed that the three spatio-temporal categories of the articles, 'presential', 'absential', 'ceased existence', apply to deictic properties of the referent, which is either visibly (or otherwise perceivably) present at the speech situation, absent from it, or not existing anymore. These three categories are illustrated with the examples in (128).

(128) a. *kay-a:-poj* ***as*** *pa:ko*
eat-DR-CAUS ART.N dog
'I fed/am feeding/will feed the dog (present).'

b. *kay-a:-poj* ***kos*** *pa:ko*
eat-DR-CAUS ART.N.AB dog
'I fed/will feed the/a dog (absent).'

c. *kay-a:-poj* ***os*** *pa:ko*

eat-DR-CAUS ART.N.PST dog
'I fed the/a dog (that is now dead and gone).'

An important concept regarding nominal tense has been introduced in the seminal paper by Nordlinger and Sadler (2004): the difference between "independent" and "propositional" nominal tense. Independent nominal tense means that the temporal marking of the nominal constituent locates the time at which the property denoted by the noun holds of the referent (in some languages, this may involve the functioning or the possessed state of the referent). In propositional nominal tense, the temporal marking of the nominal serves as a tense marker for the entire proposition.

The examples in (128) above show that in Movima, the article indicates the actual existence or ceased existence of the referent (here, a dog). At the same time, however, the choice of the article influences the temporal interpretation of the entire clause: nonpast in (128a)-(128b), past in (128)c. However, this is not evidence for propositional tense marking: The past interpretation of a clause like (128)c is caused by the fact that there is no contradictory element marking a different tense in the main clause, and because it is most common to talk in a past context about entities that have ceased to exist. So far, the temporal interpretation of the proposition is an effect of implicature only and should not be considered tense marking (Comrie 1985).

Evidence for independent nominal tense marking comes from the possibility to cancel the implicational effect of the article on the clause, as in (129) (see (136) for a similar example). The clause contains an RP with a past article, but it also contains a TAM particle implying future reference (*loy*). The past article indicates ceased existence of the referent, but this is independent of the interpretation of the clause as a whole. This and similar examples, therefore, identify the past article as a marker of independent nominal tense.

(129) ***loy*** *it* *to'baycho-poj-cheł* *n-**isnos*** *nonok=Ø*
ITN 1INTR remember-CAUS-R/R OBL-ART.F.PST grandparent=1SG
'I'll remember my late grandmother.' [elicited]

However, examples like (129) are rare, and the implicational power of the tense-marking function of the article cannot be underestimated. Moreover, speakers have a certain degree of freedom in choosing a past or nonpast form of the article when narrating past-tense events. Especially absent non-human referents, whose actual existence at the time of speaking is often not considered relevant, tend to be referred to according to the discourse tense. For instance,

example (130) illustrates the use of the past-tense article with reference to a spider that has explicitly not been killed (the event occurred just the day before) and may thus still be alive. However, the past article is used because what is relevant to the speaker is to maintain temporal reference in discourse, and not the possible ongoing existence of the spider.

(130) *ka=s* *rey* *tikoy-na:-wa=Ø--as,* ***os*** *si:wa* *merek*
COP.NEG=DET EPIST kill-DR-NMZ=1SG--3N.AB ART.N.PST spider big
'I didn't kill it, the big spider.' [EAO araña 009]

In principle, therefore, nominal tense marking in Movima is of the independent type, the choice of the article depending on deictic properties of the referent and having only an implicational effect on the temporal interpretation of the clause. However, depending on the relevance of the referent and on the priority the speaker may give to the marking of discourse tense, the past article can also be used to mark discourse tense directly, thereby functioning as a marker of propositional nominal tense. It is therefore possible to speak of a gradual shift of nominal tense marking in Movima from purely independent towards propositional. This can be represented as in Figure 35.

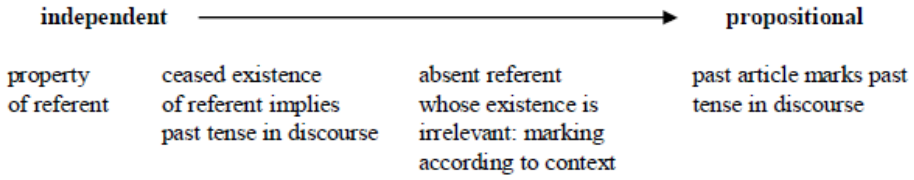


Figure 35. The shift from independent to propositional nominal tense marking

In Figure 35, prototypical independent nominal tense marking is given on the left-hand side. Here, a temporal property of the referent is marked, independently of the discourse tense. The fact that this marking usually has an effect on the interpretation of the proposition as a whole (as shown in (128)) is a step towards propositional temporal marking. When entities are referred to by the past form, even though they are still in existence and even present or relevant to the speaker, the article does not indicate a temporal property of the referent anymore. It serves as a propositional nominal tense marker that provides temporal information of the clause as a whole. This shift can be seen as an instance of what Traugott has termed the

"subjectification" effect in grammaticalization (Traugott 1995; 1999): the article is used to express the speaker's attitude towards the propositional content, and through conventionalization of this use, the article may grammaticalize into a marker of propositional tense.

7.1.2 Absolute concepts, concrete entities, and states/events

The possibility to indicate temporal distinctions with the article in Movima also depends on the type of referent. Three types of referents are distinguished by the spatio-temporal categorization of the Movima article, as will be detailed below: "absolute concepts", "concrete concepts", and "times/events". The likelihood of their being marked for past tense is depicted in Figure 36.

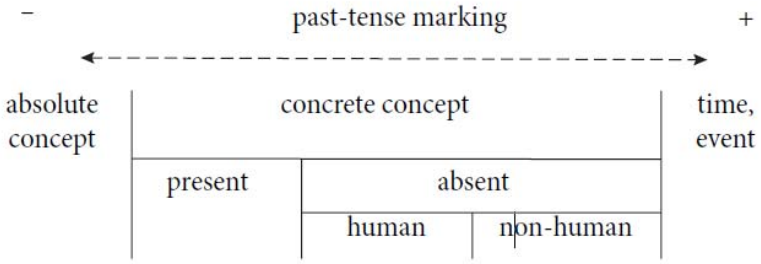


Figure 36. Likelihood of past-tense marking of referents in a past-tense context (Haude 2004: 88)

Absolute concepts are those concepts that are not expected to change over time, like fixed locations or habitually recurring situations. These are always referred to with the presential article. In (131), for instance, the town Trinidad – which is not the place where the sentence was uttered – is referred to with the presential article, while the person that is in that town is referred to as absent (in this extract, with the absential form of the pronoun).

(131) *bo* *rey* *isne* *kino'* ***n-as*** *Tirinra*
REAS EPIST PRO.3F.AB DEM.F.AB OBL-ART.N Trinidad
'Because she (absent) is in Trinidad.' [EAO Barredoras 024]

Concrete concepts are entities such as persons, animals, or objects. RPs referring to concrete concepts contain a noun denoting the entity, an adjective describing it (e.g. 'big (one)'), or a

verb attributing to it a particular role in a particular event (e.g. 'his watched (one)'; see 4.2.2, example (51)). RPs referring to concrete entities indicate the referent's presence at or absence from the speech situation, or their ceased existence, as was illustrated in (128) above.

Points in time (e.g. *ima:yoj* 'morning'), events ('walking'), and states ('being a child') can be characterized as non-time-stable referents. RPs referring to times, events, or states indicate the temporal rather than spatial location of their referents: occurrence in the present or future, in the past on the same day, or in the past before the same day. When referring to a time or event/state, the "absential" article marks "hodiernal past", i.e. a time span that occurred on the same day, but before the moment of speaking. This is shown in (132), where the absential article is used in the temporal adverbial clause 'in the morning'. By contrast, in (133), the same adverbial clause contains the past article, which signals a point in time before the day of speaking; and the presential article in the two adverbial clauses in (134) indicates that the event takes place in the future, i.e. on the next day.

