Guglielmo Inglese and Silvia Ballarè

1 Analyzing language variation: Where sociolinguistics and linguistic typology meet

1 Introduction

This book offers a collection of chapters that explore the interplay between cross-linguistic and intra-linguistic perspectives to the study of language variation.[1]

Variation is an inherent property of natural languages and is pervasive across several layers: in the performance of individual speakers, within speech communities, across languages distant in time and space. As a general definition, variation can be understood as the co-existence of formally distinct linguistic forms to express the same content. Nevertheless, the nature of variation itself remains constant irrespective of its locus of manifestation. As Croft (2022: 27) observes "the patterns of variation and change found in [. . .] a particular language are in many cases simply instances of patterns of variation and change found across languages".

Language variation has constituted the core of the research agenda of at least two disciplines of linguistics, that is, variationist sociolinguistics and linguistic typology.[2] At a superficial glance, the two appear to deal with quite distinct domains. Typology "concerns itself with the study of structural differences and similarities between languages" (Velupillai 2012: 15), and is based on the study of linguistic phenomena in more or less large samples of languages representing the genetic and geographic diversity of the languages of the world.[3] By contrast, sociolinguistics focuses on "the correlation of dependent linguistic variables with independent

1 The ideas that inform this chapter have partly been published in an earlier chapter in Italian, which served as an introduction to the edited volume *Tipologia e sociolinguistica: verso un approccio integrato allo studio della variazione. Atti del Workshop SLI 2020*, published in 2021 by Officinaventuno (Milano).

2 To these, one should also add historical (comparative) linguistics, whose main focus is however variation across the diachronic dimension, as we discuss in Section 4.

3 Interestingly, typology has also been subsumed under the more general heading of *diversity linguistics* (see Haspelmath 2014-).

Guglielmo Inglese, University of Turin, e-mail: guglielmo.inglese@unito.it
Silvia Ballarè, University of Bologna, e-mail: silvia.ballare@unibo.it

https://doi.org/10.1515/9783110781168-001

social variables" (Chambers 2003: ix), and does so by typically focusing on speakers of a single language.[4]

As a matter of fact, typology and sociolinguistics approach variation with distinct premises and goals. Broadly speaking, typology is interested in crosslinguistic variation because it seeks to explore what the limits of such variation are, and, based on the empirical large-scale study of which linguistic structures are attested and which are not in the languages of the world, it aims at understanding the universal properties of human language.[5] Conversely, sociolinguistics takes a keen interest in those linguistic phenomena that have a social meaning, so as to unveil "the mechanisms which link extra linguistic phenomena (the social and cultural) with patterned linguistic heterogeneity (the internal, variable, system of language)" (Tagliamonte 2012: XIV).

What the two fields have in common is the effort to show that language variation, in all its shapes, is not random, but systematically takes place within well-defined boundaries. As a result, sociolinguistics and typology have developed over time a number of theoretical models and tools that share remarkable similarities. Nevertheless, while a number of studies attempting to address at least aspects of this relationship have appeared in recent years (see e.g. Kortmann 2004a, Trudgill 2011a), a principled discussion on how the two disciplines may interact has not yet been carried out in a programmatic way. This volume aims to fill this gap and to provide a venue for chapters discussing the bridging between sociolinguistic and typological research from various angles, with the ultimate goal of laying out the methodological and conceptual foundations of an integrated research agenda for the study of linguistic variation.

In this introductory chapter we take a closer look at the main intersections that exist between sociolinguistics and typology, with a particular focus on their theoretical and methodological apparatus. The chapter is structured as follows. Section 2 focuses on similarities among the conceptual tools developed by sociolinguistics and typology, while Section 3 deals with shared methodologies and practices. Section 4 addresses the role that diachrony plays in the study of cross- and intra-linguistic variation. Section 5 discusses a number of recent studies that showcase the mutual benefits between sociolinguistic and typological approaches, and introduces the chapters featured in this volume. Section 6 features some conclusive remarks.

4 We do not address here the thorny issue of how to distinguish languages from dialects, that is, whether two linguistic entities shall be seen as independent languages (and therefor attest to cross-linguistic variation) or instantiate varieties of the same languages (thus attesting to intra-linguistic variation). We refer to Gil (2016) for an insightful discussion.

5 The inquiry on universals carried out by typology remains fundamentally alternative to the one pursued by generative linguistics (see Daniel 2010 for discussion).

2 Analyzing variation: Theoretical models

In addressing the study of variation, sociolinguistics and typology have been confronted with a number of strikingly similar methodological issues, to address which the two fields have, quite independently, come up with comparable theoretical toolkits. In this section, we focus on three key issues, that is (i) the individuation of the linguistic trait to be taken into account when analyzing variation, that is, the notions of variables and comparative concepts (ii) the study of the aggregation of linguistic traits, be it either sociolinguistic varieties or linguistic types (iii) the relationship among linguistic traits and their systematization in terms of implicational hierarchies and universals.

2.1 How to describe variation: Sociolinguistic variables and comparative concepts

The first step in describing variation concerns the need to set the boundaries of the linguistic phenomenon under investigation, that is, the boundaries of the domain of variation. To this end, variationist sociolinguistics relies on the notion of *sociolinguistic variable*, whose definition is "is the first and also the last step in the analysis of variation" (Labov 2004: 7). A sociolinguistic variable can be defined as the set of "alternative ways of saying "the same" thing" (Labov 1972: 188). In addition to "semantic equivalence", variants as understood by variationist sociolinguistics also correlate with social factors (such as the formality of the context, the geographic origin of the speaker, etc.).

The debate around "semantic equivalence" among variants started in the '70s (see Berruto 2007 [1995]: 139–145 and Tagliamonte 2006: 70–76). Specifically, it soon became clear that while this methodology is particularly apt to describe phonetic variation, its application to other domains of linguistic analysis is far from straightforward. This is a consequence of the fact that more complex linguistic structures necessarily carry a meaning component, which poses a challenge to the individuation of semantic equivalence. This is why Sankoff (1972: 58) broadens his definition of variable and argues that "whenever there are options open to a speaker, we can infer from his or her behaviour an underlying set of probabilities". Similarly, to account for variation outside of the phonetic domain, Lavandera (1978: 181) invokes the principle of *functional comparability* of linguistic forms in individual contexts. Labov, turning to morphosyntactic variation, offers a narrower interpretation of "semantic equivalence" as pertaining to the referential level only (Labov 1972: 271; 1978; Weiner and Labov 1983).

Over the decades, the boundaries of the notion of variable have been extended so as to include variation at other levels, including morphology, syntax and the lexicon. The most problematic level undoubtedly remains that of pragmatics, and for obvious reasons, if one considers the level of abstractness of the categories used to describe pragmatic phenomena (see Pichler 2013: 6–9; on the variationist approach to pragmatic phenomena see also Tagliamonte 2005, Schneider and Barron 2008, Cameron and Schwenter 2013, and Sansò 2020: 81–104).

Linguistic typology is essentially a comparative branch of linguistics, which deals with the categorization of linguistic traits across languages (Moravcsik 2016). This means that delimiting the domain of variation is a necessary prequisite of the typological inquiry, so as to ensure the comparability and the correct categorization of linguistic traits. As Koptjevskaja-Tamm et al. (2015: 436) put it "cross-linguistic identification of studied phenomena presupposes a procedure which ensures that we compare like with like". This is a relatively straightforward task in the case of the most basic domains of lexical typology (Koptjevskaja-Tamm 2012). Things become more difficult when one tries to compare morphosyntactic entities across languages – which constitutes the core of the typological enterprise (Evans 2020; on the comparability issue see in particular two special issues of *Linguistic Typology* 20(2) [2016] and 24(3) [2020] and Alfieri et al. 2021) – since most of the notions developed in the Western grammatical thought (and on the basis of Classical languages) cannot be straightforwardly applied across languages: this concerns basic notions such as subject and object (cf. Bickel 2010; Witzlack-Makarevich 2019), subordinate clause (Cristofaro 2003) and even the notion of 'word' itself (Haspelmath 2011).

To overcome these difficulties, and achieve a meaningful comparison disregarding the structural and formal peculiarities of individual languages, scholars have argued that cross-linguistic comparison must be based on universally applicable semantic notions (cfr. Stassen 1985: 14; Croft 2003: 13–19). These have more explicitly been defined by Haspelmath (2010, 2018) as *comparative concepts* (as opposed to language-specific *descriptive categories*). Comparative concepts constitute the key tool for typologists to compare linguistic phenomena across languages and draw generalizations on the distribution of linguistic patterns, from the domain of phonetics to that of pragmatics.

