

Naturally occurring data in and beyond linguistic typology

18-19 May 2023, Bologna

Book of Abstracts

Last update: 13.03.2023

Index

CROSS-LINGUISTIC PATTERNS OF CAUSAL ADVERBIAL CLAUSES IN GRAMMAR-BASED TYPOLOGY AND CORPUS LINGUISTICS	3
<i>Holger Diessel</i>	
LINKING TYPOLOGICAL AND SOCIOLINGUISTIC DATA TO DRAW INFERENCES ABOUT CONTACT EFFECTS: A WORLDWIDE COMPARATIVE APPROACH	5
<i>Francesca Di Garbo, Eri Kashima, Ricardo Napoleão de Souza, and Kaius Sinnemäki</i>	
TYPOLGY AND VARIATION OF SUBJUNCTIVE MOOD. ON SHARED (PROBABILISTIC) CONSTRAINTS	7
<i>Salvio Digesto</i>	
TYPOLGY AND ITS DATA: FUNCTIONAL MONOCULTURE OR STRUCTURAL DIVERSITY?	10
<i>John W. Du Bois and Giorgia Troiani</i>	
HOW UBIQUITOUS IS COMPLEMENTATION? A NATURALISTIC CROSS-CORPUS STUDY	12
<i>Nicholas Evans, Wayan Arka, Danielle Barth, Henrik Bergqvist, Christian Döhler, Sonja Gipper, Dolgor Güntsetseg, Yukinori Kimoto, Dominique Knuchel, Hitomi Ōno, Eka Pratiwi, Saskia van Putten, Alan Rumsey, Andrea Schalley, Stefan Schnell, Asako Shiohara, Elena Skribnik, Yanti</i>	
WORD ORDER AND INFORMATION STRUCTURE: NEW INSIGHTS FROM CONVERSATIONAL CORPORA	14
<i>Natalia Levshina, Rowena Garcia and Yevheniy Skyra</i>	
<i>ZE MA SHE-'ANI TSARIX 'THAT'S WHAT I NEED': 'REVERSED PSEUDO-CLEFTS' IN HEBREW CONVERSATION</i>	17
<i>Yael Maschler and Hilla Polak-Yitzhaki</i>	
OBJECT OMISSION IN NATURALLY OCCURRING DATA AND ANTIPASSIVES	20
<i>Caterina Mauri, Andrea Sansò</i>	
EMBODIED TYPOLOGY: GESTURE AND PROSODY IN TRADITIONAL STORYTELLING	22
<i>Amina Mettouchi</i>	
TYPOLGY AND USAGE: EXPLANATORY PERSPECTIVES WITH SPECIAL REFERENCE TO NEGATION	24
<i>Matti Miestamo, Ksenia Shagal, Olli Silvennoinen and Chingduang Yurayong</i>	

THE MIGHTY DEMONSTRATIVE <i>Marianne Mithun</i>	26
VARIATION AND TYPOLOGY IN ROMANCE: A CROSS-LINGUISTIC STUDY OF SUBJUNCTIVE SELECTION IN SPONTANEOUS SPEECH <i>Shana Poplack, Rena Torres Cacoullos, Nathalie Dion, Rosane de Andrade Berlinck, Salvio Digesto, Dora Lacasse² and Jonathan Steuck</i>	28
RELATIVELY COMPLEX OR RELATIVELY SIMPLE? TOWARD NEW WAYS OF ANALYZING LANGUAGE VARIATION <i>Benedikt Szmrecsanyi</i>	31
SOCIOLINGUISTICS MEETS TYPOLOGY: THE VALUE OF VERNACULAR SPEECH TO ACCOUNT FOR CORSS-LINGUISTIC PATTERNS <i>Sali Tagliamonte</i>	33
LANGUAGE VARIETY AS THE MINIMAL GRAMMATICAL SYSTEM IN A TYPOLOGICAL PERSPECTIVE <i>Alessandro Vietti and Massimo Cerruti</i>	35

Cross-linguistic patterns of causal adverbial clauses in grammar-based typology and corpus linguistics

Holger Diessel
University of Jena

In this paper, I investigate cross-linguistic patterns of causal adverbial clauses in two different data sources: (i) a typological database of 80 languages and (ii) corpus data of different registers and different genres. The paper shows that while there are some conspicuous differences between the typological data and the corpus data, the two data sources complement each other in important ways. More specifically, I argue that the data from reference grammars and corpora provide converging evidence for the hypothesis that the structural patterns of causal adverbial clauses are shaped by competing functional and cognitive forces (Diessel and Hetterle 2011).

The first part of the paper describes the cross-linguistic patterns of causal clauses based on data from reference grammars and other types of grammatical descriptions. The data suggest that there is a general tendency for causal clauses to follow the main clause. However, in some OV languages, causal clauses are consistently preposed to the main clause (e.g. in Japanese). In particular, we find that causal clauses precede the main clause in languages in which semantically similar constructions (e.g. adpositional phrases and other types of subordinate clauses) are also consistently preposed.

The second part of the paper investigates the structure and distribution of causal clauses in spoken and written corpus data from five languages: English, German, French, Japanese and Chinese. The corpus data are largely consistent with the description of adverbial clauses in reference grammars, but there are some important differences between spoken and written registers. In particular, there is a much stronger tendency for causal clauses to follow the main clause in spoken registers than in writing. Moreover, the corpus data show that pre- and postposed causal clauses differ in terms of length and intonation. In all six languages, postposed causal clauses tend to be longer and intonationally more independent of the associated main clause than preposed causal clauses.

The third part of the paper seeks to integrate the findings from typology and corpus linguistics into a unified account. In particular, I argue that the structure and position of causal clauses is motivated by two competing forces: A communicative strategy to support a controversial statement by an adverbial clause of reason or cause, and a general processing preference to express semantically similar constituents in formally similar types of constructions. Since the communicative strategy is especially prominent in spontaneous speech, spoken registers tend to include a larger proportion of postposed causal clauses than written genres. The paper concludes with some general remarks on the diachronic evolution of causal clauses. In the literature it is often argued that adverbial clauses have evolved from paratactic sentences (e.g. Hopper and Traugott 2003); but while this scenario is plausible to

explain the emergence of cause and reason clauses that follow the main clause, preposed causal clauses are commonly derived from adpositional phrases, relative clauses or other semantic types of adverbial clauses that precede the associated (main) clause (Diessel 2019).

References

- Diessel, H. and Hetterle, K. 2011. Causal clauses: A cross-linguistic investigation of their structure, meaning, and use. In P. Siemund (ed.), *Linguistic Universals and Language Variation*, 21-52. Berlin: Mouton de Gruyter.
- Diessel, H. 2019. Preposed adverbial clauses: Functional adaptation and diachronic inheritance. In K. Schmidtke-Bode, N. Levshina, S. Michaelis and I. A. Seržant (eds.) *Explanations in Linguistic Typology: Diachronic Sources, Functional Motivations and the Nature of the Evidence*, 191-226. Leipzig: Language Science Press.
- Hopper, P. and Traugott, E. (2003). *Grammaticalization*. Cambridge: CUP.

Linking typological and sociolinguistic data to draw inferences about contact effects: a worldwide comparative approach

Francesca Di Garbo^{1,2}, *Eri Kashima*², *Ricardo Napoleão de Souza*^{3,2}, and *Kaius Sinnemäki*²

¹ Aix-Marseille University – CNRS laboratory Parole et Langage; ²University of Helsinki;

³University of Edinburgh

Sociolinguistic typology studies the relationship between properties of language structures and the sociohistorical contexts in which languages are learned and used (Trudgill 2011). This approach entails analyzing the distribution of selected language structures in relation to selected nonlinguistic correlates, ranging from demography to geography, and degree of contact (e.g. Lupyán and Dale 2010, Bentz and Winter 2013, Sinnemäki and Di Garbo 2018). Recent studies focusing on individual linguistic areas and families show that tailoring the selection of sociolinguistic variables to the specifics of language communities is key to detecting an effect of those variables on linguistic structures (Verkerk and Di Garbo 2022). Yet, fine-tuned variable design in sociolinguistic typology may appear costly on a global level, as it requires extensive pre-analyses of the social and linguistic history of the language communities under study.