(132) *jayna* *kino'* *chi:~chi* *no-**kos*** *ima:yoj*
DSC DEM.F.AB MD~go_out OBL-ART.N.AB morning
'She (absent) has gone out (today) in the morning.' [EAO Llegada hija 003]

(133) *jo'yaj--isne* *n-**os*** *ima:yoj*
arrive--3F.AB OBL-ART.N.PST morning
'She (absent) arrived (on some other day) in the morning.' [MCA_280806_2 397]

(134) *n-**as*** *tawakni* *n-**as*** *ima:yoj* *jop<a:>ye=Ø* *i:ri*
OBL-ART.N next_day OBL-ART.N morning send<DR>=1SG DEM.PL.SPK
'Tomorrow in the morning I'll send these.' [ERM_140806_2 095]

The Movima article provides temporal information consistently – even if through implicature – also because complement and adverbial clauses, already introduced in 6.1.1 (examples (76) and (77)), have the form of RPs. These RPs refer to events or states, which are highly sensitive for temporal location. Since states and events are inanimate, these RPs always take the neuter article, whose three forms (*as* 'presential', *kos* 'absential', *os* 'past') here distinguish three temporal categories rather than marking spatial location. Unless cancelled by some overt device in the main clause (see 7.1 below), the main-clause tense is interpreted as being identical to that of the dependent clause; this is to say, the tense of the dependent clause

has scope over the main clause.

This effect can be observed in (135a)-(135c), examples that were provided spontaneously by my main Movima teacher when she realized what I was looking for (note that the reference to the 'fence', *as chakdi*, remains constant).

(135) a. *jayna* *t* *ba:lomaj* ***n-as*** *ji:sa-na:-wa=Ø*
DSC 1INTR finish OBL-ART.N make-DR-NMZ.EVT=1SG
as *chakdi*
ART.N fence
'I'll finish making the (present) fence.' (lit.: "with my making the fence")
(I'm still building it)

b. *jayna* *t* *ba:lomaj* ***no-kos*** *ji:sa-na:-wa=Ø*
DSC 1INTR finish OBL-ART.N.AB make-DR- NMZ.EVT=1SG
as *chakdi*
ART.N fence
'I just finished making the (present) fence.' (today)

c. *jayna* *t* *ba:lomaj* ***n-os*** *ji:sa-na:-wa=Ø*
DSC 1INTR finish OBL-ART.N.Pst make-DR-NMZ.EVT=1SG
as *chakdi*
ART.N fence
'I finished making the (present) fence.' (before today)

That the past-tense form of complement and adverbial clauses signals a temporal location independent from the tense of the main clause is shown in (136). Here, the main predicate is in the imperative form, thus implying that the event will take place in the future.

(136) *ajlomaj-**ti*** ***os*** *naye-wa=n*
tell_about-IMP.DR ART.N.PST marry-NMZ.EVT=2
'Tell (her) about your (past) marriage!' [EAO Jovina 001]

The three-way temporal distinction in RPs representing complement and adverbial clauses can be explained by the fact that these RPs refer to states or events, i.e. concepts that in principle

do not have a spatial location and that are not time-stable. Therefore, the article can be employed here unambiguously for temporal instead of spatial deixis, and the absential article can be used to introduce an additional temporal category.

At the same time, there is a direct parallel with RPs denoting concrete entities, whose referent must have ceased to exist in order to be referred to with the past article. In the case of times and states/events, the absential and past forms can be used only when the situation is concluded. A situation that has started in the past but is continuing in the present is always referred to with the presential article. The spatio-temporal categories encoded by the article are summed up in Table 16.

Table 16. The spatio-temporal categories encoded by the article

article form	**entity**	**time/situation**
"presential"	present "absolute" (e.g. fixed location)	present future "absolute" (habitual)
"absential"	absent, in existence	concluded earlier on the same day
"past"	absent, out of existence (before the day of speaking)	concluded before the day of speaking

If the notion of tense is taken as a deictic category of temporality, the Movima article system is a tense-marking system in that it marks the relation between two times. It does this in analogy to spatial deixis by categorizing the physical and/or temporal presence and accessibility of a referent in the perception of the speaker. With the example pf the neuter article, which displays most distinctions because it is combined with concrete as well as with abstract nouns, this is schematized in Figure 37.

"presential"	"absential"		"past"
+ present	+ present	- present	- present
+ accessible	- accessible	+ accessible	- accessible

Figure 37. The neuter article as a marker of perceived presence and accessibility (Haude 2004: 88)

This perspective makes it possible to appreciate one of the many different strategies that languages use to encode temporal relations, and thereby provides more insight into the ways in which language organizes cognitive concepts.

7.2 Some unusual properties of Movima subordinate clauses

Haude, Katharina. 2011d. Referring to states and events: subordination in Movima. In *Subordination in South American Languages*, van Gijn, Rik, Katharina Haude, and Pieter Muysken (eds.), 141-168. Amsterdam/Philadelphia: Benjamins.

As we saw above, being referential phrases, Movima complement and adverbial clauses provide temporal information consistently, while Movima main-clause predicates do not. This is challenging for linguistic theory, since it is assumed that universally, subordinate clauses provide a background to the main line of argumentation expressed by main clauses, and therefore, contain reduced information (Hopper & Thompson 1984; Cristofaro 2003). In the paper summarized here I argued that also beyond tense marking, Movima subordinate clauses do not seem to contain reduced information.

First of all, subordinate predicates retain their argument structure: in a subordinate transitive clause, just as in a main clause, both arguments may be overtly expressed. Direct and inverse marking are retained, as illustrated in (137) and (138), respectively. (In nominalized forms, the inverse is represented by reduplication, see Haude 2014b, or zero, to which I will get further below).

(137) *yey-na=Ø* *[as* *visitar-na:-wa=Ø* *kus*
want-DR=1SG ART.N visit-DR-NMZ.EVT=1SG ART.M.AB
alkaka:ye=Ø]
relative=1SG
'I want to visit my relative.' (lit.: "I want my visiting my relative.")
[EAO Visita 047]

(138) *isko* *ka=s* *jiwa-wa=is* *[n-as* *rey*
PRO.3PL.AB COP.NEG=DET come-NMZ.EVT=3PL OBL-ART.N EPIST
vel~vel-wa=n--is]
INV~look_after- NMZ.EVT=2--3PL
'They don't come to see you, you know.' (lit.: "They, their coming for you being seen by them is not.") [EAO In between 090]

Second, nominalized predicates are marked as possessed. This means that subordinate intransitive clauses, in contrast to main intransitive clauses, always contain an overt argument expression. In (139), this overt argument encoding on the embedded form *joy-wa=sne* 'her going' clearly disambiguates between the main-clause subject (first person singular) and the subject of the embedded clause (third person singular feminine). In (140), by contrast, the unexpressed subject of the main verb *joy-cheł* must be retrieved from the context. (Here, the unexpressed main-clause argument is coreferential with that of the subordinate clause, but this is not necessarily the case.)

(139) *jayna* *yey-na=Ø* *as* *joyaj-wa=sne*
DSC want-DR=1SG ART.N arrive-NMZ.EVT=3F.AB
'Now I want her to come.' [ERM_140806_1 0202]

(140) *joy-cheł* *n-os* *te:lo-wa='ne* *nosdé*
go-R/R OBL-ART.N.PST dance-NMZ.EVT=3F there
'(She) went to dance there.' (lit.: "(She) goes for her past-dancing there.")
[EAO In between 012]

Furthermore, there are two types of "nominalization" in embedded clauses, already introduced in 6.1.1: a suffix *-wa*, which combines with verbal and nominal bases, and reduplication, which occurs with nouns and adjectives. These two processes can be taken as a criterion for distinguishing between nouns and verbs, as I have done in 6.1.1; however, they can also be interpreted as marking the distinction between two lexical aspects (*Aktionsarten*): The suffix *-wa* marks a predicate denoting an event or temporal state, and reduplication marks the predicate as denoting an existential state. This reasoning is possible because also nouns can receive the suffix *-wa* to denote an event associated with the denotee of the noun. For instance,

the noun *wa:ka* ‘cattle, cow, meat’, when suffixed with *-wa* in an embedded clause, means ‘slaughtering’, as in (141).

(141) *joy-a-łe=us* *is* *wa:ka*
go-DR-CO=3M.AB ART.PL cow
bo *os* *wa:ka-wa=us*
REAS ART.N.PST cow-NMZ.EVT=3M.AB
‘He brought (the) cows in order to slaughter (them).’ [EAO Cbba 137]

Here, we can assume that the nominalizer *-wa* itself induces the event reading, as suggested by the gloss in (141). Alternatively, we can postulate a preceding zero verbalizer. This would be plausible as well, since the zero marker would correspond to the verbalizer *-tik*, which never occurs before further suffixation. Consider the main-clause form of the verb ‘slaughter’ in (142a) and the nominalized form in (142b) analyzed as derived from this verb.