In this respect, typology echoes the notion of "semantic equivalence" discussed above for sociolinguistics, in the sense that both in the case of comparative concepts and in that of sociolinguistic variables, the comparability issue has been resolved by according primacy to semantic-functional parameters in delimiting the object of investigation. Nevertheless, it must be remarked that more recent studies have also argued that typology needs *hybrid* comparative concepts, in which formal and structural parameters are taken into account (e.g., Miestamo 2005: 39–45; Stassen

2010; Croft 2016), as long as these are not based on language-specific criteria and are in their turn based on comparative concepts (e.g., the concept of root or affix).

2.2 Systematizing variation: Sociolinguistic varieties and linguistic types

Another common trait between sociolinguistics and typology is that, once the domain of variation has been identified as discussed in Section 2.1, individual traits can be organized in more general aggregates that better capture the systematicity of variation (see Grandi 2018, 2020 *inter al.*). The two disciplines achieve this by resorting to the notions of *sociolinguistic variety* and *linguistic type*, respectively.

The notion of sociolinguistic variety has been formulated in order to account for the fact that variation in languages is not random, but that instead language should be thought of as "as an object possessing orderly heterogeneity" (Weinreich et al. 1968: 100). A linguistic variety is "the realization of the linguistic system, in, or better among, classes of speakers and usages" (Berruto 2007 [1995]: 63; see further Hudson 1980, *inter al.*). In other words, varieties are constituted by a bundle of linguistic traits – at various levels – that tend to co-occur systematically, for example, in the productions of speakers that share some social features or in a certain situational context. It must also be added at this point that varieties are not discrete entities, but rather can be placed along a continuum, giving rise to an "ordered set of elements arranged in such a way that between two adjacent entities of the set (in this case, language varieties) there are no sharp boundaries, but rather a gradual, fuzzy differentiation, each variety sharing some sociolinguistically marked features with adjacent varieties" (Berruto 2010: 235).

This view on language as an entity organized in a set of varieties, which shows an evident structuralist footprint, has been variously criticized over the years by scholars who have questioned the very existence of varieties; this debate has recently been revived by studies based on large-scale annotated corpora, which have shown, by means of quantitative methods, the reality of linguistic varieties (see Hinskens and Guy 2016 and Ghyselen and De Vogelaer 2018 for an overview).

The notion of linguistic type lies at the very historical roots of typology, and is based on the assumption that the diversity of the languages of the world is amenable to a few general types (see Ramat 2010; on the notion of type see further Round and Corbett 2020). In fact, typology itself can be understood as "the study and interpretation of linguistic or language types" (Velupillai 2012: 15). A relevant feature of linguistic types, as also argued by Grandi (2014: 11–15), is that they constitute a structured and not random set of traits. This means that the usefulness of linguistic types is that they allow predictions on the co-occurrence of specific traits. For example, as

already discussed by Greenberg (1963), knowing that in one language the object is typically placed before the verb allows one to predict that, with a frequency higher than chance, the same language will also have genitives preceding their head nouns (see Song 2010). Still, the idea that types effectively provide a holistic and discrete categorization of languages, whereby a given language fully instantiate a given type, with types neatly distinct from one another, has been largely abandoned by typologists (Croft 2003: 42–45). Types are better conceived of as ideal constructs, to which languages may adhere to various extents (Coseriu 1973: 253). For example, already Sapir (1921) casted doubt on the validity of a binary distinction between isolating vs. flexional languages, arguing that morphological complexity can be computed based on the range of synthetic/analytic strategies attested within individual languages (see Greenberg 1960; Siegel et al. 2014).

Overall, it is clear that cross-linguistic diversity in most cases cannot be reduced to a few discrete types, but should be intended as a multidimensional space, featuring a number of parameters, in which languages may be variously placed (see, e.g., Arkadiev and Klamer 2018: 444 on morphological typology). A key contribution in this respect has come from the ever-increasing inclusion in typology of large-scale quantitative data from corpora (see Section 3.2.3). For example, studies carried out on treebanks have revealed how languages do not follow a single word order pattern, but various patterns coexist (with different degrees) within individual languages (Levshina 2019; Gerdes et al. 2021). This is yet another similarity between linguistic varieties and linguistic types, as they can both be ordered in continua.

2.3 Implicational relations: Universals, hierarchies and scales

Besides the notion of variety and type, sociolinguistics and typology have also come up with powerful tools to capture the relationship between traits, within and across languages.

Implicational relations, which can be expressed as either implicational universals or hierarchies, have been introduced in modern linguistics by Joseph Greenberg (1963). Typological universals constitute perhaps the major achievement of the study of cross-linguistic variation, as it is thanks to language universal that one can assess the limit of linguistic diversity and predict which languages are possible and which not (Croft 2003: Cap. 3; Moravcsik 2010). Universals can be distinguished into absolute and implicational ones, as in (1a-b) (Croft 2003: 52–53):

(1) a. All languages have oral vowels.
    b. If a language has nasal vowels, it will also have oral vowels, but not viceversa.

The universal in (1a) is absolute because it describes a property equally shared by all (spoken) languages of the world. Universals of this type are rather the exception. Most universals are implicational, as (1b), as it establishes an implicational relationship between the occurrence of distinct linguistic traits. Implicational universals are theoretically more interesting than absolute ones, as only the former allow typologists to delimit the domain of cross-linguistic variation, so as to exclude impossible language types. Coming back to (1b), this universal also rules out the possibility that languages exist in which only nasal vowels, but not oral ones, are found.

Implicational hierarchies merely constitute a mono-dimensional chain of implicational universals, whereby the implicans of one universal is the implicatum of the following one (Croft 2003: 122). Hierarchies are particularly suitable to capture the systematic relation between a high number of traits, and have been used in various domain from the lexicon to morphosyntax (Croft 2003: Cap. 5; Corbett 2010). Famous hierarchies are the hierarchy of accessibility relativization (Keenan and Comrie 1977, 1979) and the animacy or referential hierarchy (e.g., Corbett 2000; Cristofaro 2013). In the lexical domain, mostly pertaining to the verbal lexicon, important hierarchies are the hierarchy of perception modalities (Viberg 1984), as well as transitivity (Tsunoda 1985; Malchukov 2005) and spontaneity scales (Haspelmath 2016).

Within the functional-typological approach, regular correlations among linguistic traits are not treated as random, but have traditionally been explained by resorting to specific underling functional and cognitive principles, including economy, iconicity, markedness, memory, ease of processing etc. (Moravcsik 2010; Sansò 2018). As a matter of fact, more recent research has questioned the validity of such explanations, which are strictly functional and synchronic. First – and this is an input coming from usage-based approaches to the study of language – it has been argued that frequency of occurrence, correlating with predictability and with coding efficiency, plays a key role in explaining some of the universals that have been proposed, in particular those that have to do with marking asymmetries, i.e. those cases in which a value of a given grammatical category (e.g., plural) receives extra marking as compared to another, more frequent one (e.g., singular) (see Haspelmath 2006, 2021). The second kind of criticism instead comes from adopting a diachronic perspective to cross-linguistic regularities (Section 4.1).

Comparable tools have been devised in sociolinguistics. For example, implicational scales have been traditionally regarded as an alternative model to variable rules, understood as the expression of the behavior of a variable (see Rickford 2002 inter al.). In fact, implicational scales and variable rules might be (and as a matter of fact have been) used in combination, partly because they represent distinct objects. Scales account for the relationships that may exist among variables

(Berruto 2007 [1995]: 156), by establishing implicational co-occurrence patterns among traits in a given sociolinguistic variety or in the speech of speakers with specific social connotations. In practice, a scale is made up as a cross-table in which linguistic traits are ordered on one side and (groups of) speakers or varieties on the other side. Cells of this matrix can be filled by either binary values (e.g., +/-) which account for the (non-)occurrence of a given trait in a given production, or by intervals of frequencies relative to the occurrence of a specific trait (as an example, see Cerruti 2009: 235–254).

3 Exploring variation: Empirical tools

A shared feature between typology and sociolinguistics is that these are two heavily empirically based branches of linguistics. In fact, they both proceed from the assumption that variation can only be captured by analyzing the distribution of linguistic patterns in real data. It should come as no surprise, at this point that, besides having developed similar theoretical models for the study of variation (Section 2), the two disciplines have also come up with a number of remarkably similar research tools. In this section, we highlight commonalities that can be detected among these, so as to highlight how they have been independently created as a response to similar theoretical and methodological pressures. Even though the tools are essentially the same, it is interesting to observe how they have become part of the toolkit of sociolinguists and typologists at different times, reflecting the different priorities of the two fields along their history.

3.1 The empirical turn: Corpora and questionnaires

In order to obtain comparable data on specific linguistic features from speakers of a given language, while at the same time collecting the relevant extralinguistic metadata on informants, sociolinguistics has long relied on questionnaires. Questionnaires are usually made up of two components, one concerning the linguistic data, which is typically elicited by administering specific stimuli to informants, the other one adding various type of metadata.