In this paper, we showcase a new approach to integrating sociolinguistic and typological perspectives on linguistic diversity, focusing on worldwide comparisons of language contact scenarios. The approach rests on a multi-prong research design where each step, from sampling to data collection and analysis, is geared towards modeling the dynamics of contact between pairs of genealogically unrelated languages (see Di Garbo et al. 2021 for an overview).

We present results from a dataset of 27 pairs of languages in contact (N= 54) from all around the world, which we approach both in terms of linguistic and sociolinguistic data. The linguistic data come from a selection of five linguistic variables: word prosody, syllable structure, nominal number, adnominal possession, and adnominal demonstrative systems. This choice of variables derives from their reported crosslinguistic frequency, and/or sensitivity to contact effects. The sociolinguistic data stem from a sociolinguistic questionnaire which we developed to assess the opportunities and nature of interactions that have existed between the pairs of language communities in six social domains: family and kin, ceremonial exchange, daily interactions, labor, knowledge, and trade. The questionnaire was filled out by specialists of the sampled languages and/or contact scenarios.

In the talk, we demonstrate the methods we use for linking the linguistic and sociolinguistic data, and how they allow us to draw inferences about contact effects. We also address how the different sociolinguistic scenarios may prompt linguistic effects in the languages of the sample. This demonstration aims to illustrate how this newly developed approach adds to the typologist's toolkit and contributes towards a principled understanding of linguistic structures as embedded in their broader linguistic and social ecologies.

References

- Bentz, Christian & Bodo Winter. 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change* 3. 1–27.
- Di Garbo, Francesca, Eri Kashima, Ricardo Napoleão De Souza & Kaius Sinnemäki. 2021. Concepts and methods for integrating language typology and sociolinguistics. In Silvia Ballarè & Guglielmo Inglese (eds.), *Tipologia e sociolinguistica: Verso un approccio integrato allo studio della variazione – Atti del workshop della società linguistica italiana 20 settembre 2020*, 143–176. Milano: Officinaventuno. <https://doi.org/10.17469/O2105SLI000005>.
- Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PLoS one* 5(1). 1–10.
- Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. New York: Oxford University Press.
- Sinnemäki, Kaius & Francesca Di Garbo. 2018. Language structures may adapt to the sociolinguistic environment, but it matters what and how you count: A typological study of verbal and nominal complexity. *Frontiers in Psychology* 9. <https://doi.org/10.3389/fpsyg.2018.01141>.
- Verkerk, Annemarie, & Di Garbo, Francesca (2022). Sociogeographic correlates of typological variation in northwestern Bantu gender systems. *Language Dynamics and Change*. 12(2), 155–223. <https://doi.org/10.1163/22105832-bja10017>.

Typology and variation of subjunctive mood. On shared (probabilistic) constraints

Salvio Digesto
Dalhousie University

Since early work on syntactic change, typologists have examined the indicative-subjunctive distinction in complementation, a well-attested phenomenon in several language families. Romance languages, such as Italian (1), present a dedicated verbal inflectional paradigm inherited from their ancestor language, Latin:

- (1) a. Credo che tutto ritorna_{IND}. (C-ORAL.511.264)
‘I believe that everything comes back.’
b. Ma puòd arsi che ritorni_{SUBJ}, guarda. (LIP.16.23)
‘But she may come back, look.’

The general consensus is that subjunctive mood is used to express an attitude of uncertainty and the speaker’s attitude towards the truth of the embedded proposition, as opposed to the indicative that is said to convey a more assertive statement (e.g. Whaley 1997; Schachter & Shopen 2008). This hereditary feature that is the Romance subjunctive would have inherited the set of meanings ascribed to it in the Latin era (e.g. dubitative, uncertain, irrealis; Leone 1949; Mellet 1994), and Italian is generally acknowledged as the Romance language that would have retained clear formal marking and substantial semantic motivation (Harris & Vincent 1997). It is commonly accepted that subordinate clauses originated from the supposed shift from paratactic to hypotactic syntactic solutions (Palmer 2001; Murphy 2008): independent clauses are reorganized into a more complex syntax with dependency relations (*volo; veniat* ‘I wish it; Let him come’ > *Volo ut veniat* ‘I want him to come’). The subjunctive in subordinate clauses is said to mirror the original meaning of the subjunctive form in the earlier paratactic structure (Harris 1978; Palmer 2001), and therefore making subjunctive morphology carrier of semantic meaning(s). However, a putative loss of the subjunctive in Italian discourse is widely assumed to be underway, associated with the loss of whatever semantic properties would have hypothetically triggered its use in the past. Likewise, research on grammaticalization suggested that the subjunctive can gradually lose its original meaning and become a mere marker of subordination (Bybee et al. 1994; Carlier et al. 2012). Recent variationist research has shown that subjunctive selection in Romance is devoid of semantic meaning, while lexical identity of the matrix verb is the major predictor of its selection (Poplack et al. 2018).

This variationist study assesses the putative linguistic change undergoing in Italian by tracing back subjunctive use in completive clauses diachronically and by determining the degree of retention from its ancestor language, Latin. This research sheds light on the underlying conditioning on variability using actual usage and speech-surrogate data. The variable context is determined corpus-internally:

every complement clause governed by a matrix verb that triggered the subjunctive at least once was included, yielding overall a dataset of nearly 3000 tokens, which were coded according to the mood selected as well as a series of potential explanatory factors.

Quantitative and statistical evidence shows that subjunctive selection is largely determined by lexical identity of the governor as well as embedded suppletive forms of *essere* ‘to be’, and that this pattern has been operative at least since the 16th century. Despite the significant time span targeted, no evidence of desemanticization has been found. Likewise, the analysis on the Latin subjunctive shows that subjunctive choice was already largely determined by, and restricted, to a few matrices (e.g. *rogo* ‘to ask’, *oportet* ‘it is necessary’), suggesting that a lexical pattern has been transferred and consistently retained in the daughter language.

References

- Bybee, J. L., Perkins, R., & Pagliuca, W. (1994). *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. University of Chicago Press.
- Carlier, A., De Mulder, W., & Lamiroy, B. (2012). Introduction: The pace of grammaticalization in a typological perspective. 46(2), 287–302.
- Harris, M. (1978). *The evolution of French syntax: A comparative approach*. Longman.
- Harris, M., & Vincent, N. (Eds.). (1997). *The Romance Languages*. Routledge.
- Leone, A. (1949). *L’uso del congiuntivo in Latino. Principi di sintassi ragionata*.
- Mellet, S. (1994). *Le subjonctif*. In S. Mellet, M.-D. Joffre, & G. Serbat (Eds.), *Grammaire fondamentale du latin. Le signifié du verbe* (pp. 173–209). Peeters.
- Murphy, M. D. (2008). *The role of typological drift in the development of the romance subjunctive: A study in word-order change, grammaticalization and synthesis*. The University of Texas at Austin.
- Palmer, F. R. (2001). *Mood and Modality*. Cambridge University Press.
- Poplack, S., Torres Cacoullos, R., Dion, N., de Andrade Berlinck, R., Digesto, S., Lacasse, D., & Steuck, J. (2018). Variation and grammaticalization in Romance: A cross-linguistic study of the subjunctive. In W. Ayres-Bennett & J. Carruthers (Eds.), *Manuals in Linguistics: Romance Sociolinguistics*. Mouton de Gruyter.
- Schachter, P., & Shopen, T. (2007). Parts-of-speech systems. In T. Shopen (Ed.), *Language Typology and Syntactic Description: Vol. Volume 1: Clause Structure* (2nd ed., pp. 1–60). Cambridge University Press.
- Whaley, L. (1997). *Introduction to Typology: The Unity and Diversity of Language*. SAGE Publications.

Typology and its data: functional monoculture or structural diversity?