(142) a. *wa:ka:**-tik**=us*
cow-VBZ=3M.AB
‘He slaughters.’

b. *bo* *os* *wa:ka**-Ø**-wa=us*
REAS ART.N.PST cow-VBZ-NMZ.EVT=3M.AB
‘for his slaughtering’

Evidence for a possible aspectual meaning of the two nominalizers, however, comes from some verbs that denote states, such as *dol-mi* ‘full of water’, (143a). These words do not occur with a verbalizer *-tik* (in order to denote an associated event, these words take a direct or inverse marker). In the same way as a noun, a word of this type can undergo either suffixation of *-wa*, as in (143b), or reduplication, as in (143c). The type of derivation clearly marks the difference between an event and a state.

(143) a. *dol-mi*
full-CLF.water
‘(It is) full of water.’

b. *n-as* *dol-mi-**wa**=a*
OBL-ART.N full-BE.water-NMZ.EVT=3N
'when it has been filled with water' [elicited]

c. *n-as* *dol-**<mi~>**mi=a*
OBL-ART.N full-<NMZ.ST~>BE.water=3N
'when it is full of water' [elicited]

Whichever the analysis, these morphemes make a distinction in subordinate clauses that is absent from main-clause predicates. Reduplicated embedded predicates denote existential states, while words suffixed with *-wa* denote events and non-time-stable states. Thus, independently of the analysis, all subordinate predicates overtly make a distinction of lexical aspect, something that is not the case with main-clause predicates.

The only grammatical category that is marked less consistently in embedded clauses than in main clauses is voice: on some verbs, the nominalization (just like other suffixes) cancels the expression of reflexive/reciprocal, middle, and inverse voice, so there can be an ambiguity between an intransitive and an inverse transitive reading. Example (1) from Section 3.3 above is repeated in (144): Here, only the presence of two overt arguments indicates that we are dealing with a transitive verb.

(144) *bo* *os* *kay-poj-wa=y'ɬi--k-isne* *jayna*
REAS ART.N.PST eat-CAUS-NMZ.EV=1PL--OBV-3F.AB DSC
'… so that we could (be) fed (by) her then.' [EAO In between 206]

The categories that are overtly encoded (or not) in main vs. subordinate clauses are summed up in Table 17Table 17.

Table 17. Categories encoded in main vs. subordinate clauses (Haude 2011d)

Main clause	**Category**	**Embedded clause**
–	Obligatory encoding of single argument of intransitive clause	+
–	Overt marking of lexical aspect (event/temporal state vs. existential state)	+
–	Consistent tense marking	+
+	Two overt core arguments possible	+
+	overt voice affixes	–

One reason for the high degree of overt encoding of information on person, aspect and tense in subordinate clauses is that many features that are generally considered to be typical of main-clause predication and represented by finite verb forms, are associated with subordinate predication in Movima, which occurs in the form of referential expressions. The obligatory encoding of the argument of an embedded intransitive clause is due to the fact that subordinate clauses appear as possessed RPs.

The encoding of temporal deixis is not a morphological feature of verbs, but a property of referential elements (in particular, articles) in Movima. The fact that subordinate clauses have the form of referential phrases explains why tense marking is consistently marked on subordinate, but not on main clauses. Furthermore, due to the fact that subordinate predicates do not denote concrete entities, but states and events, the article does not encode spatial deixis in these RPs and can make more fine-grained temporal distinctions than with nouns denoting concrete entities.

Furthermore, on the semantic side, we find that subordinate predicates denote situations, whereas RPs containing underived verbs refer to participants in situations (see 4.2.2). Thus, when considered from a typological perspective, we find a paradoxical situation in Movima: subordination is carried out in the form of a referential phrase, which, together with morphological marking of the predicate, can be considered a subordination-by-nominalization strategy. However, with respect to the categories encoded, the result of this operation is more "verb-like" than an underived verb, an effect that can largely be ascribed to the peculiar properties of Movima referential phrases.

8. Conclusions and further research

8.1 Summary of the above

My scientific career has centered on Movima, an endangered isolate of South America whose morphosyntactic properties are typologically challenging. One may say that Movima is one of those languages that "have it all", as Stephen Levinson (p.c.) once put it: verb-initial constituent order; split ergativity caused by hierarchical alignment; symmetrical voice with inverse features; syntactic flexibility of lexical categories; pronominal predicates presenting a copular structure; nominal tense. Many of these phenomena would have been impossible to analyze without the typological knowledge that is available today.

Based on a 30-hour annotated corpus of spontaneous oral speech that was mostly compiled during postdoctoral documentation projects (see Chapters 2 and 3), I first worked on the organization of basic clauses (Chapter 4) and on the conditions for how the arguments of a transitive predicate are expressed (Chapter 5). When analyzing basic clauses, I took into account the fact that the syntactic functions of verbs and nouns are interchangeable, and it seems that this syntactic flexibility is a key to understanding the unusual split-ergative alignment system of Movima.

I then started to investigate pragmatically marked constructions, especially the "pronominal construction", in which the external argument is expressed by a fronted pronoun (Chapter 6). This construction displays a cleft structure, but this analysis is not corroborated by the function of the construction, which is topicalizing rather than focalizing. The solution, which still requires further research, seems to be to analyze the pronoun as an emergent copula. In the context of this investigation I was also finally able to pin down the pragmatic effects of exchanging the lexical categories: a nominal predicate combined with a "verbal RP" is a focus construction.

A further line of research was the investigation of complex sentences and embedded constructions (Chapter 7). Since embedding involves nominalization, referential phrases play a central role here. This is particularly relevant as tense is indicated through determiners, which are an obligatory part of referential phrases.

My work so far has concentrated on pointing out the morphosyntactic characteristics of some central phenomena of Movima grammar and on trying to find typologically and cognitively plausible explanations for them. For instance, I see the alternation between direct and inverse clauses as an effect of discourse-semantic prominence; the apparently surprising syntactic

privileges of the nonprominent argument, by contrast, may be due to the type of syntactic privileges themselves: they involve only access to relativization-type constructions, which enhance the identifiability of the discourse referent and are therefore less relevant for already prominent referents. Also the unusual nominal tense system can be explained cognitively: In Movima, the spatial deictic system may have been extended to temporal deixis, which draws similar lines between the categories accessible – potentially accessible – inaccessible. Developments which have also taken place in the history of other language, such as the use of participles as main verbs, seem to be particularly transparent in the Movima system.

Many of these issues still merit further investigation. While continuing to work on them, I intend to pursue two larger research topics in the future, as much as possible in collaboration with colleagues. These will be elaborated on in the following sections.

8.2 The typology of predicate-initial languages: some initial steps

One larger topic that has interested me for a long time was hinted at at several occasions above. There are some grammatical features that seem to go more or less together in some languages. Basically, these are the following:

- predicate/verb in clause initial position
- syntactic flexibility of verbs and nouns
- presence of determiners
- absence of a copula
- ergative features (i.e. {P=S})

Languages and language families for which several or all of these traits have been described include the already mentioned Western Austronesian and Mayan languages, but also Classical Nahuatl (Launey 1994); Salishan languages (Davis & Matthewson 2009); and some isolates or smaller languages such as Katukina-Kanamari of Venezuela (Queixalós 2010). Movima is an additional example, not treated in previous studies of these phenomena.

Given that these languages or language families are spread over the globe and cannot be expected to have had any relationship with each other, the clustering of features can be a coincidence, but more probably they have effected each other. The idea is that if a system has one feature, this causes others to develop the others as well (see the correlates of verb-initial languages mentioned e.g. by Greenberg 1963; Myhill 1985; Clemens & Polinsky To appear).

Investigating these features involves a holistic approach to typology that identifies a particular type of language with respect to a central aspect of their grammar.

There seem to be different, sometimes implicit, hypotheses as to what causes the correlation between two or more of the above-mentioned phenomena. Some studies seem to imply that the correlations are based on the order of constituents (see Greenberg 1963; Myhill 1985; Clemens & Polinsky To appear). Others regard the use of participles as main predicates, at least at some historical stage of the language, as being the source of nominal predicate expressions and, hence, of equational-clause like structures (see e.g. Sasse 2009 on Eastern Aramaic). Perhaps for similar reasons, Kaufman's seminal paper in *Theoretical Linguistics* is titled "Austronesian nominalism and its consequences" (Kaufman 2009). Queixalós (p.c.) sees ergativity, i.e. the orientation towards the patient, as an underlying phenomenon forming the character of a linguistic system. So, what is the reason for this clustering of features, and which one may have come first? I do not yet have a clear hypothesis regarding this question, and probably it cannot be answered. Even if there is one feature that causes the appearance of the others – e.g. predicate-initial word order –, in the history of a language this feature can get lost through language contact or other developments.