Questionnaires are widely used in sociolinguistics and dialectology to gather various types of data, including acceptability judgements, information on repertories, etc. (see Meyerhoff et al. 2015: 71–73). This linguistic information comes with metadata on informants (age, geographic origin, degree of education, etc.), so as to allow the researcher to evaluate possible correlations between the distribution

of linguistic features and extra-linguistic factors. Questionnaires are particularly suitable when one wishes to explore the sociolinguistic characterization of a given trait, as they can be administered to a sample of informants that has been selected specifically for the purpose of the investigation. Nevertheless, questionnaires are not free of shortcomings, first and foremost the fact that their responses generally also represent the informants' evaluation on the stimulus, which might be distinct from their actual usage of the language (see Iannàccaro 2000 and more generally the methodological remarks in Sanga 1991).

Typological inquiries, in the most traditional sense, usually proceed with gathering a representative sample of the languages of the world and then sifting through grammatical descriptions of these languages in order to obtain data on the phenomenon under analysis. This methodology has been defined *type-based* by Levshina (2019), in the sense that languages are typically classified as having (a specific value of) vs. lacking a certain feature. However, as grammatical descriptions greatly differ in their granularity and thoroughness, in some cases, especially when one wishes to investigate less documented phenomena with a high degree of detail, questionnaires might supply a crucial additional source of data. One of the most renowned typological questionnaires has been devised by Östen Dahl to investigate the cross-linguistic realization of TAM systems (Dahl 1985).

In general, the main advantage of questionnaires is that they allow the researcher to gather highly detailed and comparable data for the linguistic trait under investigation. By directly eliciting data through standardized stimuli, one minimizes the risk of overlooking traits that exist in a language but are not described in reference grammars and/or are less documented in corpora. To further minimize the limits of linguistic stimuli, including translation biases and speakers' attitudes, a number of questionnaires featuring a set of non-linguistic (typically visual) stimuli has been elaborated. These cover both morphosyntactic and lexical domains (Koptjevskaja-Tamm et al. 2015: 441–442 for an overview).[6] Non-linguistic visual stimuli have also long been used in sociolinguistic research and still constitute a valuable tool (see for example de Benito Moreno 2022 for a recent application).

Some of the limits inherent to questionnaires can be overcome by looking for linguistic data in corpora. Corpora constitute a long-standing resource in sociolinguistics, as they enable access to a considerable amount of linguistic data naturally produced by speakers, thereby avoiding possible biases introduced by the controlled setting in which data from questionnaires is elicited. Clearly, in order to be useful for sociolinguists, corpora must also be supplied with extra-linguistic

6 A useful collection of typological questionnaires can be found at https://www.eva.mpg.de/lingua/tools-at-lingboard/questionnaires.php.

metadata. Such resources are extremely valuable in that they allow fine-grained quantitative analyses, which have been a hallmark of sociolinguistics since its earliest phase. Over the years, several sociolinguists have devoted time and effort to the creation of corpora, as well as to the discussion of the more suitable methodologies for corpus building (e.g., Tagliamonte 2006) and of the relevance of corpora for sociolinguistic research (Poplack 2021).

By contrast, even though quantitative approaches are not new to typology (see already Greenberg 1960), it is only recently that the use of corpora has opened the way for a new quantitative oriented typology based on a rich empirical basis. Levshina (2021) offers an excellent overview of the type of corpora used in typology, the typological research questions that can be approached with quantitative methods, and the advantages and limits of corpus-based, or better *token-based*, typology.

It must be remarked that the use of corpora for typology is practically and methodologically more challenging than for sociolinguistics. The main issue is that cross-linguistic corpus-based studies need a reasonable number of corpora of different languages, which additionally must be sufficiently similar to allow a meaningful comparison. Two main solutions to this issue have been proposed, that is, either parallel corpora consisting of the translation of the same text in several languages (Cysouw and Wälchi 2007), or, more recently, corpora that share a common annotation schema, such as the treebanks featured in the Universal Dependency project (Zeman et al. 2020). Another relevant ongoing project is the DoReCO, which aims at publishing annotated spoken corpora of 150 languages, with a focus on endangered and/or underdescribed languages (Paschen et al. 2020). The latter project in particular has constituted a major step forward in advocating for the need of typological studies not based on grammatical description, but rather grounded on corpora of spoken language (Haig et al. 2021).

Notably, corpora in typology do not merely constitute a source of cross-linguistic data, but the shift from type- to token-based typology also bears remarkable theoretical consequence. Specifically, while in traditional type-based typology languages are typically treated as either featuring/lacking a certain trait, token-based typology is more suitable to describe much more accurately the gradual nature of specific linguistic phenomena. This entails, among other thing, replacing the categorial approach to types with a description of variation into continua (see Section 2.1). Token-based typology has given important results in the study of word order (Levshina 2019, Gerdes et al. 2012, Choi et al. 2021), in phonetic studies (e.g., Paschen et al. 2022), and in the study of morphological complexity (e.g., Brigada Villa and Inglese 2021; Stave et al. 2021). On a more general level, as Levshina (2020: 8) remarks, once typology turns to usage- (and corpus-)based models, which entail a keen interest in the gradual nature of language and the interaction between

the realization of individual traits and their context of use, typologists effectively conduct variationist studies, which can then benefit from the methods and tools developed within the sociolinguistics tradition.

3.2 Visualizing data: Atlases

Geography constitutes a key aspect of linguistic variation. Linguistic atlases have been devised as means to observe diversity of languages in space. In dialectology, atlases have been devoted mostly to the description of phonetic, morphological and lexical variation. In particular, they have been thriving in European countries with a marked multilingual landscape, such as Germany, Switzerland, Italy, and Russia (Lameli 2010: 569). Modern linguistic atlases are made up by individual charts on which a specific phenomenon is mapped according to its realization in specific points within the area under consideration. Building on the tradition of dialectal atlases, the last decades have also witnessed a surge in typological and sociolinguistically informed atlases. Sociolinguistic atlases account for intralinguistic variation and typically display how the variants of a given variable are realized in space. As an example, consider the SAND – *Syntactische Atlas van de Nederlandse Dialecten* (Barbiers et al. 2005) or the ANAE – *Atlas of North American English* (Labov et al. 2006). The interest in the geographic and areal dimension constitutes in some respects a more recent acquisition in typology. The main typological atlas is the WALS – *Word Atlas of Language Structure* (Haspelmath et al. 2005, online version Dryer and Haspelmath 2013). WALS has served as the model for the creation of a number of analogous resources. Notable examples are the APiCS – *Atlas of Pidgin and Creole Language Structures* (Michaelis et al. 2013) and PHOIBLE, which gathers data on the phonological inventory of almost 2100 languages (Moran and McCloy 2019). Recently, a new typological atlas, *Grambank*, has been launched, which features information concerning 195 morphosyntactic parameters for almost 2500 languages (Skirgård et al. 2023). There exist also atlases restricted to particular linguistic areas, as is the case of SAILS – *South American Indigenous Language Structures* (Muysken et al. 2016), an atlas featuring several grammatical traits of South American languages.

As we have discussed in this section, both typology and sociolinguistics have developed practical tools to display and analyze variation in space. A peculiar resource in this respect, which shows how tools developed by one discipline can easily and fruitfully be transferred to the other, is the eWAVE – *electronic World Atlas of Varieties of Varieties of English* (Kortmann and Lunkenheimer 2013, digital version Kortmann et al. 2020). eWAVE shows a structure entirely analogous to that of WALS, but instead collects data on morphosyntactic structures in 77 geograph-

ical varieties of English. There is however one key difference with respect WALS: while WALS largely disregards intralinguistic variation, as each language is represented by only one variety, in the case of eWAVE each chart represent a given trait and for each variety the relative frequency of the trait is given (with values ranging from pervasive to absence of feature).

4 Variation and change: The role of diachrony

In the structuralist tradition initiated by de Saussure (2016 [1916]: 174), synchrony and diachrony are two complementary perspectives to the study of language. According to this view, variation pertains to the synchronic level only, while diachrony is rather concerned with change. Such a dichotomic approach has been largely superseded, and scholars nowadays agree that the two dimensions are rather "two sides of the same coin" (Seiler 2018: 82; see also Giacalone Ramat et al. 2012). The interplay between the two dimensions is twofold (e.g. Croft 2000; Luraghi 2010; Seiler 2018). On the one hand, variation is the result of change, in that change may result in the creation of new forms for a given function, thereby bringing about a new variant within a system; on the other hand, synchronic variation is also the pool from which language change, understood as a process of selection and conventionalization of a new variant against an older one, takes places. A key contribution in offering a more dynamic understanding of the interface between synchrony and diachrony comes from grammaticalization studies (see Hopper and Traugott 2003; Lehmann 2015 [1995]). Grammaticalization processes, at least in their more traditional formulation, typically involved lexical elements that progressively acquire grammatical meaning, losing their morphosyntactic autonomy in the process. Crucially, these processes take place on-line in the concrete production of speakers, thanks to the reinterpretation of specific source constructions into new target constructions (see below).