John W. Du Bois and Giorgia Troiani
University of California, Santa Barbara

Typology sets itself ambitious theoretical goals, as it seeks to document and generalize about the diversity and commonalities of linguistic structures across the languages of the world. Many typologists would further subscribe to the goal of explanation, posing the fundamental functional question: Why are languages the way they are? On a methodological level, typology strives to improve its working methods by developing, for example, a more representative sampling of the world's languages in all their diversity. But the actual linguistic data that are called on to *represent* this diversity are often treated almost as an afterthought. Data may be collected using whatever method is most convenient, such as elicitation of isolated sentences, or recording of monologues. Any limitations of the data are excused by pointing to the limited resources available for documenting a given language, or the fact that the data were collected by other people for other purposes, or both. Even honoring the Boasian injunction to collect “text” leaves the deeper issues unresolved; better sampling of the same old data will not address the needs of current typology. The time is ripe for typology to rethink what it means to represent a language, and to represent the use of language.

To be sure, typology's data must simultaneously satisfy an imposing list of requirements, including cross-linguistic comparability, accuracy, coverage, and representativeness. To meet these demands efficiently, methods like sentence elicitation will doubtless remain indispensable tools in the methodological repertoire – used for convenience, for clarifying meanings, and for coverage of the full range of expected structures known to be worth looking out for. This is not the problem. But to get beyond the what is expected, the already known, the researcher must cede some control to the language users themselves, who can be counted on to use their language to fulfill their own goals, if given the chance. Typology must be prepared to be surprised by its own data. The potential to reveal new forms of structural diversity will be proportional to the functional diversity of the collected data.

Consider the case of GRAID, a cross-linguistic corpus designed to serve typology, which has many admirable features, including open access, comparability of data across languages, and more. But all the data represent just one genre of monologue. This reflects a design decision intended to create a body of comparable data that are indigenous to, and hence representative of, each sampled language. The problem is that functional monoculture breeds structural homogeneity. Such a corpus is likely to leave undocumented the functions that are found only in the back-and-forth of conversation, as well as the structures that express them.

We begin our presentation with a diagnosis of the mismatch between the laudable goals of typology and its current methods and data. We then survey some recent efforts by typologists and

others to collect corpora of supposedly “natural” discourse. We identify problems with the criteria used to assess the naturalness of speech, which can be traced back to certain assumptions about what “natural” means when applied to culture, and how it is operationalized in linguistic research. We call into question the widespread practice of treating fast speech phonetic reductions as a sufficient diagnostic, arguing instead for the need to target naturally occurring language use, defined as what language users do when they are using language for their own reasons. Situating our proposals within the broader enterprise of typology, we offer a new classification of typology’s data, distinguishing between naturally occurring language use – which we argue is the gold standard – and its useful but limited approximation in merely naturalistic speech. We locate these two along a continuum that acknowledges a continuing role for sentence elicitation and controlled experiments, as we clarify the distinctive contributions that each can make to the work of typology. To ground our proposals, we give an inside look at the methods used in two corpus construction projects we are currently involved in, for Kazakh (Turkic, Kazakhstan) and Sakapultek (Mayan, Guatemala). We present practical suggestions for how typological projects at any scale can collect the kind of discourse data that will effectively document the functional and structural diversity of naturally occurring language use.

In sum, we argue that by tapping into the rich vein of functional diversity found in naturally occurring conversation, typology will be better positioned to discover new forms of structural diversity, and to learn how what language users do drives the emergence of linguistics structure in the world’s languages.

How ubiquitous is complementation? A naturalistic cross-corpus study

Nicholas Evans, Wayan Arka, Danielle Barth, Henrik Bergqvist, Christian Döhler, Sonja Gipper, Dolgor Güntsetseg, Yukinori Kimoto, Dominique Knuchel, Hitomi Ōno, Eka Pratiwi, Saskia van Putten, Alan Rumsey, Andrea Schalley, Stefan Schnell, Asako Shiohara, Elena Skribnik, Yanti
Australian National University

Complementation constructions are one of the cornerstones of accounts of language, and language evolution, that place recursion at centre stage. Sentences like *Caterina suspects that [many typologists don't believe that [convincing accounts of need to be grounded in naturally occurring data]]]* exhibit multiple levels of complementation, and the rewrite rules they are believed to need (of the type (i) S → ... NP; (ii) NP → S) which were used by Chomsky (1958) and subsequent developments of generative grammar to argue the need for powerful transformational grammars. Psychologists of social cognition, such as de Villiers (1999), argued that the acquisition of 'theory of mind' and mastery of 'false belief tasks' in developing children went hand-in-hand with the acquisition of complement constructions that allowed them to represent states of affairs as being held in the minds of particular social actors.

Against this background, we need to ask a more basic empirical question: how widespread are complement constructions, in fact? Do all languages have them? Are they used more often for some types of complement than for others? How deeply embedded do they get, in natural discourse? These are the questions we set out to answer in this paper. We draw on two distinct types of corpus. One is the 'parallax' SCOPIC task (San Roque et al 2012, Barth & Evans 2017), which presents participants with a series of 16 picture cards, enriched for themes of social cognition (including thought, speech, fear, memory and wishes for the future), whose task structure allows speakers the chance to construct their own stories, in the process generating a series of 'parallax' corpora across different languages which treat broadly similar themes while allowing scope for different individuals to make different expressive choices, and which involves description, narrative and conversation. While we have been building the cross-linguistic SCOPIC corpus from around 30 languages of all continents, we draw on a subset of a dozen here. For each language represented, we also use a 'free' corpus – a naturalistic corpus assembled for other purposes. Taken together, this gives us a rich and naturalistic data set from a wide range of languages (Arta, Avatime, Balinese, Dalabon, German, |Gui, Indonesian, Kogi, Komnzo, Ku Waru, Matukar Panau, Mongolian, Murrinhpatha, Vera'a, Yurakare).

We then develop a set of construction-type labels for 'propositional framing' constructions – basically all constructions which can frame propositions in the way that is done by complement constructions in at least some languages. This set of functional definitions allows us to code up alternative constructional realisations, so that alongside complement constructions we also include such alternatives as paratactic quoted speech, inflections (e.g. desideratives instead of want-verbs),

particles (e.g. counterfactual particles instead of verbs like ‘believe’). We likewise develop a set of functions (e.g. ‘utterance’, ‘thought’, ‘fear’) based on the functional range of complement-taking predicates in Noonan (1985).

This approach then allows us to go back to our corpus and examine how far languages adopt the same coding strategies to express propositional framing. As we will show, the proportion of propositionally-framed constructions that use complement constructions is highly variable cross-linguistically, and while used commonly in some languages there are others in which they appear very rarely in our corpus, their place being taken by other constructions like paratactically-linked direct quotation taking its place.

Word order and information structure: new insights from conversational corpora

Natalia Levshina¹, Rowena Garcia^{1,2} and Yevheniy Skyra¹
¹MPI for Psycholinguistics, Nijmegen; ²University of Potsdam

The given-before-new order is considered as the default in many languages (e.g., Halliday 1967; Gundel 1988). Psycholinguists explain this fact by the ease of retrieval of accessible information (the principle “Easy First” by MacDonald 2013; see also Bock 1982), although a rigid word order can be advantageous for language production (cf. MacDonald’s [2013] principle “Plan Reuse”). Notably, some languages, such as Cayuga (Iroquoian, the USA) and Ngandi (Gunwinyguan, Australia) put new or newsworthy constituents first (Chafe 1994; Mithun 1992), which raises a question about how necessary the given–new order is for interaction. It has also been argued that languages with more flexible ordering of constituents implement better the principle “Easy first” and therefore can employ their word order to express information structure. Moreover, it has been claimed that language users resort to word order “freezing” (i.e., use the canonical order) in the absence of other cues to grammatical roles (Bouma 2011; Lee 2001).

In our study we investigate these popular claims using spoken corpora of German, Russian and Tagalog—three languages with some degree of A and P order flexibility. More specifically, we ask:

- 1) To what extent do these three languages with relatively flexible word order prefer given–new order of A and P arguments?
- 2) Is accessibility more important than the syntactic role in determining the order of A and P?
- 3) Does the canonical (A – P) order help to disambiguate the constituents?