Since Movima is an isolate, my "discovery" that this language is yet another member of this – crosslinguistically rare – type calls for further research on this clustering of features. An obstacle is that there are, of course, no historical data of Movima that might provide evidence for how the system has developed in this language. Furthermore, the language at its present stage of attrition does not permit in-depth experimental research (as done, for instance, by Norcliffe et al. 2015 on the Mayan language Tzeltal), let alone the investigation of the speech of younger generations. Typological research is all there is. Some questions I would like to approach in collaboration with colleagues working on languages with similar systems include the following:

- Do these languages show the same tendency to express relativization through nominalization?
- Given the absence of a copula, how are identificational clauses formed, especially those with pronominal elements? Which element is the predicate in that case?
- Which non verb-initial languages show similar traits?

The best kick-off to investigate the question would be to organize a workshop with experts on the above-mentioned languages or language families. I have been brainstorming on this idea

for some time, but I still have not found the catching idea that would be the point of departure: syntactic flexibility of lexical categories? verb initiality? equation without a copula? etc.

I hope that these ideas will take shape during my planned collaboration in the project "Prominence-related structures in symmetrical-voice and Papuan languages" directed by Nikolaus Himmelmann and Sonja Riesberg within the second phase of the Collaborative Research Center *Prominence in Language* in Cologne (see Chapter 5) starting in 2021. The project is conducted by experts on Austronesian symmetrical-voice languages (Tagalog, Totoli), and we will discuss our annotations and different construction types. Based on the primary discourse data we all work with, we expect to gain a better understanding of the essence of these very similar grammatical systems.

8.3 Morphosyntax in discourse

8.3.1 Corpus annotation of "construction types"

While the topic outlined in 8 is close to my heart, but still vague in terms of research activities and results, my plans are more concrete regarding collaborative work on the comparison of the discourse functions of language-particular constructions. In contrast to research on predicate-initial languages, which can only be carried out by experts on languages with very specific traits, the possibilities for language comparison on this topic are almost unlimited, and I expect it to be relatively easy to arrive at interesting results.

As outlined in the preceding sections, in the past I have concentrated on sentence-level syntax and analyzed the possible structures that clauses and sentences can have. I worked most of all on the so-called "unmarked" constructions, i.e. the verb-initial sentences that probably make up almost 90 per cent of all Movima clause types.

With the article on the pronominal construction in the volume on information structure (Haude 2018a; see 6.2.1), I first used contextual data systematically to identify the *function* of a construction, which I would not have been able to identify otherwise. The same happened with the paper I wrote for the volume on "Clefts and other focus constructions" (see 6.2.4), where the analysis of the larger context allowed me to identify a "real cleft" and to distinguish it from the simple predicate-nominal focus construction. In my future research I plan to continue investigating the uses and functions of "marked" vs. "unmarked" constructions, which involve the placement of a constituent in preverbal position, but also the predicative use of a nonverbal element.

In order to do so, I plan to annotate all types of constructions that appear in the corpus. These annotations will make it possible, among other things, to

- make a quantitative analysis of the different construction types and their "marked" status
- assess the distribution of transitive and intransitive constructions
- analyze the function of a particular construction type by taking into account the preceding and subsequent context
- investigate the use of different types of argument expressions (RP, bound pronoun, free pronoun, zero)

I will, first of all, continue with the annotation of intransitive clauses according to the scheme described in 3.3 (Figure 9). I will then turn to the annotation of transitive clauses with first- and second-person arguments, which were not part of the quantitative study described in 5.1. This amount of corpus annotation has not yet been carried out very often on little-documented languages. Quantitative studies of argument expressions, construction types, or the distribution of transitive vs. intransitive clauses are usually based on a limited number of texts (see the corpora in the MultiCAST project, https://multicast.aspra.uni-bamberg.de/). To achieve more robust results, however, I consider it necessary to make the entire corpus available for this type of study. Once it is fully annotated, it can serve as a basis for an unlimited number of studies, some of which can potentially be carried out as part of an MA or even a PhD thesis.

8.3.2 The LABEX-EFL project "Morphosyntax in discourse"

The corpus annotation will be part of a collaborative research project *Morphosyntax in discourse: Comparing the representation of events and their participants in discourse corpora of typologically diverse languages* that I coordinate together with Tatiana Nikitina (LLACAN) within the LABEX-EFL. This project brings together linguists who have compiled annotated corpora on genealogically and typologically different languages (see Table 18). We envisage international collaboration with, for example, Eva Schultze-Berndt (Manchester), Sonja Riesberg (Cologne), and Stefan Schnell (Bamberg), who all have similar research questions and are experienced in corpus analyses.

All members of this research group are experts on typologically and genetically diverse languages, of which they have assembled corpora of spontaneous oral discourse and which

represent different types of constructions that allow rearrangement of event and participant expressions in discourse. Some of the participating languages and their constructions most relevant for the investigation are listed in Table 18 (in alphabetical order).

Table 18. Languages and phenomena to be investigated

language or smaller family	**(larger) family**	**special phenomenon to be studied**	**researcher**
Amis	Austronesian	symmetrical voice	Bril
Kabyle	Berber	"absolute" vs. "annexed" state marking on nouns	Mettouchi
Kiranti	Tibeto-Birman	direct-inverse	Lahaussois
Movima	Isolate (Amerindian)	direct-inverse + valency decrease	Haude
Pesh	Chibchan	"topic" morpheme	Chamoreau
Romanian	Romance	differential object marking	Mardale
Wan	Mande	overt marker for focus + A argument	Nikitina

Like myself, the participating colleagues have worked for many years on their respective languages of study and have acquired an excellent knowledge of the morphological, syntactic, and also semantic properties of different syntactic constructions. Our aim now is to get a better idea of the *conditions* under which a particular construction or argument expression is used in the individual language. This topic has been of interest at least since the 70s (Chafe 1980; Givón 1983b), but it has usually been investigated on the basis of more limited corpora. Based on our larger corpora of spontaneous speech we will investigate the grammatical constructions that speakers actually use when describing events – in particular, events involving more than one participant – in a larger context. In particular, we will analyze features such as voice alternations (8.3.2.1), the ways in which event participants are encoded (8.3.2.2), and information structure (8.3.2.3).

8.3.2.1 Voice alternations

In many languages, predicates can undergo voice alternations such as active, passive, middle etc. (see Kulikov 2011), while other languages possess symmetrical alternations such as actor/undergoer voice (see Haude & Zúñiga 2016). Both systems can grammatically promote or demote arguments according to their discourse prominence; so, for instance, a passive can be used to maintain the discourse topic as a syntactic subject when it represents the undergoer in a two-participant event description.

Functional similarities and differences between a direct/inverse and an active/passive alternation can be illustrated easily. In languages with a direct/inverse system, like Movima, the direct voice is chosen when the internal argument – which represents the more topical (and/or animate, etc.) event participant – is the actor, and the external argument – less topical, inanimate, etc. – the undergoer (145); the inverse is chosen in the opposite case (146). As the translations of these examples show, this alternation is comparable to the active-passive alternation in languages like English.

(145) Movima direct/English active:
*sal-na**=us**--k-is*
look_for-DR=3M.AB--OBV-3PL.AB
'He looked for them.'

(146) Movima inverse/English passive:
*sal-kay-a**=us**--k-is*
look_for-INV=3M.AB--OBV-3PL.AB
'He was being looked for by them.'

In both the Movima inverse and the English passive (146), the grammatical relation of the third-person masculine (=*us*) is retained with respect to the related direct viz. active construction (145). In this way, both devices can serve a topic-maintaining or reference-tracking function. At the same time, there are obvious differences, such as the difference in transitivity (inverse=transitive, passive=intransitive) and the corresponding difference in the syntactic status of the actor (direct vs. oblique argument, respectively). Comparing the overt expression of actors in inverse vs. in passive constructions in spontaneous discourse may yield interesting results regarding the way in which typologically different languages present events and their participants.

8.3.2.2 Expression of arguments and reference tracking

Partly independently of the sentence in which they appear, discourse participants can be encoded in different ways: participants that are "accessible" (i.e. identifiable from the context) are more naturally encoded by pronouns (e.g. Ariel 1990; Gundel, Hedberg & Zacharski 1993; Kibrik 2011) than newly introduced and less easily accessible discourse participants, which tend to be encoded by full (and maybe also modified or relativized) noun phrases. Many

languages – with or without obligatory verbal person indices – allow arguments to remain unexpressed, which can either signal high accessibility of their referents or, on the contrary, a lack of relevance to the overall discourse. Many languages use overt marking on nominal constituents, e.g. definiteness, to indicate more fine-grained distinctions between different discourse referents. Likewise, differential argument marking is known to be at least partially related to discourse prominence (e.g. von Heusinger & Kaiser 2007). For example, in Spanish, differential argument marking – through the dative preposition *a* – is largely restricted to animate referents, with which it is obligatory in many contexts; however, with indefinite animate noun phrases, differential marking indicates a specific referent, as in the well-known example *Busco **(a)** una secretaria qui sabe inglés* 'I look for a **(specific)** secretary who speaks English.')