In the reminder of this section, we briefly sketch the ways in which sociolinguistics and typology have integrated the diachronic dimension within their synchronic account of intra- and cross-linguistic variation.

4.1 Diachronic typology

In spite of its apparently predominant synchronic orientation, typology has taken the diachronic dimension into account since its earliest days. Let us consider the case of morphological typology: already Humboldt, in formulating his clas-

sification of languages into isolating, agglutinative, inflected and incorporating, not only proposed that these constitute four distinct synchronic language types, but also established a diachronic relation among these, in the sense that each stage follows the other one along an ideal evolutionary process. The need for a diachronic dimension has been more recently advocated by Greenberg himself (e.g., Greenberg 1995). Nevertheless, it is only over the last couple of decades that the diachronic dimension has been more systematically integrated in the typological framework, thanks to a fruitful interaction with historical linguistics and grammaticalization studies (e.g. Narrog 2017; Heine and Narrog 2018; Bisang et al. 2020). In particular, grammaticalization studies have shown how grammatical structures recurrently arise from a well-defined set of historical sources and following a restricted number of processes across languages (es. Narrog 2017, Kuteva et al. 2019).). In this respect, diachrony offers a key to understand variation, by answering to the question "why languages have come to be the way they are" (Croft 2022: 27).

The diachronic perspective is also crucial to semasiological approaches, in the sense that grammaticalization studies offer an important tool to explain the co-existence of different functions performed by the same grammatical form, by establishing a diachronic connection among these and by providing a straightforward explanation for why certain polyfunctionality patterns recur across languages and others do not.

Studies in diachronic typology have led to an overall reassessment of the role of the synchronic functional principles that underlie language universals and implicational hierarchies. This approach to typology, which can be defined as *source-oriented*, is grounded in the assumption that cross-linguistic regularities in the distribution of linguistic forms cannot be solely based on synchronic functional motivations but can be better understood by taking into account constraints on the types of sources and diachronic processes that lead to the emergence of such forms in the first place (cf. Blevins 2004; Moravcsik 2010; Cristofaro 2019, 2021; Mithun 2018; see Haspelmath 2019 for a critical overview). As a result of this line of research, diachronic explanations have become an increasingly essential component in explaining the nature and constraints of language diversity.

4.2 Sociolinguistic varieties and patterns of change

Similarly to typology, also within sociolinguistics the study of variation has been increasingly related to historical processes. Specifically, it has been observed how variants of individual variables analyzed across different varieties of the same language might instantiate distinct diachronic stages within a single grammaticaliza-

tion process (see Nevalainen and Palander-Collin 2011; Poplack 2011; Cerruti 2021). Since synchronic variation is the place where every process of change begins, one could say that the different patterns of variation may be put in relation with mutation patterns (see Lehmann 1985). In particular, sub-standard varieties show linguistic features that can be considered to be more grammaticalized when compared to the ones attested in standard (or *supra*-standard) varieties (see Cerruti 2021 for a discussion).

From our perspective, it is interesting to link the notion of *(socio)linguistic variable* to the one of *layering*, which Hopper (1991: 22) describes as follows: "withing a broad functional domain, new layers are continually emerging. As this happens, the older layers are not necessarily discarded, but may remain to coexist with and interact with the newer layers". Clearly, from a sociolinguistic point of view, these so-called "layers" can be considered variants of the same variable, and the *layering* portrays the co-existence in a certain moment in time of different ways of conveying the same functional meaning. Several variationist studies have been carried out in this perspective; it is worth mentioning, among others, Tagliamonte (2000), Poplack (2011), and Torres Cacoullos (2011).

Furthermore, to a diachronic dimension is linked the "inverse" semasiological approach to sociolinguistic variables, whereby the focus is not on how a given meaning is realized by different forms (and what governs such variation) but rather on how different functions of the same form may correlate with extra-linguistic parameters (see Cerruti 2011). In this case, the main issue is how to securely establish which and how many distinct functions should be associated to a given form, which is usually left to the individual researcher's intuition. This difficulty was already clear to Weiner and Labov (1983: 31), who argued that "theoretically, it should be possible to draw equal profit from cases where a single form is used with several meanings. But the possibility of accurate measurement is less immediate with semantic variation". Nevertheless, such an approach offers the significant advantage that it does not require the researcher to follow the principle of "semantic equivalence" and to individuate in a straightforward way all contexts to be considered as instantiating variation (that is, all contexts of occurrence of a given form). This perspective, which stems from studies in historical linguistics in which the diachronic connections among various meanings/functions of a given form are studied, has also been successfully applied to sociolinguistic studies.

5 Towards a unitary framework for the study of variation

In spite of the fact that structural variation has been the main object of interest of both typology and sociolinguistics, the kinds of explanations for variation that the two disciplines have formulated are rather different. Patterns of cross-linguistic distributions are usually explained in terms of functional properties (economy, iconicity, processing, etc.). Conversely, language internal variation is often explained by variationist sociolinguistics by also appealing to extra-linguistic socio-demographic factors (speakers' age and education, register, etc.). Likewise, in a historical perspective, typology has been concerned with the general mechanisms of language change that bring about specific cross-linguistic patterns of distribution (Cristofaro 2019), while sociolinguistics has put emphasis on the extra-linguistic factors behind the progressive diffusion of linguistic innovations within communities (Labov 2001).

Nevertheless, typological and sociolinguistic explanations of variation are in principle not mutually exclusive and should be integrated into a general explanatory framework of linguistic variation. As a matter of fact, building on the theoretical and methodological similarities between the two disciplines, as outlined in Sections 2 to 4, scholars have started to highlight the possible mutual benefits of a more systematic and programmatic interaction between the two (e.g., Trudgill 2011a; see also the papers in Kortmann 2004a and Ballarè and Inglese 2021). In what follows, we highlight aspects of this relationship, focusing on the way in which typology has influenced sociolinguistics and vice versa. In addition, we also discuss how the chapters included in this volume variously address these lines of research from different angles, bringing their own contribution to the overall quest of integrating language-specific and cross-linguistic research on linguistic variation on both a methodological and a conceptual level.

Sociolinguistics (and, more generally, the study of intralinguistic variation) has benefited from the influence of typology first and foremost in that the latter has offered useful *tertia comparationis* which, by virtue of their comparative cross-lingusitic nature, are independent from individual languages (or families). As already remarked by Bossong (1991: 143) "broad typological comparison of data from many genetically and structurally different languages is necessary in order to be able to describe phenomena of single languages as what they really are".

One must mention in this respect the volume edited by Kortmann (2004a), whose programmatic objective was to foster the dialogue among scholars working on inter- and intra-linguistic variation, as explicitly stated in the introduction: "The purpose of this invitation is to bring together for the first time two research traditions in the study of language variation (and change) which so far have largely

worked independently of each other, to make them enrich and provide new vistas for each other" (Kortmann 2004b: 1).

The benefit of a typological angle to sociolinguistic research is well exemplified in the chapters by Lorenzo Ferrarotti and Konstantinos Sampanis. In his chapter, *Isolation, complexification, and development of unusual features: a case study from some Gallo-Italian dialects of Northern Italy*, Lorenzo Ferrarotti sets out from the observation that these dialects show a number of traits, both at the phonetic and at the morphological level, which not only sets them apart from the urban variety spoken in Novara, the main city in the area, and from Lombard (and Milanese) dialects, but are also unique among Romance languages more generally. These features include "hardening" of word-final [ŋ] > [k], preservation of final vowels (which also entails preservation of richer inflectional paradigms in the nominal domain), and generalized enclisis of weak object pronouns in all syntactic contexts. Viewed in their Romance context, these traits strike as rather unusual, and could be dismissed as oddities. Instead, Ferrarotti argues that a better understanding of these traits can be achieved once these are considered in a more general context. In particular, following Trudgill (2011a), he argues that these traits represent more *complex* strategies with respect to neighboring dialects and Romance languages more generally, and that these processes of complexification can only take place due to the specific social and geographic settings in which these dialects are spoken.