Our study focuses on the order of A and P arguments in spontaneous informal conversations. We annotated at least three hundred examples of complete transitive clauses in each of the three languages for a range of variables: animacy, person, length (in words and syllables), discourse status, situational status of A and P; as well as the tense and aspect form of the predicate, the communicative type of the utterance and other information. In order to annotate the discourse and situational status, we developed a special annotation schema based on previous theoretical and practical work on information structure (Chafe 1994; Dipper et al. 2007; Prince 1982). The discourse status variable has the following values: mentioned close in discourse, mentioned far in discourse, inferable from a mentioned referent (close or far in discourse) and new. The situational status variable has the values present in the extralinguistic situation and salient, present and non-salient, inferable from a present referent (salient or non-salient) and absent. The new/absent referents are further subdivided into familiar and unfamiliar. In addition, the arguments can be non-referential.

Our descriptive analyses show that the more accessible (in discourse or situation) argument comes first in only about 50% of the clauses (see, for example, discourse accessibility in Figure 1). In many

cases, there is no difference in accessibility. The order of A and P is more systematic (more than 75% AP). For multivariate analyses, we used non-parametric machine learning techniques (conditional random forests and artificial neural networks). The reason is that the number of variables (and their values) is considerable, relative to the number of examples. The accuracy and predictive power of the trained models applied to test datasets is high (about 90%). We observe little effect of the discourse or situational status on the order of A and P. Instead, the most influential variable is the role itself (A or P). This is an unexpected finding, which puts the role of information structure in a new light. Moreover, we find that we have virtually no ambiguity in formal expression of the roles in the data. This means that the canonical order is used without the need for disambiguation. We explain our findings from different perspectives, such as the noisy channel communication account (Gibson et al. 2013) and MacDonald’s (2013) principle of “Plan Reuse”.

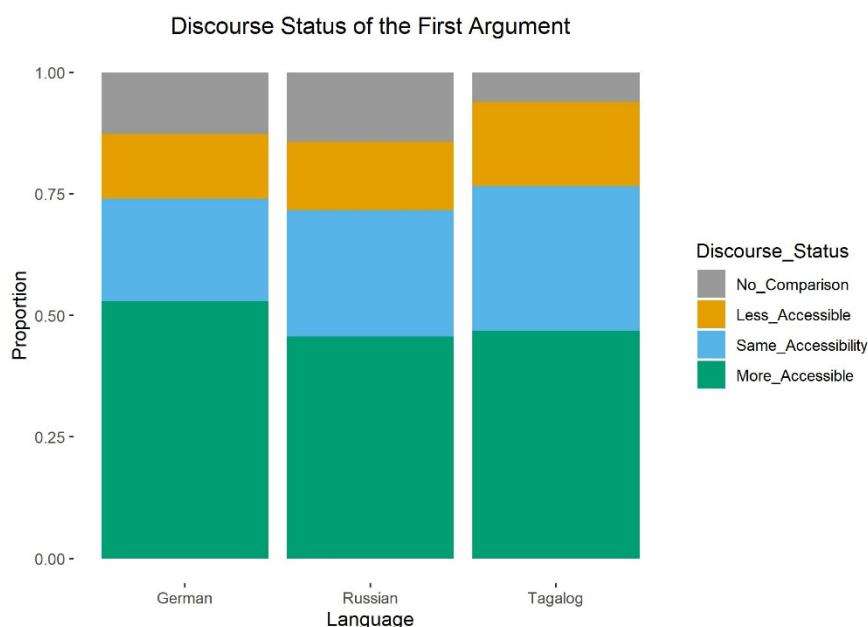


Figure 1. Discourse status of the first argument (A or P) in relation to the second argument.

References

- Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review* 89(1), 1-47.
- Bouma, G., (2011). Production and Comprehension in Context: the Case of Word Order Freezing. In A. Benz & J. Mattausch (eds.), *Bidirectional Optimality Theory*, 169-190. Amsterdam: Benjamins.
- Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press.

- Dipper, S., Goetze, M. & Skopeteas, S. (2007). Information Structure in Cross-Linguistic Corpora: Annotation Guidelines for Phonology, Morphology, Syntax, Semantics and Information Structure. ISIS Working Papers of the SFB 632. Potsdam: Universitätsverlag Potsdam.
- Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences* 110(20), 8051–8056.
- Gundel, J. K. (1988). Universals of topic-comment structure In M. Hammond, E. Moravczik and J. Wirth (eds.), *Studies in syntactic typology*, 209-239, Amsterdam: Benjamins.
- Halliday, M. (1967). Notes on transitivity and theme in English, Part 2. *Journal of Linguistics* 3, 199-244.
- Lee, H. (2001). Markedness and word order freezing. In P. Sells (ed.), *Formal and empirical issues in optimality theoretic syntax*. Stanford: CSLI Publications.
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, 226.
- Mithun, M. (1992). Is basic word order universal? In Payne, D. L. (ed.), *Pragmatics of word order flexibility*, 15–61. Amsterdam: Benjamins.
- Prince, E. F. (1981) Toward a taxonomy of given/new information. In P. Cole (ed.), *Radical Pragmatics*, 223–254. New York: Academic Press.

Ze ma she-'ani tsarix 'that's what I need': 'Reversed pseudo-clefts' in Hebrew conversation

Yael Maschler and Hilla Polak-Yitzhaki
University of Haifa

Employing Interactional Linguistic (Couper-Kuhlen/Selting 2018) methodology, this study investigates Hebrew [*ze ma she-* 'this is what'+ clause] structures, 'equivalent' to structures termed 'reversed pseudo-clefts' in, e.g., English (Collins 1991, Oberlander/Delin 1996) and French (Lambrecht 2001). The corpus consists of 9 hours of video-recorded casual conversation among friends and relatives, manifesting 71 *ze ma she-* tokens.

Such structures are analyzed according to traditional Hebrew syntax as belonging in the realm of complex syntax, with *ze* considered subject of a nominal (copular) matrix clause and the [*ma she-* + clause] – a nominalized clause embedded as predicate. However, we argue that the *ze ma she-* construction has become a fixed chunk, grammaticized from repetitive discourse actions. Information Structure based studies of reversed pseudo-clefts claim that the demonstrative pronoun denotes Given information, while the clause, which contains the prosodic nucleus of the tone unit, denotes New information (e.g., Collins 1991), and that the discourse function of reversed pseudo-clefts is predominantly a summative one. We have found that the Hebrew structure serves mainly other functions, can even be deployed as a projecting construction (Auer 2005; cf. Küttner 2020), and may manifest additional prosodic patterns. Despite similarities in lexical and syntactic features with pseudo-clefts (e.g., Hopper/Thompson 2008, Maschler/Pekarek Doehler in press), the functions of the *ze ma she-* construction are different.

The construction is usually deployed in our corpus at a moment of (slight) dispute, and it implements one of five actions: 'claim-backing' (Antaki/Leudar 1990); getting back to a previous topic; framing prior-talk metalingually; disclaiming responsibility; seeking clarification. For example, tokens functioning as 'claim-backing' present evidence supporting the soundness of a previous disputable assertion, as in the following suggestion to heat up water in the microwave:

```
1 Ya'ir:      ....'efshar      lexa^mem mayim ba-mikro.
              it's_possible warm.INF water in-the-micro
              ....it's possible to heat up water in the micro.
              ^left hand in PUOH--> 1. 5

2 Inbal:     .../ə/

3 Omri:      (0.34) # (0.67) ken?
              yes
              ....really?
              #tilts head-> 1. 5

4 Ya'ir:     ..%m m--%
              ..m m--
```

%moves hand in PUOH horizontally to the left while shaking head%

→ 5 ze ma she-hu yode'a la'a#sot. ^
 this.M.SG what that-he know.PRS.M.SG do.INF
 that's what it can do.

...

12 Ya'ir: ...hu mexamem 'otam.
 he warm.PRS.M.SG OBJ-3M.PL
 ...**it heats them** [the water molecules] **up.**

Following Ya'ir's suggestion (line 1), Omri doubts the plausibility of heating up water in the microwave with *ken?* 'really?' (line 3). In response, Ya'ir produces a [*ze ma she-* + clause] structure manifesting two primary stresses – one on the demonstrative pronoun *ze*, the other on the verb *do* – supporting the soundness of his suggestion: 'that's what it can do' (line 5). The construction is accompanied here by a head shake and moving the hand horizontally away from the body while holding it in the PUOH (Palm Up Open Hand) gesture (Kendon 2004).