One problem for studying these phenomena is that it is difficult to identify the conditions under which a particular construction is used without risking circularity – e.g., is a pronoun used because the referent is considered "accessible" in the addressee's mind, or is it the use of the pronoun that suggests its accessibility? This problem was discussed, among others, by Chafe (1976; 1994). Givón (1983b) was an important proposal for identifying accessible referents independently by counting their frequency of mention. Our ELAN corpora will provide us with the possibility to follow this path. A new function "Groups and Links" in the programme ELAN-CorpA developed by Christian Chanard (http://llacan.vjf.cnrs.fr/res_ELAN-CorpA.php) will allow us to mark and list referential expressions.

In studies of the representation of event participants by full noun phrases, Bickel (2003) and Stoll & Bickel (2009) suggest a morphosyntactically based typological correlation between elaborate event descriptions and relatively poor information on participants, on the one hand, and simpler event descriptions and highly elaborate information on event participants, on the other hand. This is particularly relevant when also grammatical phenomena not directly related to the participant constellation of events, but to the event's internal structure (e.g. verbal aspect marking), are taken into account (see Payne & Shirtz 2015): a language whose aspect-marking system also serves an information structuring purpose may belong to those languages that spend less efforts on elaborating on the expression of event participants. With the corpora represented in the project it will be possible to compare discourse data of similar types (e.g. length, genre, amount and person features of discourse protagonists) in different languages to see if this hypothesis is supported by our data.

The starting point of the research in this project will be the analysis of how the event representation patterns identified previously for each language are distributed in actual discourse. This will necessarily entail an analysis of discourse-structuring strategies in the respective languages, e.g. of the ways in which new protagonists are introduced in a narrative and the prosodic, morphological and/or syntactic markers that indicate a new passage in a text. The structuring of discourse influences the perspective from which an event is presented, but it is usually not the first thing morphosyntacticians or descriptive linguists look at. Therefore, this is a new topic for most of us, and colleagues with an expertise in the organization of discourse (e.g. of the Cologne *CRC Prominence in Language*, see 5.2), will occasionally be invited.

Once the criteria that may influence the choice of a construction in an individual language are identified, the functions of the morphosyntactic strategies used by one language can be compared with those of others (e.g. a passive with an inverse; omission of argument expressions in different languages; elaborate event vs. participant descriptions), while bearing in mind the overall morphosyntactic system and the grammatical possibilities of each language. We will select chunks of discourse that allow cross-linguistic comparison: they should be of the same genre (e.g. narrative), be of approximately equal length, contain a roughly equal type of discourse protagonists (e.g. third persons) and a comparable number of referents, since a larger number of referents demands more explicit disambiguation and, hence, a more fine-grained lexical specification of discourse referents.

In addition, the time distance between coreferring expressions, measurable with the time-aligned annotations that are available now (e.g. in ELAN), can be revealing, since a reason for referring to an already introduced entity with a full noun phrase may simply be the time that has elapsed since its last mention.

The analysis of discourse data in order to find out how argument expressions and syntactic structures are employed to guide the hearer through a text will build directly on the research I have done so far on prominence/salience (see Chapter 5) and information structure (Chapter 6). I have also already explored some of the most relevant literature on this topic, especially when preparing the course “Reference in Discourse” that I gave together with Stavros Skopeteas at the University of Göttingen in 2018, which was largely based on the works by Chafe (1976; 1994), DuBois (1987) and Kibrik (2011).

8.3.2.3 Information structure

Finally, the project team will investigate information structure, in continuation of the project GD1 *The Typology and Corpus Annotation of Information Structure and Grammatical Relations* of the first phase of the LABEX EFL (with which it shares most of its members; see 2.2.3). Alternations in information structure involve different ways to express the same propositional content, be it by prosodic, morphological, or syntactic devices. Constructions may also differ simply in the ordering of their constituents. In many languages, especially those with nominal case marking or verbal argument affixes, a useful device to reflect pragmatic status is syntactic ordering (Downing 1995), as in German SVO (*Sie suchten ihn* 'they looked for him') vs. OVS (*Ihn suchten sie* 'They looked for HIM [as opposed to somebody else]'). Importantly, all these strategies – voice, argument expression and marking, constituent order (and more, such as aspect marking, lexical choices, and prosody) – interact and thereby provide a speaker with a complex toolkit to present the events in the most appropriate way. Studying these constructions in their contextual embedding will help us to understand their choice.

8.3.3 Outlook

This project will give me the opportunity to focus more deeply on questions I already approached in previous research on my data: Which factors play a role for the zero encoding of the external argument (see 5.1), for pronominal encoding vs. encoding as an RP, of for the choice of the valency-decreasing construction rather than the inverse? In addition, I will investigate issues that I had not concentrated on before, like discourse units ("paragraphs") and the constructions with which new referents are introduced (e.g. existential constructions). I hope to develop a better-founded idea of what a "protagonist" is (see 5.2) and how it is treated linguistically. At the same time, through the collaboration with my colleagues I hope to be able to compare the grammatical strategies of Movima with those found in other languages. This could result, for instance, in the empirical discovery that the inverse has indeed many functional similarities (or not) with the passive or with differential object marking in another language. If our collaboration within this project proves fruitful, I envisage to submit a ANR-DFG funded project together with Stefan Schnell and Geoffrey Haig (Bamberg University).

This investigation overlaps partly with the project *De l'intrapropositionnel à l'interpropositionnel: syntaxe et construction du texte* (dir. Claudine Chamoreau), which is part of Axe 3 of the SeDyL starting in 2020. This project investigates, on the one hand, subordination strategies in different languages and the way in which expressions of states and

events are hierarchically ordere; on the other hand, it focuses on reference tracking and the effect that referential expressions have on information structure.

My interest in investigating discourse data of Movima, furthermore, matches perfectly with the aims of the above-mentioned project "Prominence-related structures in symmetrical-voice and Papuan languages" directed by Sonja Riesberg and Nikolaus Himmelmann. This project is primarily concerned with prominence relations in Austronesian symmetrical-voice languages, but it will also involve the comparison with Movima, described in the proposal as "one of the few well-established symmetrical voice languages outside western Austronesia". The project departs from the assumption that besides the better-studied sentence-level factors, also more global discourse structures may have a major influence on the choice of one voice over another, and therefore, the investigations will be extended from the local level to the global level of discourse in symmetrical-voice languages. In the project proposal I am listed as an associated member, and I hope that the interaction between the team of this project and the LABEX project will enhance fruitful synergetic effects. Last but not least, the collaboration with this project will connect the more concrete research I described in 8.3 with my interest in predicate-initial languages and the additional features they share.

REFERENCES

Aissen, Judith. 1997. On the syntax of obviation. *Language* 73(4). 705–750.

Aldridge, Edith. 2012. Antipassive and ergativity in Tagalog. *Lingua* 122(3). 192–203.

Alvarez González, Albert. 2012. Relative clauses and nominalizations in Yaqui. In Bernard Comrie & Zarina Estrada-Fernández (eds.), *Relative Clauses in Languages of the Americas: A typological overview*, 67–95. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London/New York: Routledge.

Arnold, Jennifer E. 2008. Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes* 23(4). 495–527.

Authier, Gilles & Katharina Haude (eds.). 2012. *Ergativity, valency and voice* (Empirical Approaches to Language Typology 48). Berlin/New York: Mouton de Gruyter.

Bickel, Balthasar. 2003. Referential density in discourse and syntactic typology. *Language* 79(4). 708–736.

Bickel, Balthasar. 2011. Grammatical relations typology. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 399–444. Oxford/New York: Oxford University Press.

Bisang, Walter. 2015. *Hidden complexity: the neglected side of complexity and its consequences*. Linguistics Vanguard.

Bloomfield, Leonard. 1946. Algonquian. In Harry Hoijer (ed.), *Linguistic structures of Native America*, 85–129. New York: Viking Fund Publications in Anthropology 6.

Bohnemeyer, Jürgen. 2004. Split intransitivity, linking, and lexical representation: the case of Yukatek Maya. *Linguistics* 42(1). 67–107.