In a similar vein, in the chapter *On typological shift in Inner Anatolian Greek*, Kostantinos Sampanis takes a look at variation in dialects of Greek spoken in Anatolia (Inner Anatolian Greek) through a typological lens. These dialects show a number of morphosyntactic peculiarities with respect to Modern Greek and to dialects spoken in Greece, which have traditionally been explained as the result of intensive contact with Turkish (contact among these varieties is such that some scholars have also analyzed Inner Anatolian Greek as a mixed Greek and Turkish language, an analysis which, however, the author of the chapter refutes). As Sampanis points out, these peculiarities may be better understood if placed in a typological perspective, as these include the agglutinative nominal morphology, with separate case and number endings, and a tendency towards left-branching structures at the phrasal and sentence level. In other words, in these domains, one may consider Inner Anatolian Greek as displaying a distinct typological profile from that of other Greek varieties. Crucially, the typological divergence between Inner Anatolian and Modern Greek is to a large extent triggered by contact with Turkish, but, as the author observes, different traits have developed under different social settings. In particular, one may imagine that the reshaping of nominal paradigms in a more agglutinative-like direction is the result of a long-standing contact scenario within a bilingual society, whereas left-branching structures may have been more readily

introduced, even within the span of a generation, due to the pressure of dominant Turkish upon local Greek varieties.

The notion of mixed language also plays a key role in the chapter by Anja Hasse and Guido Seiler. In this chapter, Hasse and Seiler discuss this notion in light of data from Amish Shwitzer. Amish Shwitzer is a language spoken by a group of Old Order Amish in Adams County, Indiana (US), which, as the author argue, has emerged as a result of language contact between Bernese Swiss German and Pennsylvania Dutch. By framing their analysis within a more general typology of mixed languages and the possible socio-cultural conditions that may favor their emergence, the authors argue that Amish Shwitzer is to be considered an intertwined language. In particular, the authors single out two socio-historical factors as crucial in determining the rise (and success) of this language: the role of mixed marriages and the role of this language as a marker of identity, as it constitutes a means whereby speakers differentiate themselves not only from American majority society but also from Non-Swiss Amish communities and from Non-Amish Swiss communities. Interestingly, the two factors are linked to two distinct stages in the history of Amish Shwitzer: while mixed marriage (and the consequent incomplete adult L2 acquisition of Bernese Swiss German by Pennsylvania Dutch-speaking Amish) are the main reason why the new mixed lect emerged, its success within the community is mostly linked to its function as a marker of identity. In doing so, this chapter makes a compelling case that distinct socio-historical factors might account for different stages of language change, namely actuation and diffusion.

The main lesson that sociolinguistics has taught typology concerns the need to take into account intra-linguistic variation and its underlying extra-linguistic factors, along two lines. The first line of research focuses on the interaction between language-internal and cross-linguistic variation. On a methodological level, what typologists may learn from sociolinguistics is the opportunity to consider, for languages with a complex sociolinguistic situation, (also) non-standard varieties, often neglected in the practice of building typological samples. This point is well illustrated in the chapter by Sampanis, where remarkable typological differences among dialects of Greek are discussed. In addition, the comparison between non-standard, typically oral, varieties may reveal the existence of common features even across typologically distant languages (cf. e.g. Auer 1990; Auer and Maschler 2013; Ballarè and Inglese 2022) and show patterns of variation that cannot be observed taking into consideration standard varieties only (Bossong 1985, 1991 and more recently, Filppula et al. 2009; Murelli and Kortmann 2011; Kortmann and Lunkenheimer 2013; Seiler 2019; Grandi 2021).

The second line of research, exemplified by Trudgill (2011), aims at integrating non-linguistic factors in understanding and explaining linguistic diversity, much in the same vein as sociolinguistics looks at extra-linguistic factors to explain

intra-linguistic variation. While non-linguistic factors investigated by sociolinguistics generally pertain to individual speakers within a speech community, the focus of typological studies evidently lies in how the distribution of language structures is shaped by factors that instead pertain to population structure and more generally the ecology of speech communities as a whole (see Di Garbo et al. 2021 and Becker et al. this volume for an overview).

Among the factors that have been proposed, one might mention demographic factors such as the structure of social networks (open vs. close-knit) within communities (Wray and Grace 2007; Trudgill 2011b), marriage practices in terms of endogamy vs. exogamy (Pakendorf et al. 2021), population size and number of linguistic neighbors (Lupyan and Dale 2010), as well proportion of L2 speakers in a community (Bentz and Winter 2013), but also non-social factors, including environmental and climatic considerations (Everett et al. 2015). Factors of various kinds may also be combined: for example, Urban (2020) discusses the way in which environmental constraints affect social and demographic structures in mountainous areas, and the consequences for the structure of languages spoken in these areas. As also remarked by Grandi (2021), such considerations also have non-negligible methodological consequences, for example on the guidelines for building typological samples. More generally, over time, also thanks to the development of typological atlases (Section 3.2), it has become increasingly clear how geographic and areal considerations must be accounted for in explaining the distribution of linguistic traits (see Nichols 1992; for an overview see Bickel 2017). As Becker et al. (this volume) remark, one of the reasons why socio-historical factors have not yet systematically been taken into account in typological studies is a practical one. As already remarked by Trudgill (2011b), quantitative studies aiming at establishing correlations between cross-linguistic distributions of linguistic traits and socio-cultural factors have predominantly relied on population size as a predictor. This is essentially a practical choice, as population size is relatively easy demographic data to access, but there is in principle no reason to assume that this factor indeed displays a higher explanatory power. Building on these premises, in the chapter *Socio-linguistic effects on conditional constructions: A quantitative typological study*, Laura Becker, Matías Guzmán Naranjo, and Samira Ochs try to assess, by means of fine-grained quantitative analyses, the impact of various social factors on the inventory of conditional markers in a sample of the languages of the world. Their finding is that population size remains the better predictor, in the sense that languages with a higher number of speakers will also likely have more than conditional markers. Other factors, such as use of the language in education and existence of a written tradition do not appear to be as informative, partly due the fact that they often correlate with population size, but also partly because this is data that is more difficult to come by in a detailed enough way to measure its relevance with quantitative

techniques. Therefore, the chapter makes a compelling point that language description and documentation constitutes the crucial missing link between typology and sociolinguistics: the more grammar writers include socio-cultural information in their description, the more typologists will be able to take these into account when investigating language diversity.

Conditionals are also the topic of the chapter authored by Alessandra Barotto, Simone Mattiola and Caterina Mauri, titled *Counterfactual conditionals: linguistic variation in Italian and beyond*. The aim of this chapter is two-fold. On the one hand, the authors investigate the strategies to express counterfactual conditional in different varieties of Italian. To this aim, adopting a sociolinguistic perspective, they conduct a corpus analysis of counterfactuals in a spoken corpus of Italian and find that the symmetric strategy involving the imperfective past tense (*imperfetto*) in both the protasis and the apodosis is more likely to occur in non-formal situations and is favored by less educated speakers. This sociolinguistic investigation, which highlights patterns of intralinguistic variation, is then connected to a cross-linguistic investigation of counterfactuals in a sample of 203 languages. Cross-linguistic data shows the widespread use of symmetrical constructions, especially those involving past habitual forms. This means that the strategy most frequently found in non-standard Italian is in fact the manifestation of a recurrent cross-linguistic type. This raises the issue whether such symmetric counterfactuals are in fact typical of non-standard spoken varieties cross-linguistically. In this respect, the authors echo Becker et al. (this volume) and remark that, unless cross-linguistic data also comes with detailed metadata on the sociolinguistic context of the variety recorded, bridging studies of inter- and intra-linguistic variation remains a difficult task.

It must be remarked that, in some cases, the sociohistorical factors that have been so far explored remain to some extent *distal* to the mechanisms underpinning language variation (and change). An attempt to bridge this gap takes center stage in the chapter by Bert Cornillie and Malte Rosemeyer, *Syntactic elaboration in the domain of periphrasticity: evidence from Spanish.* The authors focus on diamesic variation, and frame their analysis from the perspective of the contrast between communicative distance vs. immediacy, as developed in Romance linguistics. Based on quantitative analyses of extensive historical documentation from Spanish, the authors compellingly show that periphrastic verb constructions, especially those of Latinate origin, are more likely to occur in texts that belong to the real of communicative distance. In particular, they establish a correlation between a higher percentage of clause linking strategies, which are typical of written texts in a communicative distance setting, and occurrences of verbal periphrases. In doing so, they offer clear evidence of how the distribution of a particular typological trait, that is, analytic verbal strategies, may be the historical result of a precise socio-cultural setting. This contrasts with the rise of analytic strategies in other contexts, such as

in pidgin and creole languages, which is rather motivated by the need for clarity in contexts of communicative immediacy. These findings offer further evidence that even the same typological trait may arise for different reasons in different languages, thus corroborating the findings of source-oriented diachronic typology.

The lines of research mentioned in this section, as well as the chapters presented in this volume, show that a lot has already been achieved in the direction of integrating sociolinguistic and typological approaches to variation, but that much is yet to be done. Understanding the link between sociohistorical and typological variation ultimately requires a twofold effort, which future research will have to take into account: on the one hand, conducting in-depth studies of language evolution and change and of the role of contact and language ecology in the dynamics of language; on the other hand, using evidence from these studies to develop new methods and variables for large-scale comparisons of language structures, social structures and interactions thereof.