We explore the correlation between the functions of the construction and its lexical, syntactic, and prosodic features, its sequential position, as well as the speaker's embodied conduct, thus shedding light on the interlaced nature of grammar, the body, and interaction.

References

- Antaki, Charles & Leudar, Ivan. 1990. Claim-backing and other explanatory genres in talk. *Journal of Language and Social Psychology* 9: 279-292.
- Auer, Peter. 2005. Projection in interaction and projection in grammar. *Text* 25: 7-36.
- Collins, Peter C. 1991. *Cleft and Pseudo-Cleft Constructions in English*. London and New York: Routledge.
- Couper-Kuhlen, Elizabeth & Selting, Margret. 2018. *Interactional Linguistics. Studying Language in Social Interaction*. Cambridge: Cambridge University Press.
- Hopper, Paul J. & Thompson, Sandra A. 2008. Projectability and clause combining in interaction. In: Ritva Laury (ed.), *Crosslinguistic Studies of Clause Combining*, Amsterdam: John Benjamins, 99-123.
- Kendon, Adam. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Küttner, Uwe-A. 2020. Tying sequences together with the [*That's* + *Wh*-Clause] format: On (retro-) sequential junctures in conversation, *Research on Language and Social Interaction* 53 (2): 247-270. DOI: <https://doi.org/10.1080/08351813.2020.1739422>

- Lambrecht, Knut. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39(3): 463–516.
- Maschler, Yael and Pekarek Doehler, Simona. In press. Pseudo-cleft-like structures in Hebrew and French conversation: The syntax-lexicon-body interface. *Lingua*.
- Oberlander, Jon & Delin, Judy. 1996. The function and interpretation of reverse *wh*-clefts in spoken Discourse. *Language and Speech* 39(2-3): 185–227.

Object omission in naturally occurring data and antipassives

Caterina Mauri, Andrea Sansò

University of Bologna / University of Insubria

Research on antipassives has long recognized some features of these constructions for which various functional and formal explanations have been proposed (see Sansò 2018 for a survey). In almost half of the languages with an antipassive, for instance, the antipassive is lexically restricted, i.e. it is possible with only a subset of transitive verbs (see Polinsky 2013; Heaton 2017: 161). Cases of lexically restricted antipassives include Eton (van de Velde 2008: 129), Hup (Epps 2005: 405-407), and Ainu (Bugaeva 2021). Typically, in these languages the antipassive is possible with verbs of ingestion, interaction, communication, or traditional activities, but is incompatible with action verbs. A related feature is the correlation between antipassives and imperfective states of affairs: if an antipassive construction can have a perfective interpretation, it must also have an imperfective one (Polinsky 2017: 215-216). While the latter universal tendency is generally explained in functional-typological terms as an effect of imperfectivity on the identifiability of object arguments, the frequent lexical restrictions on antipassive derivation have fostered lexicalist approaches to this type of constructions (see Polinsky 2017).

In diachronic-typological terms, these features have been explained as a case of persistence of features of the source constructions from which the antipassive emerges through grammaticalization (see Sansò 2017, 2018). Some of these sources, for instance, exclusively or predominantly appear in imperfective contexts: a case in point is the construction formed by means of the verb *do* combined with an action noun (as in *do the washing*), which more often than not appears in imperfective contexts such as habitual, iterative or durative predications. In other cases, the source construction implies an element that automatically triggers a non-specific (and thus, less perfective) configuration: this is what happens when generic objects such as ‘thing(s)’ are grammaticalized into antipassive markers (Sansò 2018).

The aim of this paper is to test the hypothesis whether these two features of antipassives correlate with tendencies with respect to object omission in naturally occurring data. We will do so by analyzing object omission in a corpus of spoken Italian (The KIParla corpus, cf. Mauri et al. 2019, Ballarè et al. 2022), a language that has no dedicated antipassive construction. We will show that object omission in spoken Italian is significantly more frequent with the same verb classes that allow the antipassive derivation more straightforwardly in languages with lexically restricted antipassives. Moreover, we will also show that these verb classes tend to combine more frequently than other classes with generic objects such as *cosa/cose* (‘thing(s)'), *roba* (‘stuff’) and *gente* (‘people’) and to form constructions involving the verbs *dare* (‘give’) and *fare* (‘do’) and a verbal noun (e.g. *dare una*

pulita, lit. ‘give a cleaning’, i.e. ‘to clean up (intr.)’). The results of the survey suggest that typological tendencies are deeply rooted in language usage and that most of the diversity characterizing antipassives across languages can find a sound explanation by combining diachronic insights with the analysis of naturally occurring data.

References

- Ballarè, Silvia, Eugenio Gorla & Caterina Mauri. 2022. *Italiano parlato e variazione linguistica. Teoria e prassi nella costruzione del corpus KIParla*. Bologna: Pàtron.
- Bugaeva, Anna. 2021. Unspecified participant. A case of antipassive in Ainu. In Katarzyna Janic & Alena Witzlack-Makarevich (eds.), *Antipassive. Typology, diachrony, and related constructions*, 213-246. Amsterdam: John Benjamins.
- Epps, Patience. 2005. *A grammar of Hup*. Berlin: Mouton de Gruyter.
- Heaton, Raina. 2017. *A typology of antipassives, with special reference to Mayan*. PhD Dissertation, University of Hawai‘i at Mānoa.
- Mauri, Caterina, Silvia Ballarè, Eugenio Gorla, Massimo Cerruti & Francesco Suriano. 2019. KIParla corpus: a new resource for spoken Italian. In Raffaella Bernardi, Roberto Navigli, Giovanni Semeraro (eds.), *Proceedings of the 6th Italian Conference on Computational Linguistics CLiC-it*. (URL: <http://ceur-ws.org/Vol-2481/paper45.pdf>)
- Polinsky, Maria. 2013. Antipassive constructions. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures online*. Leipzig: The Max Planck Institute for Evolutionary Anthropology. (URL: <https://wals.info/chapter/108>)
- Polinsky, Maria. 2017. Antipassive. In Jessica Coon, Diane Massam & Lisa deMena Travis (eds.), *The Oxford handbook of ergativity*, 307-331. Oxford: Oxford University Press.
- Sansò, Andrea. 2017. Where do antipassive constructions come from? A study in diachronic typology. *Diachronica* 34: 175-218.
- Sansò, Andrea. 2018. Explaining the diversity of antipassives: Formal grammar vs. (diachronic) typology. *Language and Linguistics Compass* 12(6): 1-22. (URL: <https://doi.org/10.1111/lnc3.12277>)
- Van de Velde, Mark L. O. 2008. *A grammar of Eton*. Berlin: Mouton de Gruyter.

Embodied typology: gesture and prosody in traditional storytelling

Amina Mettouchi

EPHE, Paris/CNRS-LLACAN

Typology has so far been concerned mostly about structural and systemic diversity, focusing on phonological, grammatical and lexical properties of languages (cf. WALS: <https://wals.info/>, GLOTTOBANK: <https://glottobank.org/>). But as more and more spoken and signed corpora of lesser-described languages are being compiled (cf Haig, Schnell & Seifart (eds) 2022), more and more fine-grained comparative studies can be conducted (Mello, H., A. Mettouchi, M. Mithun, A. Panunzi, T. Raso 2021, Mettouchi & Vanhove 2022).

Effects of genre, as well as individual and cultural variables, are now being considered, paving the way for more complex and subtle representation of language variation. This zooming in is accompanied with a focus on the more embodied dimensions of speech, voice and gesture, thus potentially putting to the fore a major dimension of indigenous languages, their oral tradition.

It is this cultural dimension of variation I am interested in, and I would like to propose ways in which it can be included in corpus-based typological studies, centering on the corpus-encoding and analysis of phenomena pertaining to prosody (audible breath intakes, prosodic segmentation) and gesture (open-hand palm gestures), in Berber/Amazigh traditional storytelling (Mettouchi 2018, Ferré & Mettouchi 2020, Mettouchi & Vanhove 2022).