Bricker, Victoria R. 1979. Wh-questions, relativization, and clefting in Yucatec Maya. In *Papers in Mayan linguistics*, 107–136. Columbia, Missouri: Lucas Brothers.

Cardús, Fray José. 1886. *Las Misiones Franciscanas entre los infieles de Bolivia. Descripción del estado de ellas en 1883 y 1884, con una noticia sobre los caminos y tribus salvajes, una muestra de varias lenguas, curiosidades de historia natural y un mapa para servir de ilustración*. Barcelona: Librería de la Inmaculada Concepción.

Castillo, Marius del. 1929. *El corazón de la América meridional (Bolivia)*. Barcelona: Imprenta Comercial.

Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles Li (ed.), *Subject and topic*, 25–55. New York: Academic Press.

Chafe, Wallace (ed.). 1980. *The pear stories: cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

Chafe, Wallace. 1994. *Discourse, consciousness, and time: the flow and displacement of conscious experience in speaking and writing*. Chicago/London: The University of Chicago Press.

Chamberlain, Alexander F. 1910. Sur quelques familles linguistiques peu connues ou presque inconnues de l'Amérique du Sud. *Journal de la Société des Américanistes* 7. 179–202.

Clemens, Lauren Eby & Maria Polinsky. To appear. Verb-initial word orders, primarily in Austronesian and Mayan languages. In Martin Everaert & Henk C. van Riemsdijk (eds.), *The Wiley Blackwell Companion to Syntax, Second Edition*. Hoboken, NJ: John Wiley & Sons, Inc.

Comrie, Bernard. 1978. Ergativity. In Winfried P. Lehmann (ed.), *Syntactic Typology: Studies in the Phenomenology of Language*, 329–394. Austin, TX: University of Texas Press.

Comrie, Bernard. 1985. *Tense*. Cambridge: Cambridge University Press.
Comrie, Bernard. 1989. *Language universals and linguistic typology: syntax and morphology*. 2nd edn. Chicago: The University of Chicago Press.
Comrie, Bernard & Sandra A. Thompson. 2007. Lexical nominalization. In Timothy Shopen (ed.), *Language Typology and Syntactic Description, vol. 3: Grammatical Categories and the Lexicon*, 334–381. 2nd edn. Cambridge: Cambridge University Press.
Créqui-Montfort, Georges de & Paul Rivet. 1914. La langue mobima. *Journal de la Société des Américanistes de Paris* 11. 183–211.
Cristofaro, Sonia. 2003. *Subordination*. Oxford: Oxford University Press.
Croft, William. 2003. *Typology and universals*. 2nd edn. Cambridge: Cambridge University Press.
Crystal, David. 2008. *A dictionary of linguistics and phonetics* (Arbeitspapier). 6th edn. Malden, MA/Oxford: Blackwell Pub.
Davis, Henry & Lisa Matthewson. 2009. Issues in Salish syntax and semantics. *Language and Linguistics Compass* 3(4). 1097–1166.
Declerck, Renaat. 1988. *Studies on copular sentences, clefts, and pseudo-clefts*. Leuven: Leuven University Press, Foris Publications Holland/USA.
Dixon, R. M. W. 1994. *Ergativity* (Cambridge Studies in Linguistics 69). Cambridge: Cambridge University Press.
Dixon, R.M.W. 1972. *The Dyirbal language of North Queensland* (Cambridge Studies in Linguistics 9). Cambridge: Cambridge University Press.
Dixon, R.M.W. & Alexandra Y. Aikhenvald. 1997. A typology of argument-determined constructions. In Joan Bybee, John Haiman & Sandra A. Thompson (eds.), *Essays of Language Function and Language Type. Dedicated to T. Givón*, 71–113. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Downing, Pamela. 1995. Word order in discourse: by way of introduction. In Pamela Downing & Michael Noonan (eds.), *Word order in discourse* (Typological Studies in Language 30), 1–28. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.
Dryer, Matthew S. 1997. Are grammatical relations universal? In Joan Bybee, John Haiman & Sandra A. Thompson (eds.), *Essays of Language Function and Language Type. Dedicated to T. Givón*, 115–143. Amsterdam/Philadelphia: John Benjamins Publishing Company.
DuBois, John W. 1987. The discourse basis of ergativity. *Language* 63(4). 805–855.
Eberhard, David M., Gary F. Simons & Charles D. Fennig. 2019. *Ethnologue: Languages of the World*. Dallas, TX: SIL International.
Falk, Simone. 2014. On the notion of salience in spoken discourse - prominence cues shaping discourse structure and comprehension. *Travaux interdisciplinaires sur la parole et le langage* (30).
Fox, Barbara A. 1987. The noun phrase accessibility hierarchy reinterpreted: subject primacy or the absolutive hypothesis? *Language* 63(4). 856–870.
Ganenkov, Dmitry. 2016. Relativization in a morphologically ergative language: a corpus study. Presented at the 36th Meeting of the Deutsche Gesellschaft für Sprachwissenschaft, Konstanz. https://www2.hu-berlin.de/syntax/themp/uploads/PDF/DGfS2016_ganenkov.pdf (18 May, 2020).
Gildea, Spike. 2012. The referential hierarchy and attention. *Faits de Langues* 39(1). 33–47.
Givón, Talmy. 1983a. Topic continuity in discourse: an introduction. In Talmy Givón (ed.), *Topic continuity in discourse: a quantitative cross-linguistic study*, 1–41. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Givón, Talmy (ed.). 1983b. *Topic continuity in discourse: a quantitative cross-linguistic study* (Typological Studies in Language 3). Amsterdam/Philadelphia: John Benjamins Publishing Company.
Gordon, Peter C. & Randall Hendrick. 2005. Relativization, ergativity, and corpus frequency. *Linguistic Inquiry* 36(3). 456–463.
Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of Language*, 73–113. Cambridge, MA: MIT Press.
Grimshaw, Jane. 1990. *Argument structure*. Cambridge, MA: MIT Press.
Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.
Haig, Geoffrey & Stefan Schnell. 2016. The discourse basis of ergativity revisited. *Language* 92(3). 591–618.
Haude, Katharina. 2004. Nominal tense marking in Movima: nominal or clausal scope? In Leonie Cornips & Jenny Doetjes (eds.), *Linguistics in The Netherlands 2004* (AVT Publications 21), 80–90. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://www.benjamins.com/#catalog/journals/avt.21.11hau/fulltext.
Haude, Katharina. 2006. *A Grammar of Movima*. Nijmegen: Radboud University Doctoral dissertation. https://repository.ubn.ru.nl/bitstream/handle/2066/41395/41395.pdf.
Haude, Katharina. 2008. Clause structure and alignment in Movima. *Amerindia* 32. 89–107.
Haude, Katharina. 2009a. Hierarchical alignment in Movima. *International Journal of American Linguistics* 75(4). 513–532.
Haude, Katharina. 2009b. Reference and predication in Movima. In Patience Epps & Alexandre Arkhipov (eds.), *New Challenges in Typology: Transcending the Borders and Refining the Distinctions*, 323–342. Berlin: Mouton de Gruyter. http://hal.archives-ouvertes.fr/docs/00/67/07/14/PDF/NCT2-13-haude-20090120-HAL.pdf.
Haude, Katharina. 2010a. The intransitive basis of Movima clause structure. In Spike Gildea & Francesc Queixalós (eds.), *Ergativity in Amazonia* (Typological Studies in Language 89), 285–316. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.89.11hau. http://halshs.archives-ouvertes.fr/docs/00/67/05/53/PDF/haude_ErgAmaz-singlespace.pdf.
Haude, Katharina. 2010b. "She kisses her late husband" = "She kissed her husband": nominal tense in Movima. In Jan Wohlgemuth & Michael Cysouw (eds.), *Rara & Rarissima*. Berlin/New York: De Gruyter Mouton. http://halshs.archives-ouvertes.fr/docs/00/67/09/25/PDF/Rara_Haude-revision-corrected.pdf.
Haude, Katharina. 2011a. *El idioma movima*. La Paz: Ministerio de Educación/UNICEF. http://www.proeibandes.org/wp-content/uploads/2018/12/lengua-movima.pdf.
Haude, Katharina. 2011b. Argument encoding in Movima: the local domain. *International Journal of American Linguistics* 77(4). 559–575. https://doi.org/10.1086/662156.
Haude, Katharina. 2011c. Tense marking on dependent nominals in Movima. In Renate Musan & Monika Rathert (eds.), *Tense across Languages*. Berlin/Boston: De Gruyter. http://halshs.archives-ouvertes.fr/docs/00/67/22/48/PDF/haude-TenseAcrossLanguages-revision_1.pdf.
Haude, Katharina. 2011d. Referring to states and events: subordination in Movima. In Rik van Gijn, Katharina Haude & Pieter Muysken (eds.), *Subordination in South American Languages* (Typological Studies in Language 97), 141–168. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.97.06hau. http://halshs.archives-ouvertes.fr/docs/00/67/12/04/PDF/Movima-subordination-HAL.pdf.