Acknowledgements: This book collects some of the papers presented at the workshop "Integrating sociolinguistic and typological perspectives on language variation" which took place (online) during the 54th annual meeting of the *Societas Linguistica Europaea* (31 August – 3 September 2021). We would like to express our gratitude to the participants to the workshop, and especially to our co-convenors, Francesca Di Garbo and Eri Kashima, whose ideas also largely inspired this introductory chapter. We also thank Chiara Gianollo and Daniel Van Olmen for having encouraged us to submit this volume to the TiLSM series, as well as Barbara Karlson for her kind assistance during the editorial process. Finally, we thank all the authors who have contributed to the current volume for believing in our editorial project and all the anonymous reviewers for their invaluable help. Silvia Ballarè acknowledges that this contribution is the result of the research carried out within the PRIN 2017 UniverS-Ita project Written Italian of university students: sociolinguistic framework, typological trends, didactic implications (ERC SH4, Prot. 2017 LAP429).

Reference

Alfieri, Luca, Giorgio Francesco Arcodia & Paolo Ramat (eds.). 2021. *Linguistic Categories, Language Description and Linguistic Typology*. Amsterdam/Philadelphia: John Benjamins.

Arkadiev, Peter & Marian Klamer. 2018. Morphological theory and typology. In Jenny Audring & Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*, 436–454. Oxford: Oxford University Press.

Auer, Peter. 1990. Einige umgangssprachliche Phänomene der türkischen Syntax und Möglichkeiten ihrer Erklärung aus 'natürlichen' Prinzipien. In Norbert Boretzky, Werner Enninger & Thomas Stolz (eds.), *Spielarten der Natürlichkeit–Spielarten der Ökonomie. Beiträge zum 5, Essener Kolloquium, vol. 2*, 271–298. Bochum: Brockmeyer.

Auer, Peter & Yael Maschler. 2013. Discourse or grammar? VS patterns in spoken Hebrew and spoken German narratives. *Language Science* 37. 147–181.

Ballarè, Silvia & Guglielmo Inglese (eds.). 2021 *Tipologia e sociolinguistica: verso un approccio integrato allo studio della variazione. Atti del Workshop SLI 2020*. Milano: Officinaventuno.

Ballarè, Silvia & Guglielmo Inglese. 2022. The development of locative relative markers. From typology to sociolinguistics (and back). *Studies in Language* 46(1). 220–257. https://doi.org/10.1075/sl.20013.bal

Barbiers, Sjef, Hans Bennis, Gunther de Vogelaer, Magda Devos & Margreet van der Ham. 2005. *Syntactische Atlas van de Nederlandse Dialected, Deel I*. Amsterdam: Amsterdam University Press.

Bentz, Christian & Bodo Winter. 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change* 3. 1–27.

Berruto, Gaetano. 2007 [1995]. *Fondamenti di sociolinguistica*. Rome/Bari: Laterza.

Berruto, Gaetano. 2009. Περί συντάξεως. Sintassi e variazione. In Angela Ferrari (ed.), *Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione, giustapposizione, Atti del X Congresso della Società Internazionale di Linguistica e Filologia italiana* (Basilea, 30 giugno-3 luglio 2008), 21–58. Firenze: Cesati.

Berruto, Gaetano. 2010. Identifying dimensions of linguistic variation in a language space. In Peter Auer & Jürgen Erich Schmidt (eds.), *Language and Space, Part I: Theories and Methods. An International Handbook of Linguistic Variation*, 226–241. Berlin/New York: De Gruyter Mouton.

Bickel, Balthasar. 2010. Grammatical relations typology. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 399–444. Oxford: Oxford University Press.

Bickel, Balthasar. 2017. Areas and universals. In Raymond Hickey (ed.), *The Cambridge Handbook of Areal Linguistics*, 40–55. Cambridge: Cambridge University Press.

Bisang, Walter, Andrej Malchukov & the Mainz Grammaticalization Project team (Iris Rieder, Linlin Sun, Marvin Martiny, Svenja Luell). 2020. *Position paper: Universal and areal patterns in grammaticalization*. In Walter Bisang & Andrej Malchukov (eds.), *Grammaticalization Scenarios: Cross-linguistic Variation and Universal Tendencies*, 1–88. Berlin/New York: De Gruyter Mouton.

Blevins, Juliette. 2004. *Evolutionary phonology: the emergence of sound patterns*. Cambridge: Cambridge University Press.

Bossong, Georg. 1985. *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Narr.

Bossong, Georg. 1991. Differential Object Marking in Romance and Beyond. In Douglas Kibbee & Dieter Wanner (eds.), *New Analyses in Romance Linguistics*, 143–170. Amsterdam/ Philadelphia: John Benjamins.

Brigada Villa, Luca & Guglielmo Inglese. 2021. Inferring Morphological Complexity from Syntactic Dependency Networks: A Test. *Proceedings of the Third Workshop on Computational Typology and Multilingual NLP, 10 June 2021, 10–22.* Association for Computational Linguistics. Available at <https://aclanthology.org/2021.sigtyp-1.pdf>

Cameron, Richard & Scott Schwenter. 2013. Pragmatics and variationist sociolinguistics. Robert Bayley, Richard Cameron & Celi Lucas (eds.), *The Oxford Handbook of Sociolinguistics*, 464–483. Oxford: Oxford University Press.

Cerruti, Massimo. 2009. *Strutture dell'italiano regionale. Morfosintassi di una varietà diatopica in prospettiva sociolinguistica*. Frankfurt am Main: Lang.

Cerruti, Massimo. 2011. Il concetto di variabile sociolinguistica a livello del lessico. *Studi italiani di linguistica teorica e applicata* 40(2). 211–231.
Cerruti, Massimo. 2021. Variazione sociolinguistica e processi di grammaticalizzazione. In Silvia Ballarè & Guglielmo Inglese (eds.), *Tipologia e sociolinguistica: verso un approccio integrato allo studio della variazione. Atti del Workshop SLI 2020*, 53–80. Milano: Officina 21.
Chambers, Jack K. 2003. *Sociolinguistic theory: linguistic variation and its social significance*. Oxford: Blackwell Publishers.
Choi, Hee-Soo, Bruno Guillaume & Karën Fort. 2021. Corpus-based language universals analysis using universal dependencies. In Radek Čech & Xinying Chen (eds.), *Proceedings of the Second Workshop on Quantitative Syntax (Quasy, SyntaxFest 2021)*, 33–44. Sofia: Association for Computational Linguistics.
Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.
Corbett, Greville G. 2010. Implicational hierarchies. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 43–68. Oxford: Oxford University Press
Coseriu, Eugenio. 1973. Sulla tipologia linguistica di Wilhelm von Humboldt: contributo alla critica della tradizione linguistica. *Lingua e stile* 8. 235–266.
Cristofaro, Sonia & Fernando Zúñiga. 2018. Synchronic vs. diachronic approaches to typological hierarchies. In Sonia Cristofaro & Fernando Zúñiga (eds.), *Typological hierarchies in synchrony and diachrony*, 4–27. Amsterdam/Philadelphia: John Benjamins.
Cristofaro, Sonia. 2003. *Subordination*. Oxford: Oxford University Press.
Cristofaro, Sonia. 2013. The referential hierarchy: Reviewing the evidence in diachronic perspective. In Dik Bakker & Martin Haspelmath (eds.), *Languages Across Boundaries: Studies in the Memory of Anna Siewierska*, 69–93. Berlin/New York: De Gruyter Mouton.
Cristofaro, Sonia. 2019. Taking diachronic evidence seriously: Result-oriented vs. source-oriented explanations of typological universals. In Karsten Schmidtke-Bode, Natalia Levshina, Susanne Maria Michaelis & Ilja A. Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, 25–46. Berlin: Language Science Press.
Cristofaro, Sonia. 2021. Towards a source-oriented approach to typological universals. In Peter Arkadiev, Jurgis Pakerys, Inesa Šeškauskienė & Vaiva Žeimantienė (eds.), *Studies in Baltic and Other Languages*, 97–117. Vilnius: Vilniaus Universiteto Leidykla.
Croft, William. 2000. *Explaining language change: an evolutionary approach*. Harlow: Pearson Longman.
Croft, William. 2003. *Typology and universals*. Cambridge: Cambridge University Press.
Croft, William. 2016. Comparative concepts and language-specific categories: Theory and practice. *Linguistic Typology* 20(2). 377–393.
Croft, William. 2022. *Morphosyntax. Constructions of the World's Languages.* Cambridge: Cambridge University Press.
Cysouw, Michael & Bernhard Wälchli. 2007. Parallel texts: Using translational equivalents in linguistic typology. *Sprachtypologie und Universalienforschung* (*STUF*) 60(2). 95–99.
Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.
Daniel, Michael. 2010. Linguistic typology and the study of language. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 43–68. Oxford: Oxford University Press.
de Benito Moreno, Carlota. 2022. *The Middle Voice and Connected Constructions in Ibero-Romance: A Variationist and Dialectal Account.* Amsterdam/Philadelphia: John Benjamins.
Di Garbo, Francesca, Eri Kashima, Ricardo Napoleão de Souza & Kaius Sinnemämi. 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 SLI 2020*, 143–176. Milano: Officina 21.