References

- Atkinson, Q. and R. Gray (co-leaders). GLOTTOBANK. (Project page: <https://glottobank.org/>)
- Dryer, Matthew S. & Haspelmath, Martin (eds.) 2013. The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <https://wals.info/>).
- Ferré, G. & A. Mettouchi (2020) A Cultural Study of Open-Palm Hand Gestures and their Prosodic Correlates, Proceedings of 10th International Conference on Speech Prosody SP2020, 285-289.
- Haig, Geoffrey & Schnell, Stefan & Seifart, Frank (eds.), Doing corpus-based typology with spoken language data: State of the art. Language Documentation and Conservation °SP25. Honolulu, HI: University of Hawai'i Press. <https://nflrc.hawaii.edu/lcd/sp25/>
- Mettouchi, A. (2018). Audible breath intakes in monologues. JOSS (Journal of Speech Sciences) 7(2):93-106.
- Mello, H., A. Mettouchi, M. Mithun, A. Panunzi, T. Raso (2021). Prosody and Corpora. Cadernos de Linguística, v. 2, n. 1, e385.

Mettouchi, A. & M. Vanhove. (2021) Prosodic segmentation and cross-linguistic comparison in CorpAfroAs and CorTypo: Corpus-driven and corpus-based approaches. In Haig, Geoffrey & Schnell, Stefan & Seifart, Frank (eds.), *Doing corpus-based typology with spoken language data: State of the art*, 59–113. Honolulu, HI: University of Hawai'i Press.

Typology and usage: explanatory perspectives with special reference to negation

Matti Miestamo, Ksenia Shagal, Olli Silvennoinen and Chingduang Yurayong
University of Helsinki

Usage-based approaches have recently gained ground in the field of linguistic typology. On the one hand, more and more typological work is being done based on corpus data and, on the other, the role of language use as an explanatory factor in typology is growing (for recent discussion on these two perspectives, see Levshina 2022 and Miestamo et al. 2022, respectively). In this talk, we will focus on the latter perspective, addressing the question of how naturally occurring discourse data can be used to account for cross-linguistic generalizations. For concrete illustration, we will draw from our recent typological and corpus-based work on negation. We argue that negation is a particularly suitable domain for this kind of discussion, since (a) it is universally present in the world's languages and sufficiently frequent in discourse, (b) it is relatively well-studied from a typological point of view, and (c) due to its special pragmatic properties, it shows a variety of discourse effects that can be possibly linked to cross-linguistic generalizations.

Usage-based explanation of cross-linguistic patterns has been an important element in typological research ever since Greenberg 1966. Well-known examples of this line of thinking include Givón 1979 and Hopper and Thompson 1980. More recently, especially Haspelmath (2008, 2021) and Diessel (2019) have developed these ideas further. The relationship between usage and grammar is crystallized in Du Bois's (1985: 363) famous words, "[g]rammars code best what speakers do most" and further specified in Hawkins's (2004) Performance Grammar Correspondence Hypothesis: "Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments." (Hawkins 2004: 3). Crucially for the present discussion, it is to be expected that the discourse frequency of a linguistic feature will correlate with its cross-linguistic frequency, i.e. what is frequent in discourse, will also be frequently grammaticalized across languages.

Even though the relationship between discourse frequency and cross-linguistic frequency has received quite a lot of attention and explanations have been proposed along these lines in many typological studies, systematic discourse data has rarely been presented to support the proposed explanations. In this talk, we will address the possibilities and potential problems of usage-based explanations for cross-linguistic generalizations put forward in language typology. Focusing on the typology of negation, we will briefly chart the discourse-based explanations of cross-linguistic findings proposed in the literature (see also Miestamo & al. 2022) and then present a few case studies that allow us to discuss the relationship between grammar and discourse in more concrete terms.

References

- Diessel, Holger. 2019. *The grammar network: How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.
- Du Bois, John W. 1985. Competing motivations. In John Haiman (ed.), *Iconicity in syntax* (Typological Studies in Language 6), 343–365. Amsterdam/Philadelphia: John Benjamins.
- Givón, Talmy. 1979. *On understanding grammar* (Perspectives in Neurolinguistics and Psycholinguistics). New York: Academic Press
- Greenberg, Joseph H. 1966. *Language universals, with special reference to feature hierarchies*. The Hague: Mouton.
- Haspelmath, Martin. 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19 (1). 1–33.
- Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. *Journal of Linguistics* 57 (3). 605–633.
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56: 251–299.
- Levshina, Natalia. 2022. Corpus-based typology: Applications, challenges and some solutions. *Linguistic Typology* 26 (1). 129-160.
- Miestamo, Matti, Ksenia Shagal & Olli O. Silvennoinen. 2022. Typology and usage: The case of negation. *Linguistic Typology at the Crossroads* 2 (1). 121-154.

The mighty demonstrative

Marianne Mithun

University of California, Santa Barbara

A well-known source of grammatical constructions is the routinization of frequently-occurring discourse patterns. Such processes can unfold over long spans of time, however, and philological records of the kinds of unscripted, interactive spoken language in which they take place are rarely available at such time depths. Examination of speech in modern languages, however, can sometimes provide insight into discourse patterns that might evolve into grammatical constructions. Modern documentation can add the important element of prosodic structure.

The point of departure here is demonstratives. Demonstratives can of course serve as pronouns: *That is my hat*. In many languages they also serve as determiners, and in some they have also developed into third person pronouns, relativizers, complementizers, and more. Both their frequency in speech and the discourse functions they serve show interesting variation across languages, differences which might set the stage for different grammatical developments. Some of these are examined here in three areally and genealogically unrelated languages: Central Pomo (Pomoan) of Northern California, Central Alaskan Yup'ik (Inuktitut-Yupik-Unangan) of Alaska, and Navajo (Dene) of the North American Southwest.

In all three, demonstratives function to maintain coherence in discourse within and across speaker turns, referring to a previously introduced referent. The Central Pomo demonstrative has developed in a direction the others have not. In this language, unlike the others, core arguments are not identified within verbs. First and second persons are identified by independent pronouns, but continuing topical third persons are not generally mentioned overtly. When topic continuity has been interrupted, however, the demonstrative is used for accessible referents. This marker is now moving into a position in the pronominal paradigm and is inflected for number and case when referring to humans.

Demonstratives play important roles in regulating the flow of information in all three languages. As elsewhere, speakers tend to introduce one significant new idea at a time, in separate intonation units. In many languages demonstratives serve as place holders while the speaker formulates the next idea. These tendencies have crystallized into various recurring discourse patterns and grammatical constructions. In Yup'ik, topic shifts are often introduced with a demonstrative pronoun in one intonation unit, followed by an appositive nominal in the next: 'I'm enjoying staying at the house. **Those** (known to both of us), your parents, I take care of them'. Navajo contains a set of demonstratives which function both as pronouns and determiners, but the most common demonstrative functions only as a pronoun. It frequently enters into topic shift constructions as well. Here a new topic is typically introduced in one intonation unit followed by this demonstrative and a

pause, then a comment in the next: ‘We herded goats and sheep. We usually didn’t plant. We made that our meat. We rode horses around. **My father that**, um, he was a silversmith.’ The demonstrative has become a marker of a topic shift.

In Yup’ik and Central Pomo the demonstratives also appear frequently in antitopic constructions, re-identifying a topical referent following the nuclear clause. These may consist just of the demonstrative pronoun, or the demonstrative + nominal together representing an established referent, as in the following Yupik exchange. A: ‘I want to talk about X. B: ‘**That** is probably a story, **that**. We’ll do it later, **that story**.’ Similar antitopic constructions occur in Central Pomo. A: Oh, that kind of people. they leave each other and throw the kids away. B: I always feel sorry for them, **those kids**. These frequent constructions appear to be a gateway into much less frequent determiner + nominal constructions, which can identify a well-established referent within the nuclear clause.

Variation and typology in Romance: a cross-linguistic study of subjunctive selection in spontaneous speech

Shana Poplack¹, Rena Torres Cacoullos², Nathalie Dion¹, Rosane de Andrade Berlinck^{3/1}, Salvio Digesto⁴, Dora Lacasse² and Jonathan Steuck²

¹University of Ottawa, ²The Pennsylvania State University, ³Universidade Estadual Paulista,

⁴Dalhousie University

Most scholars view mood selection as semantically motivated (with real, asserted, etc. predicates expressed in indicative, and irrealis complements selecting subjunctive). But typological studies featuring a diachronic perspective suggest that subjunctives grammaticalize into concomitants of subordination (Bybee, Perkins & Pagliuca 1994). Such grammaticalization would entail lower rates of subjunctive morphology and “vacuous” variability (i.e. alternation that is *not* semantically motivated). In this paper, we make novel use of *variable conditioning profiles* to address these competing perspectives, focusing on Romance complement clauses, and two major parameters of grammaticalization: semantic bleaching (desemanticization) and “obligatorification” (Lehmann 1995).