Haude, Katharina. 2012a. Undergoer orientation in Movima. In Gilles Authier & Katharina Haude (eds.), *Ergativity, Valency and Voice* (Empirical Approaches to Language Typology 48). Berlin/New York: De Gruyter. http://halshs.archives-ouvertes.fr/docs/00/69/20/77/PDF/08_Haude.pdf.

Haude, Katharina. 2012b. Saillance inhérente et saillance discursive en movima. *Faits de Langues* 39(1). 169–180. https://doi.org/10.1163/19589514-039-01-900000013.

Haude, Katharina. 2014a. Animacy and inverse in Movima: a corpus study. *Anthropological Linguistics* 56(3–4). 294–314.

Haude, Katharina. 2014b. Reduplication in Movima: a prosodic morphology approach. In Gale Goodwin Gómez & Hein van der Voort (eds.), *Reduplication in Indigenous Languages of South America* (Brill's Studies in the Indigenous Languages of the Americas 7). Leiden/Boston: Brill. http://halshs.archives-ouvertes.fr/docs/00/96/11/34/PDF/13_RED_in_Movima-final.pdf.

Haude, Katharina. 2018a. Nonverbal predication in Movima. In Simon E. Overall, Rosa Vallejos & Spike Gildea (eds.), *Nonverbal predication in Amazonian languages* (Typological Studies in Language 122), 217–244. Amsterdam: John Benjamins Publishing Company. https://halshs.archives-ouvertes.fr/halshs-01848221v1.

Haude, Katharina. 2018b. A topic-marking cleft? Analyzing clause-initial pronouns in Movima. In Evangelia Adamou, Katharina Haude & Martine Vanhove (eds.), *Information structure in lesser-described languages: Studies in prosody and syntax* (Studies in Language Companion Series 199), 217–244. Amsterdam: John Benjamins Publishing Company. https://halshs.archives-ouvertes.fr/hal-01875573v1.

Haude, Katharina. 2019a. On discourse-semantic prominence, syntactic prominence, and prominence of expression: the case of Movima. *Journal of Pragmatics* 154. 22–38.

Haude, Katharina. 2019b. From predication to reference: on "verbal DPs" in Movima. In Christine Bonnot, Outi Duvallon & Hélène de Penanros (eds.), *Individuation et référence nominale à travers les langues*, 53–77. Paris: Editions Lambert-Lucas. https://halshs.archives-ouvertes.fr/hal-01971969v1.

Haude, Katharina. 2019c. Clefting and nominal predication: two focus-marking constructions in Movima. *Linguistic Discovery* 17(1). 106–127.

Haude, Katharina. 2019d. Grammatical relations in Movima: Alignment beyond semantic roles. In Alena Witzlack-Makarevich & Balthasar Bickel (eds.), *Typological Studies in Language*, vol. 123, 213–256. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.123.07hau. https://benjamins.com/catalog/tsl.123.07hau (3 February, 2020).

Haude, Katharina & Annie Montaut (eds.). 2012. *La Saillance* (Faits de Langues 39(1)). https://brill.com/view/journals/fdl/39/1/fdl.39.issue-1.xml.

Haude, Katharina & Alena Witzlack-Makarevich (eds.). 2016a. *Referential hierarchies and alignment* (Linguistics 54(3)).

Haude, Katharina & Alena Witzlack-Makarevich. 2016b. Referential hierarchies and alignment: an overview. *Linguistics* 54(3). 433–441.

Haude, Katharina & Fernando Zúñiga. 2016. Inverse and symmetrical voice: on languages with two transitive constructions. *Linguistics* 54(3). 443–481.

Heath, Edwin R. 1883. Dialects of Bolivian Indians. A philological contribution from material gathered during three years residence in the Department of Beni in Bolivia. *Kansas City Review Science and Industry* 6. 679–687.

Hengeveld, Kees. 1992. *Non-verbal predication: theory, typology, diachrony.* Berlin/Boston: De Gruyter Mouton.

Hervás y Panduro, Lorenzo. 1787. *Vocabulario poligloto con prolegomeni sopra piu' di CL. lingue [Vocabulario políglota con prolegómenos sobre más de 150 lenguas]*. Cesena: Gregorio Biasini.

Heusinger, Klaus von & Georg Kaiser. 2007. Differential object marking and the lexical semantics of verbs in Spanish. In Georg A. Kaiser & M. Leonetti (eds.), *Proceedings of the Workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages"* (Arbeitspapier 122), 85–110. Konstanz: Fachbereich Sprachwissenschaft der Universität Konstanz.

Himmelmann, Nikolaus P. 1991. The Philippine challenge to universal grammar. Köln: Institut für Sprachwissenschaft.

Himmelmann, Nikolaus P. 2005a. Tagalog. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 350–376. London/New York: Routledge.

Himmelmann, Nikolaus P. 2005b. The Austronesian languages of Asia and Madagascar: typological characteristics. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 110–181. London/New York: Routledge.

Himmelmann, Nikolaus P. 2008. Lexical categories and voice in Tagalog. In Peter K. Austin & Simon Musgrave (eds.), *Voice and Grammatical Functions in Austronesian Languages*, 247–293. Stanford, CA: CSLI Publications.

Himmelmann, 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.

Hockett, Charles F. 1966. What Algonquian is really like. *International Journal of American Linguistics* 32(1). 59–73.

Holmstedt, R. D. & A. R. Jones. 2014. The pronoun in tripartite verbless clauses in Biblical Hebrew: resumption for left-dislocation or pronominal copula? *Journal of Semitic Studies* 59(1). 53–89.

Hopper, Paul J. & Sandra A. Thompson. 1984. The discourse basis for lexical categories in universal grammar. *Language* 60(4). 703–752.

Jasinskaja, Katja, Sofiana Chiriacescu & Stefan Hinterwimmer. 2015. Prominence in discourse. In Amedeo De Dominicis (ed.), *pS-prominenceS: Prominences in Linguistics. Proceedings of the International Conference*, 134–153. Viterbo: Disucom Press.

Judy, Judith E. 1965. Independent verbs in Movima. *Anthropological Linguistics* 7(7). 10–15.

Judy, Judith E. & Robert A. Judy. 1962a. *Movima y castellano* (Vocabularios Bolivianos 1). Cochabamba: Instituto Lingüístico de Verano.

Judy, Judith E. & Robert A. Judy. 1962b. Fonemas del Movima, con atención especial a la serie glotal. In *Notas Lingüísticas de Bolivia, 5*, 1–35. Cochabamba: Instituto Lingüístico de Verano.

Judy, Judith E. & Robert A. Judy. 1967. Movima. In *Bolivian Indian Grammars* (Summer Institute of Linguistics Publications in Linguistics and Related Fields 16), 353–408. Oklahoma: SIL publications.

Judy, Robert A. 1965. Pronoun introducers in Movima clauses. *Anthropological Linguistics* 7(7). 5–9.

Katz, Aya. 1996. *Cyclical Grammaticalization and the cognitive link between pronoun and copula*. Austin, Texas: Rice University PhD dissertation.

Kaufman, Daniel. 2009. Austronesian Nominalism and its consequences: A Tagalog case study. *Theoretical Linguistics* 35(1). 1–49.

Kaufman, Daniel. 2018. Austronesian predication and the emergence of biclausal clefts In Indonesian languages. In Sonja Riesberg, Asako Shiohara & Atsuko Utsumi (eds.), *Perspectives on information structure in Austronesian languages*, 207–245. Berlin: Language Science Press.

Keenan, Edward L. 1976. Towards a universal definition of "subject." In Charles Li (ed.), *Subject and Topic*, 247–301. New York: Academic Press.

Kibrik, Andrej A. 2011. *Reference in discourse*. Oxford: Oxford University Press.

Klaiman, M. H. 1991. *Grammatical voice* (Cambridge Studies in Linguistics 59). Cambridge ; New York: Cambridge University Press.

Korn, Agnes. 2011. Pronouns as verbs, verbs as pronouns: demonstratives and the copula in Iranian. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in Iranian linguistics* (Beiträge zur Iranistik 34), 53–70. Wiesbaden: Reichert.