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology (wals.info).

Evans, Nicholas. 2020. Introduction: Why the comparability problem is central in typology. *Linguistic Typology* 24(3). 417–425.

Everett, Caleb, Damián E. Blasi & Seán G. Roberts. 2015. Climate, vocal folds, and tonal languages: connecting the physiological and geographic dots. *PNAS* 112. 1322–1327.

Filppula, Markku, Juhani Klemola & Heli Paulasto (eds.). 2009. *Vernacular Universals and Language Contacts*. *Evidence from Varieties of English and Beyond*. New York: Routledge.

Gerdes, Kim, Sylvain Kahane & Xinying Chen. 2021. Typometrics: From Implicational to Quantitative Universals in Word Order Typology. *Glossa: a journal of general linguistics* 6(1). 17.

Ghyselen, Anne-Sophie & Gunther De Vogelaer. 2018. Seeking systematicity in variation: theoretical and methodological considerations on the "variety" concept. *Frontier in psychology* 9. 385.

Giacalone Ramat, Anna, Caterina Mauri & Piera Molinelli (eds.). 2012. *Synchrony and Diachrony. A dynamic interface*. Amsterdam/Philadelphia: John Benjamins.

Gil, David. 2016. Describing languoids: When incommensurability meets the language-dialect continuum. *Linguistic Typology* 20(2). 439–462.

Grandi, Nicola. 2014. *Fondamenti di tipologia linguistica*. Roma: Carocci.

Grandi, Nicola. 2018. Su alcune possibili interazioni tra tipologia e sociolinguistica. In Chiara Gianollo & Caterina Mauri (eds.), *CLUB Working Papers in Linguistics*, 257–265. Bologna: Alma Mater Studiorum – Univesità di Bologna.

Grandi, Nicola. 2021. Fattori sociolinguistici e costruzione del campione tipologico. Su alcune possibili interazioni tra tipologia e sociolinguistica. In Silvia Ballarè & Guglielmo Inglese (eds.), *Tipologia e sociolinguistica: verso un approccio integrato allo studio della variazione. Atti del Workshop SLI 2020*, 81–101. Milano: Officina 21.

Greenberg, Joseph H. 1960. A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26(3). 178–194.

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: MIT Press.

Greenberg, Joseph H. 1995. The diachronic typological approach to language. In Masayoshi Shibatani & Theodora Bynon (eds.), *Approaches to language typology*, 145–166. Oxford: Clarendon Press.

Haig, Geoffrey, Stefan Schnell & Frank Seifart (eds.). 2021. *Doing Corpus-Based Typology with Spoken Language Corpora: State of the art.* Honolulu: University of Hawai'i Press.

Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42(1). 25–70.

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.

Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1). 31–80.

Haspelmath, Martin (ed.). 2014-. *Studies in Diversity Linguistics.* Berlin: Language Science Press.

Haspelmath, Martin. 2016. Universals of causative and anticausative verb formation and the spontaneity scale. *Lingua Posnaniensis* 58(2). 33–63.

Haspelmath, Martin. 2018. How comparative concepts and descriptive linguistic categories are different. In Daniël Van Olmen, Tanja Mortelmans & Frank Brisard (eds.), *Aspects of linguistic variation*, 83–114. Berlin/New York: De Gruyter Mouton.

Haspelmath, Martin. 2019. Can cross-linguistic regularities be explained by change constraints? In Karsten Schmidtke-Bode, Natalia Levshina, Susanne Maria Michaelis & Ilja A. Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, 1–23. Berlin: Language Science Press.

Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form-frequency correspondencies and predictability. *Journal of Linguistics* 57(3). 605–633.

Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie (eds). 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.

Hinskens, Frans & Gregory R. Guy (eds.). 2016. *Coherence, covariation and bricolage. Various approaches to the systematicity of language variation*. Amsterdam: Elsevier. [= Special issue of *Lingua* 172/173].

Heine, Bernd & Heiko Narrog (eds.). 2018. *The Oxford Handbook of Grammaticalization*. Oxford: Oxford University Press.

Hopper, Paul. 1991. On some principles of grammaticalization. In Elizabeth C. Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, *Vol. 1, Theoretical and methodological issues*, 17–35. Amsterdam/Philadelphia: John Benjamins.

Hopper, Paul J. & Elizabeth Closs Traugott (eds.). 2003. *Grammaticalization*. 2nd edn. Cambridge: Cambridge University Press.

Hudson, Richard A. 1980. *Sociolinguistics*. Cambridge: Cambridge University Press.

Iannàccaro, Gabriele. 2000. Per una semantica più puntuale del concetto di 'dato linguistico'. *Quanderni di semantica* 21(1). 51–79.

Keenan, Edward & Bernard Comrie. 1977. Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8. 63–99.

Keenan, Edward & Bernard Comrie. 1979. Noun phrase accessibility revisited. *Language* 55(3). 649–664.

Koptjevskaja-Tamm, Maria. 2012. New directions in lexical typology. *Linguistics* 50(3). 373–394.

Koptjeskaja-Tamm, Maria, Ekaterina Rakhilina & Martine Vanhove. 2015. The semantics of lexical typology. In Nick Riemer (ed.), *The Routledge Handbook of Semantics*, 434–454. London: Routledge.

Kortmann, Bernd (ed.). 2004a. *Dialectology Meets Typology: Dialect Grammar from a Cross-Linguistic Perspective*. Berlin/New York: De Gruyter Mouton.

Kortmann, Bernd. 2004b. Introduction. In Bernd Kortmann (ed.), *Dialectology Meets Typology: Dialect Grammar from a Cross-Linguistic Perspective*. Berlin/New York: De Gruyter Mouton.

Kortmann, Bernd & Krestin Lunkenheimer. 2013. *The Mouton World Atlas of Variation in English*. Berlin/New York: De Gruyter Mouton.

Kortmann, Bernd, Krestin Lunkenheimer & Katharina Ehret (eds.). 2020. *The Electronic World Atlas of Varieties of English* (https://ewave-atlas.org/).

Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog & Seongha Rhee. 2019. *World lexicon of grammaticalization. 2nd, extensively revised and updated edition*. Cambridge: Cambridge University Press.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 2001. *Principles of Linguistic Change. Vol. II. Social Factors*. Oxford: Blackwell.

Labov, William. 2004. Quantitative reasoning in linguistics. In Ulrich Ammon, Norbert Dittmar & Klaus J. Mattheier (eds.), *Sociolinguistics/Soziolinguistik: An International Handbook of the Science of Language and Society*. Volume 1, 6–22. Berlin/New York: De Gruyter Mouton.

Labov, William, Sharon Ash & Charles Boberg. 2006. *The Atlas of North American English: Phonetics, phonology and sound change*. Berlin/New York: De Gruyter Mouton.

Lameli, Alfred. 2010. Linguistic Atlases – traditional and modern. In Peter Auer & Jürgen Erich Schmidt (eds.), *Language and Space, Part I: Theories and Methods. An International Handbook of Linguistic Variation*, 567–592. Berlin/New York: De Gruyter Mouton.

Lavandera, Beatriz. 1978. Where does the sociolinguistic variable stop? *Language in Society* (7)2. 171–182.

Lehmann, Christian. 1985. Grammaticalization: Synchronic variation and diachronic change. *Lingua e Stile* 20. 303–318.

Lehmann, Christian. 2015 [1995]. *Thoughts on grammaticalization*. 3rd edn. Berlin: Language Science Press.

Levshina, Natalia. 2019. Token-based typology and word order entropy: A study based on Universal Dependencies. *Linguistic Typology* 23(3). 533–572.

Levshina, Natalia. 2021. Corpus-based typology: applications, challenges and some solutions. *Linguistic Typology* 26(1). 129–160.

Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PLOS one* 5(1). 1–10.

Luraghi, Silvia. 2010. Causes of language change. In Silvia Luraghi & Vit Bubenik (eds.), *Continuum Companion to Historical Linguistics*, 354–366. London – New York: Continuum.

Malchukov, Andrej. 2005. Case pattern splits, verb types and construction competition. In Mengistu Amberber & Helen de Hoop (eds.), *Competition and variation in natural languages: The case for case*, 73–117. Amsterdam: Elsevier.

Meyerhoff, Miriam, Erik Schleef & Laurel MacKenzie. 2015. *Doing sociolinguistics. A practical guide to data collection and analysis*. London: Routledge.

Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.). 2013. *Atlas of Pidgin and Creole Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology*. (http://apics-online.info).

Miestamo, Matti. 2005. *Standard negation: the negation of declarative verbal main clauses in a typological perspectiv*e. Berlin/New York: De Gruyter Mouton.

Mithun, Marianne. 2018. Deconstructing teleology: The place of synchronic usage patterns among processes of diachronic development. In Sonia Cristofaro & Fernando Zúñiga (eds.), *Typological hierarchies in synchrony and diachrony*, 111–128. Amsterdam/Philadelphia: John Benjamins.

Moran, Steven & Daniel McCloy (eds.). 2019. *PHOIBLE 2.0*. Jena: Max Planck Institute for the Science of Human History. https://phoible.org/

Moravcsik, Edith A. 2010. Explaining language universals. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 69–89. Oxford: Oxford University Press.

Moravcsik, Edith A. 2016. On linguistic categories. *Linguistic Typology* 20(2). 417–425.

Murelli, Adriano & Bernd Kortmann. 2011. Non-standard varieties in the areal typology of Europe. In Bernd Kortmann & Joan van der Auwera (eds.), *The Languages and Linguistics of Europe. A Comprehensive Guide*, 525–544. Berlin/New York: De Gruyter Mouton.

Muysken, Pieter, Harald Hammarström, Olga Krasnoukhova, Neele Müller, Joshua Brichall, Simon van de Kerke, Loretta O'Connor, Swintha Danielsen, Rik van Gijn & George Saad. 2016. *South American Indigenous Language Structures (SAILS) Online*. Jena: Max Planck Institute for the Science of Human History. (http://sails.clld.org).

Narrog, Heiko. 2017. Typology and grammaticalization. In Alexandra Aikhenvald & R. M. W. Dixon (eds.), *The Cambridge Handbook of Linguistic Typology*, 151–177. Cambridge: Cambridge University Press.

Nevalainen, Terttu & Minna Palander-Collin. 2011. Grammaticalization and sociolinguistics. In Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Grammaticalization*, 118–128. Oxford: Oxford University Press.

Nichols, Johanna. 1992. *Linguistic Diversity in Space and Tim*e. Chicago: University of Chicago Press.
Paschen, Ludger, François Delafontaine, Christoph Draxler, Susanne Fuchs, Matthew Stave & Frank Seifart. 2020. Building a time-aligned cross-linguistic reference corpus from language documentation data (DoReCo). *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 2657–2666. Marseille: European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2020/pdf/2020.lrec-1.324.pdf
Paschen, Ludger, Susanne Fuchs & Frank Seifart. 2022. Final Lengthening and vowel length in 25 languages. *Journal of Phonetics* 94. 101179. https://doi.org/10.1016/j.wocn.2022.101179.
Pakendorf, Brigitte, Nina Dobrushina & Olesya Khanina. 2021. A typology of small-scale multilingualism. *International Journal of Bilingualism* 25(4). 835–859.
Pichler, Heike. 2013. *The structure of discourse-pragmatic variation*. Amsterdam/Philadelphia: John Benjamins.
Poplack, Shana. 2011. Grammaticalization and linguistic variation. In Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Grammaticalization*, 209–224. Oxford: Oxford University Press.
Poplack, Shana. 2021. Le corpus comme portail pour l'étude de la variation (socio)linguistique. *Corpus* [en ligne] 22.
Rickford, John R. 2002. Implicational scale. In Jack K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 142–167. Oxford: Blackwell.
Round, Erich R. & Greville G. Corbett. 2020. Comparability and measurement in typological science: The bright future for linguistics. *Linguistic Typology* 24(3). 489–525.
Sanga, Glauco. 1991. I metodi della ricerca sul campo. *Rivista Italiana di Dialettologia* 15. 165–181.
Sankoff, Gillian. 1972. Above and beyond phonology in variable rules. In Charles-James Bailey & Roger W. Shuy (eds.), *New ways of analyzing variation in English*, 44–61. Washington: Georgetown University Press.
Sansò, Andrea. 2018. Explaining the diversity of antipassives: Formal grammar vs. (diachronic) typology. *Language and Linguistics Compass* 12(6). e12277.
Sansò, Andrea. 2020. *I segnali discorsivi*. Roma: Carocci.
Sapir, Edward. 1921. *Language. An Introduction to the Study of Speech*. New York: Harcourt, Brace & World.
Saussure, Ferdinand de 2016 [1916]. *Cours de linguistique générale.* Edited by Charles Bally & Albert Sechehaye. Paris: Payot & Rivages.
Schneider. Klaus P. & Anne Barron (eds.). 2008. *Variational pragmatics*. Amsterdam/Philadelphia: John Benjamins.
Seiler, Guido. 2018. Synchrony and diachrony: two outdated dimensions? In Martin Glessgen, Johannes Kabatek & Harald Völker (eds.), *Repenser la variation linguistique. Actes du Colloque DIA IV à Zurich (12–14 sept. 2016)*, 77–96. Strasbourg: Éditions de linguistique et de philologie.
Seiler, Guido. 2019. Non-Standard Average European. In Andreas Nievergelt & Ludwig Rübekeil (eds.), *'athe in palice, athe in anderu sumeuuelicheru stedi'. Raum und Sprache. Festschrift für Elvira Glaser zum 65 Geburtstag*, 541–554. Heidelberg: Winter.
Skirgård, Hedvig et al. 2023. Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances* 9(16). eadg6175. doi: 10.1126/sciadv.adg6175.
Siegel, Jeff, Benedikt Szmrecsanyi & Bernd Kortmann. 2014. Measuring analyticity and syntheticity in creoles. *Journal of Pidgin and Creole Languages* 29(1). 49–85.
Song, Jae Jung. 2010. Word order typology. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 253–279. Oxford: Oxford University Press.
Stassen, Leon. 1985. *Comparison and Universal Grammar*. Oxford: Blackwell.

Stassen, Leon. 2010. The Problem of Cross-Linguistic Identification. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 90–99. Oxford: Oxford University Press.
Stave, Matthew, Kilu von Prince & Frank Seifart. 2021. A usage-based approach to morphological typology. Paper presented at the *Societas Linguistica Europaea (SLE) 2021. Workshop: Integrating sociolinguistic and typological perspectives on language variation*, *30 August – 3 September 2021*, online.
Tagliamonte, Sali A. 2000. The grammaticalization of the present perfect in Eglish: Tracks of change and continuity in a linguistic enclave. In Olga Fischer, Anette Rosenbach & Dieter Stein (eds.), *Pathways of Change: Grammaticalization in English*, 329–354. Amsterdam/Philadelphia: John Benjamins.
Tagliamonte, Sali A. 2005. So who? Like how? Just what? Discourse markers in the conversations of young Canadians. *Journal of Pragmatics* 37(11). 1896–1915.
Tagliamonte, Sali A. 2006. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.
Tagliamonte, Sali A. 2012. *Variationist sociolinguistics. Change, observation, interpretation*. Malden: Wiley-Blackwell.
Torres Cacoullos, Rena. 2011. Variation and Grammaticalization. In Manuel Díaz Campos (ed.), *The Handbook of Hispanic Sociolinguistics*, 148–167. Oxford: Blackwell.
Trudgill, Peter. 2011a. *Sociolinguistic typology*. Oxford: Oxford University Press.
Trudgill, Peter. 2011b. Social structure and phoneme inventories. *Linguistic Typology* 15(2). 155–160.
Tsunoda, Tasaku. 1985. Remarks on transitivity. *Journal of Linguistics* 21. 385–396.
Urban, Matthias. 2020. Mountain linguistics. *Language and Linguistics Compass* 14(9). e12393.
Velupillai, Viveka. 2012. *An introduction to linguistic typology*. Amsterdam/Philadelphia: John Benjamins.
Viberg, Åke. 1984. The Verbs of Perception: a Typological Study. *Linguistics* 21. 123–162.
Weiner, Judith & William Labov. 1983. Constraints on the agentless passive. *Linguistics* 19. 29–58.
Weinreich, Uriel, William Labov & Marvin I. Herzog. 1968. Empirical foundations for a theory of language. In Winfred P. Lehmann & Yakov Malkiel (eds.), *Directions for historical linguistics: A symposium*, 95–188. Austin: University of Texas Press.
Witzlack-Makarevich, Alena. 2019. Argument Selectors. A new perspective on grammatical relations: An introduction. In Alena Witzlack-Makarevich & Balthasar Bickel (eds.), *Argument Selectors: A new perspective on grammatical relations,* 1–38. Amsterdam/Philadelphia: John Benjamins.
Wray, Alison & George W. Grace. 2007. The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117(3). 543–578.
Zeman, Daniel, Joakim Nivre et al. 2020. *Universal Dependencies 2.6*. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. Available at: http://hdl.handle.net/11234/1-3226. See also http://universaldependencies.org.