The data on which our cross-linguistic comparison is based come from four corpora of spoken French (FR; Poplack 1989), Italian (IT; Cresti & Moneglia 2005), Spanish (SP; Martín Butragueño & Lastra 2011; 2012; in prep) and Portuguese (PTG; Gonçalves 2003). Systematic extraction of every clause embedded under a matrix verb that governed at least one unambiguous subjunctive resulted in a dataset of nearly 5000 tokens. These were coded according to the mood actually selected, as well as a number of potentially explanatory conditioning factors.

Results show robust variability in subjunctive selection under putative governors, not only cross-linguistically, but also within languages. In all four, the same matrix verb can co-occur with both subjunctive and indicative in the same context, as exemplified with ‘believe’ in (1) – (4):

(1) a. Je **crois** pas que ce *soit*_[SUBJ] la fin du monde. (FR.060.195)

‘I don’t think that it would be the end of the world.’

b. Je **crois** pas que l’âge *a*_[IND] tant à faire que ça. (FR.003.189)

‘I don’t think that age has that much to do with it.’

(2) a. Eu **acredito** que *vá*_[SUBJ] *sair*. (PTG.143.356)

‘I believe that it will come out.’

b. Eu **acredito** que ele *devia*_[IND] *ter* em torno de setenta anos de idade. (PTG.99.148)

‘I believe that he must be around seventy years old.’

(3) a. **Credo** che tutti lo *sappiate*_[SUBJ]. (IT.438.218)

‘I believe that everyone knows it.’

b. **Credo** che tutto *ritorna*_[IND]. (IT.511.264)

‘I believe that everything comes back.’

(4) a. No **creo** que *haya*_[SUBJ] nadie aquí que no pague la renta. (SP.073.668)

‘I don’t think there’s anybody here who doesn’t pay rent.’

b. No **creo** que *hay*_[IND] que firmar. (SP.086.555)

‘I don’t think you have to sign’.

Moreover, contrary to received wisdom, semantic considerations turn out to play a minor role, if any, in variant choice. Instead, lexical bias is the major predictor. Tellingly, however, the “same” governor (whether determined etymologically or as translation equivalents) fails to display consistent associations cross-linguistically. Thus ‘*fear*’ co-occurs with subjunctive 33% of the time in PTG, 64% of the time in FR and 100% of the time in IT and SP.

To gauge productivity of the subjunctive, therefore, we introduce three measures: 1) the contribution of the governor (as instantiated by independent measures of its associated rate, the proportion it represents of the governor pool, and how much subjunctive morphology it accounts for), 2) the dispersion of subjunctive morphology across embedded verb types and 3) the extent to which the subjunctive is associated with particular structural contexts.

Results show that in each language, a handful of governors accounts for at least half of the governor pool and a large proportion of all subjunctive morphology. In addition, the cohort of embedded verbs carrying subjunctive morphology, though theoretically unrestricted, is also extremely limited. With near categorical subjunctive selection under the smallest number of governors and embedded verbs, FR is indisputably the least productive. But the usage facts militate against productivity in all four languages, with specific governors either highly or rarely associated with subjunctive. Importantly, these associations are *speech community*-specific and not a function of meaning. Our analysis confirms the utility of variable conditioning profiles in determining typological differences and similarities among languages, even when these are characterized by robust and deceptively chaotic surface variability.

References

- Bybee, Joan/Perkins, Revere/Pagliuca, William (1994), *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*, Chicago, University of Chicago Press.
- Cresti, E. & Moneglia, M. (eds.) (2005). *C-ORAL-ROM. Integrated reference corpora for Spoken Romance Languages*. Amsterdam, Philadelphia: John Benjamins.

- Gonçalves, S.C.L. (2003). Banco de dados Iboruna: amostras eletrônicas do português falado no interior paulista. <http://www.iboruna.ibilce.unesp.br/>.
- Lehmann, C. 1995. *Thoughts on Grammaticalization*, 2nd revised edition. Munich: Lincom.
- Martín Butragueño, P. & Lastra, Y. (eds.), 2011; 2012; in preparation. *Corpus sociolingüístico de la ciudad de México*. México: El Colegio de México.
<http://lef.colmex.mx/Sociolingüística/CSCM/Corpus.htm>
- Poplack, S. 1989. The care and handling of a mega-corpus. In R. Fasold & D. Schiffrin (eds.), *Language change and variation*. Amsterdam: Benjamins. 411–451.

Relatively complex or relatively simple? Toward new ways of analyzing language variation

Benedikt Szmrecsanyi
KU Leuven

In this presentation, I investigate the relative complexity (i.e., difficulty, see Miestamo 2009) incurred by having to choose between competing grammatical variants. While variational linguists provide overwhelming evidence for the existence, ubiquity, and systematicity of variable patterns — or “alternate ways of saying ‘the same’ thing” (Labov, 1972: 188), as in *Tom picked up the book* versus *Tom picked the book up* — there are still language mavens and theoretical linguists who dismiss or deplore variability as a matter of doctrine or explain it away (erroneously) as noise or negligible. Nonetheless, the assumption that grammatical variation *could* create undue complexity for language users is not entirely unreasonable. The idea that grammatical variation might burden language production deserves scrutiny not primarily because language users are forced to make grammatical choices — after all, using language *always* entails plenty of choice-making — but additionally because grammatical variation (as opposed to e.g., lexical variation) is typically conditioned probabilistically by any number of contextual constraints (constituent length, animacy, information status, etc.). Even before language users can make a choice as a function of the naturalness of a grammatical variant in a specific linguistic context, they need to check that linguistic context for the various constraints that regulate the variation at hand. It follows that this extra cognitive work must increase cognitive load. Or does it?

Against this backdrop, I report on the extension of a study (Gardner et al. 2021) that explores the link between production difficulty and grammatical variability using a corpus-based research design. The idea is that if isomorphism à la Haiman (1985) and No Synonymy à la Goldberg (1995) are design features of human languages, then variation — to the extent that it exists — should be suboptimal. Suboptimality, in turn, should be measurable by quantifying the extent to which variation contexts attract production difficulties.

Contrary to expectation, analysis based on the Switchboard Corpus of American English (542 speakers, 240 hours of recording) shows that the presence of variable contexts does not positively correlate with two metrics of production difficulty, namely filled pauses (*um* and *uh*) and unfilled pauses (speech planning time). When 20 morphosyntactic variables are considered collectively ($N = 57,660$ choice contexts), there is no positive effect. In other words, choice contexts do not correlate with measurable production difficulties. These results challenge the view that grammatical variability is somehow sub-optimal for speakers.

I will conclude by speculating that the putative difficulties introduced by optionality in syntactic structure or morphological realization are offset by a number of counterbalancing benefits of the flexibility to express the same grammatical concept using more than one form.

References

- Gardner, Matt Hunt, Eva Uffing, Nicholas Van Vaeck & Benedikt Szmrecsanyi. 2021. Variation isn't that hard: Morphosyntactic choice does not predict production difficulty. (Ed.) Stefan Th. Gries. PLOS ONE 16(6). e0252602. <https://doi.org/10.1371/journal.pone.0252602>.
- Goldberg, Adele E. 2003. Constructions: a new theoretical approach to language. Trends in Cognitive Sciences 7(5). 219–224. [https://doi.org/10.1016/S1089-1768\(03\)00050-0](https://doi.org/10.1016/S1089-1768(03)00050-0).
- Haiman, John. 1980. The Iconicity of Grammar: Isomorphism and Motivation. Language 56(3). 515. <https://doi.org/10.2307/414448>.
- Labov, William. 1972. Sociolinguistic patterns. Philadelphia: University of Philadelphia Press.
- Miestamo, Matti. 2009. Implicational hierarchies and grammatical complexity. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), Language Complexity as an Evolving Variable, 80–97. Oxford: Oxford University Press.