Kroeger, Paul. 1993. *Phrase Structure and Grammatical Relations in Tagalog*. Stanford, CA: CSLI Publications.

Kulikov, Leonid. 2011. Voice typology. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*, 368–398. Oxford: Oxford University Press.

Kuno, Susumu. 1987. *Functional syntax: anaphora, discourse and empathy*. Chicago/London: The University of Chicago Press.

Lambrecht, Knud. 1994. *Information structure and sentence form: topic, focus, and the mental representations of discourse referents*. Cambridge/New York: Cambridge University Press.

Lambrecht, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39(3). 463–516.

Launey, Michel. 1994. *Une grammaire omniprédicative*. Paris: CNRS Editions.

Launey, Michel. 2004. The features of omnipredicativity in Classical Nahuatl. *STUF - Language Typology and Universals* 57(1). 49–69.

Li, Charles N. & Sandra A. Thompson. 1977. A mechanism for the development of copula morphemes. In Charles N. Li (ed.), *Mechanisms of syntactic change*, 419–444. Austin/London: University of Texas Press.

Lowder, Matthew W. & Peter C. Gordon. 2015. Natural forces as agents: reconceptualizing the animate–inanimate distinction. *Cognition* 136. 85–90.

Margetts, Anna & Peter K. Austin. 2007. Three-participant events in the languages of the world: towards a crosslinguislic typology. *Linguistics* 45(3). 393–451.

Margetts, Anna, Katharina Haude, Nikolaus P. Himmelmann, Dagmar Jung, Sonja Riesberg, Stefan Schnell, Frank Seifart, Harriet Sheppard & Claudia Wegener. To appear. Cross-linguistic patterns in the lexicalisation of BRING and TAKE. *Studies in Language*.

Matthews, Peter. 1997. *The Concise Oxford Dictionary of Linguistics*. Oxford/New York: Oxford University Press.

Moravcsik, Julia E. & Alice F. Healy. 1998. Effect of syntactic role and syntactic prominence on letter detection. *Psychonomic Bulletin and Review* 5(1). 96–100.

Myhill, John. 1985. Pragmatic and categorial correlates of VS word order. *Lingua* 66(2–3). 177–200.

Nagaya, Naonori. 2012. On the syntactic transitivity of Tagalog actor-focus constructions. *NINJAL Research Papers* 4. 49–76.

Nice, Kathy Y. van & Rainer Dietrich. 2003. Task sensitivity of animacy effects: evidence from German picture descriptions. *Linguistics* 41(5).

Nichols, Johanna, David A. Peterson & Jonathan Barnes. 2004. Transitivizing and detransitivizing languages. *Linguistic Typology* 8(2). 149–211.

Norcliffe, Elisabeth, Agnieszka E. Konopka, Penelope Brown & Stephen C. Levinson. 2015. Word order affects the time course of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience* 30(9). 1187–1208.
Nordlinger, Rachel & Louisa Sadler. 2004. Nominal tense in crosslinguistic perspective. *Language* 80(4). 776–806.
Orbigny, Alcide Dessalines d'. 1839. *L'homme américain (de l'Amérique Méridionale): considéré sous ses rapports physiologiques et moraux*. Paris: Pitois-Levrault et C.
Overall, Simon E., Rosa Vallejos & Spike Gildea (eds.). 2018a. *Nonverbal predication in Amazonian languages* (Typological Studies in Language 122). Amsterdam/Philadelphia: John Benjamins Publishing Company.
Overall, Simon E., Rosa Vallejos & Spike Gildea. 2018b. Nonverbal predication in Amazonia: typological and diachronic considerations. In Simon E. Overall, Rosa Vallejos & Spike Gildea (eds.), *Nonverbal predication in Amazonia* (Typological Studies in Language 122), 1–49. Amsterdam: John Benjamins Publishing Company.
Payne, Doris L. & Shahar Shirtz (eds.). 2015. Discourse structuring and typology: how strong is the link with aspect? In *Beyond Aspect: The expression of discourse functions in African languages* (Typological Studies in Language 109), 1–22. Amsterdam/Philadelphia: John Benjamins Publishing Company.
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.
Payne, Thomas E. 1997. *Describing morphosyntax*. Cambridge: Cambridge University Press.
Prince, Ellen F. 1981. Toward a taxonomy of given-new formation. In Peter Cole (ed.), *Radical pragmatics*, 223–255. New York: Academic Press.
Queixalós, Francesc. 2010. Grammatical relations in Katukina-Kanamari. In Spike Gildea & Francesc Queixalós (eds.), *Ergativity in Amazonia*, 235–284. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Rialland, Annie & Stéphane Robert. 2001. The intonation system of Wolof. *Linguistics* 39(5). 893–939.
Riesberg, Sonja. 2014. *Symmetrical voice and linking in Western Austronesian languages*. Berlin/Boston: De Gruyter.
Sasse, Hans-Jürgen. 1991. Predication and sentence constitution in universal perspective. In Dietmar Zaefferer (ed.), *Semantic Universals and Universal Semantics*, 75–95. Berlin: Floris.
Sasse, Hans-Jürgen. 2009. Nominalism in Austronesian: a historical typological perspective. Comments on Daniel Kaufman's "Austronesian Nominalism and its consequences." *Theoretical Linguistics* 35(1). 167–181.
Schachter, Paul. 1976. The subject in Philippine languages: topic, actor, actor-topic, or none of the above? In Charles Li (ed.), *Subject and topic*, 491–518. New York: Academic Press.
Schachter, Paul & Fé Otanes. 1972. *Tagalog reference grammar*. Berkeley, CA: University of California Press.
Schultze-Berndt, Eva. 2006. Linguistic annotation. In Jost Gippert, Nikolaus Himmelmann & Ulrike Mosel (eds.), *Essentials of Language Documentation*, 213–251. Berlin: Mouton de Gruyter.
Shank, Scott. 2003. A preliminary semantics for pronominal predicates. In *Working Papers for the 38th International Conference on Salish and Neighbouring Languages*, 215–236. Vancouver, BC: University of British Columbia.
Shibatani, Masayoshi. 2019. What is nominalization? Towards the theoretical foundations of nominalization. In Roberto Zariquiey, Masayoshi Shibatani & David W. Fleck (eds.), *Nominalization in Languages of the Americas* (Typological Studies in Language 124), 15–167. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert M.W. Dixon (ed.), *Grammatical categories in Australian languages*, 112–171. Atlantic Highlands, NJ: Humanities Press.

Stoll, Sabine & Balthasar Bickel. 2009. How deep are differences in referential density? In Jiansheng Guo, Elena Lieven, Susan Ervin-Tripp, Nancy Budwig, Şeyda Özçalikan & Keiko Nakamura (eds.), *Crosslinguistic approaches to the psychology of language: research in the tradition of Dan Isaac Slobin*, 543–555. London: Psychology Press.

Talmy, Leonard. 2007. Attention phenomena. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford Handbook of Cognitive Linguistics*, 264–293. Oxford: Oxford University Press.

Tonhauser, Judith. 2003. F-constructions in Yucatec Maya. In Jan Anderssen, Paula Menéndez-Benito & Adam Werle (eds.), *The Proceedings of SULA 2*, 203–223. Vancouver, BC: GLSA, University of Massachusetts, Amherst.

Traugott, Elizabeth. 1995. Subjectification in grammaticalization. In Dieter Stein & Susan Wright (eds.), *Subjectivity and subjectivisation: linguistic perspectives*, 31–54. Cambridge: Cambridge University Press.

Traugott, Elizabeth. 1999. The rhetoric of counter-expectation in semantic change: a study in subjectification. In Andreas Blank & Peter Koch (eds.), *Historical semantics and cognition*, 177–196. Berlin: Mouton de Gruyter.

Van Valin, Robert D. & Randy J. LaPolla. 1997. *Syntax: structure, meaning, and function* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Vendler, Zeno. 1967. *Linguistics in philosophy*. Ithaca, NY: Cornell University Press.

Verhoeven, Elisabeth & Stavros Skopeteas. 2015. Licensing focus constructions in Yucatec Maya. *International Journal of American Linguistics* 81(1). 1–40.

Williams, Edwin. 1981. Argument structure and morphology. *The Linguistic Review* 1. 81–114.

Zúñiga, Fernando. 2006. *Deixis and alignment: inverse systems in indigenous languages of the Americas* (Typological Studies in Language 70). Amsterdam/Philadelphia: John Benjamins Publishing Company.