Sociolinguistics meets typology: the value of vernacular speech to account for cross-linguistic patterns

Sali Tagliamonte
University of Toronto

In this paper, I explore how vernacular speech data exposes cross-dialectal contrasts, which can, in turn, be used for insights into cross-linguistic patterns. The data come from the Ontario Dialects Project (ODP), an 11 million word archive of conversations with individuals from 21 communities across the largest province of Canada, Ontario. Because the ODP archive comprises geographic, social, and linguistic contrasts, the analyst is able to probe language features in these materials from varying perspectives. Comparative sociolinguistics (Tagliamonte, 2002) and variationist analysis (Tagliamonte, 2006; 2013) provide the theoretical approach and methodological toolkit, offering a quantitative, empirical means to understand the systems underlying language variation.

Using several different grammatical systems as case studies, I will demonstrate how cross-dialectal regularities expose typological tendencies. A study of negation, (1), reveals that the soft syntax of English indefinites and negative objects lines up with the hard syntax of the same expressions in closely related languages (Burnett et al., 2018:103). A situation of community bilingualism and similar patterns in the doubling of subjects between Anglophones and Francophones, (2), elucidates how generational trajectories of alignment towards a common constraint hierarchy is able to disentangle linguistic from discourse influences (Tagliamonte & Jankowski, under revision). The broader context of the ODP also offers geographic and social comparisons that are also key opportunities for insight. For example, the use of possessive *the* (3) (e.g. Gardner & Tagliamonte, 2020) shows there is a contrast between Canadian and British patterns based on communal possession. The choice relative pronoun *who* over *that* (4), shows how the influence of historically prestigious French and Latin in the UK at earlier times has led to longitudinal prestige norms demarcating socio-stylistic continua (D'Arcy & Tagliamonte, 2010).

1. a. There were *no* jobs to be had
b. There weren't *any* great places to eat.
2. a. **The onions, *they*** last about two months.
b. The onions last about two months.
3. 'Cause *the* house, I mean *our* house, it's not that big.
4. There's one lady *that* lives in my building *who* had been in a concentration camp...

These findings provide glimpses into the mechanisms of frequency, preference and constraints to account for cross-linguistic variation. While the evidence is multi-faceted, often situated in cultural distinctions and deeply tied to influences of context, linguistic patterns are fundamental. The findings arising from these studies of language use in vernacular interaction, employing statistical techniques and a comparative perspective offer important means to detect complex and subtle linguistic patterns. Critically, patterns in single variety, across communities and across dialects and languages seem to be a systematic continuum of interlocking patterns, which can lead to more integrated explanations for understanding universal grammar.

References:

- Burnett, H., Tagliamonte, S. A. & Koopman, H. (2018). Soft Syntax and the Evolution of Negative and Polarity Indefinites in the History of English. *Language Variation and Change* 30(1): 83-107.
- D'Arcy, A. & Tagliamonte, S. A. (2010). Prestige, accommodation and the legacy of relative *who*. *Language in Society* 39(3): 383-410.
- Gardner, M. & Tagliamonte, S. A. (2020). The bike, the back, and the boyfriend: Confronting the "definite article conspiracy" in Canadian and British English. *English World Wide* 41(2): 226-255.
- Tagliamonte, S. A. (2002). Comparative sociolinguistics. In Chambers, J. K., Trudgill, P. & Schilling-Estes, N. (Eds.), *Handbook of language variation and change*. Malden and Oxford: Blackwell Publishers. 729-763.
- Tagliamonte, S. A. (2006). *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.
- Tagliamonte, S. A. (2013). Analysing and interpreting variation in the sociolinguistic tradition. *Research Methods in Language Variation and Change*. Cambridge: Cambridge University Press. 382-401.
- Tagliamonte, S. A. & Jankowski, B. (under revision). Doubled subjects in spoken English: Insights from sociolinguistic typology.

Language variety as the minimal grammatical system in a typological perspective

Alessandro Vietti¹ and Massimo Cerruti²

¹Free University of Bozen/Bolzano, ²University of Turin

Linguistic typological research based on corpora has increasingly considered the internal variation within each language. This has mostly been done by representing the presence of a linguistic feature as a frequency or probability of occurrence (Szmrecsanyi 2009, Levshina 2016, Gerdes et al. 2021), or by incorporating internal variability of a language through measures of entropy (Levshina 2019). In this strand of research, a minor role has been played by language variety as a conceptual tool to describe the organization of a language into sub-systems, each of which being distinct by social indexing and characterized by different ways of clustering elements (according to the notions of coherence and cohesion, cf. Guy & Hinskens 2016). We present here two empirical case studies from which we derive implications for the analysis of language-internal variation in a typological perspective. (a) Allophonic relations in contact phonology (cf. Vietti & Spreafico 2018; Spreafico & Vietti 2016): the study of a corpus as reflecting a single grammatical system may lead to erroneous inferences about the nature of grammatical relations, in this case the high variability in the realization of /r/ (in Italian and Tyrolean dialect) and the consequent problem of defining its contextual distribution is resolved by introducing speaker information. This produces the partitioning of variation into socially based varieties, i.e., distinct subsystems of allophonic relations emerge. Relevant to this case is the idea that language variety may function as the minimum field of organization for grammatical relations. (b) The coexistence of two standard varieties: as with most European languages, Italian is currently characterized by the coexistence of the traditional, codified, literary standard and a newly-emerged standard which includes many informal spoken features, each consisting of bundles of features that actually co-occur in speakers' usage (cf. Cerruti & Vietti 2022); moreover, most features of the 'old' standard appear to characterize a language type which differs in some respects from that of the 'new' standard (see also Grandi 2019); this provides us with an opportunity to elaborate on the problem of choosing a point of reference in typological research (cf. Round & Corbett 2020). The implications that can be formulated from the two cases will therefore aim to answer the following question: what are the theoretical and empirical consequences if we do not subdivide language-internal variation into varieties, but treat it as a uniform, unstructured space?

References

Cerruti, Massimo & Vietti, Alessandro. 2022. Identifying language varieties: Coexisting standards in spoken Italian, in: Karen Beaman, Gregory R. Guy (Eds.), *The coherence of linguistic communities: Orderly heterogeneity and social meaning*, Routledge, London, pp. 261-280.

- Gerdes, Kim, Sylvain Kahane and Xinying Chen. 2021. Typometrics: From Implicational to Quantitative Universals in Word Order Typology. *Glossa: a journal of general linguistics* 6(1): 17. 1–31.
- Grandi, Nicola. 2019. Che tipo, l'italiano neostandard!. In B. Moretti, A. Kunz, S. Natale, & E. Krakenberger (Eds.), *Le tendenze dell'italiano contemporaneo rivisitate*, Officinaventuno, Milano, pp. 59-74.
- Guy, Gregory R., and Frans Hinskens. 2016. "Linguistic Coherence: Systems, Repertoires and Speech Communities." *Lingua* 172–173:1–9.
- Levshina, Natalia. 2016. Why we need a token-based typology: A case study of analytic and lexical causatives in fifteen European languages. *Folia Linguistica* 50(2): 507–542.
- Levshina, Natalia. 2019. Token-based typology and word order entropy: A study based on Universal Dependencies. *Linguistic Typology* 23(3): 533–572.
- Round, Erich R. & Corbett, Greville G. 2020. Comparability and measurement in typological science: The bright future for linguistics. *Linguistic Typology* 24(3). 489-525.
- Spreafico, L., & Vietti, A. (2016). The sociophonetics of /r/ in Bozen: Modelling linguistic and social variation. *International Journal of Linguistics*, 8(5), 72–88.
- Szmrecsanyi, Benedikt. 2009. Typological parameters of intralingual variability: Grammatical analyticity versus syntheticity in varieties of English. *Language Variation and Change*, 21/3, pp. 319-353.
- Vietti, A., & Spreafico, L. (2018). Sprachkontakt in der Phonologie bilingualer Sprecher des Tirolischen. In S. Rabanus (Ed.), *Deutsch als Minderheitensprache in Italien. Theorie und Empirie kontaktinduzierten Sprachwandels* (pp. 49–78). Hildesheim: Georg Olms Verlag