



Universiteit
Leiden
The Netherlands

Linguistic areas

Gijn, E. van; Wahlstrom, M.; Ruch, H.; Wahlström, M.; Hasse, A.

Citation

Gijn, E. van, & Wahlstrom, M. (2023). Linguistic areas. In H. Ruch, M. Wahlström, & A. Hasse (Eds.), *Language contact* (pp. 179-219). Berlin: Language Science Press. doi:10.5281/zenodo.8269244

Version: Publisher's Version
License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)
Downloaded from: <https://hdl.handle.net/1887/3663452>

Note: To cite this publication please use the final published version (if applicable).

Chapter 7

Linguistic areas

Rik van Gijn^a & Max Wahlström^b

^aLeiden University ^bUniversity of Helsinki

Linguistic area research has received ample attention in the last century. Nevertheless, methodology remains somewhat underdeveloped, and there seem to be few, if any, generalizations about the relation between the processes underlying area formation and their outcomes. The main challenge is that, in most cases, the past is not directly accessible and therefore has to be reconstructed. Linguistic area research, therefore, stands to gain immensely from a firm embedding into a framework that includes both other strands of contact linguistics and extra-linguistic disciplines to complete the picture.

1 Introduction

A linguistic area is a geographical region where several languages are spoken that have become similar to each other as a result of sustained contact between the speech communities. Although this description is intuitively simple, finding a satisfactory definition of what is and what is not a linguistic area is extremely difficult, if it is possible at all (see, e.g., Masica 2001, Stolz 2002, 2006, Campbell 2006, 2017). There are three central problems in defining a linguistic area: the boundary problem, the language problem, and the feature problem.

1. **The boundary problem:** Establishing the geographical boundaries of a linguistic area is often based on the distribution of features. This is problematic because the distributions of different features rarely overlap completely.
2. **The language problem:** There seems to be no non-arbitrary way to determine the minimum number of languages required to speak of a linguistic area.



3. **The feature problem:** There are no established criteria to determine the diagnostic value of features for particular linguistic areas, nor of the minimum number of features required.

These and other problems have led some researchers to suggest that the term linguistic area should be abandoned, as it is a mental construct rather than a reality (see in particular Stolz 2006, Campbell 2006, 2017). However, linguistics is full of mental constructs (a language is one, to start with) that have still proven their worth as workable concepts. Therefore, the question to answer is whether or not linguistic areas are useful concepts. The concept of linguistic area, in our view, has a number of useful applications.¹

- They are informative to typologists for sampling purposes (especially larger areas) – see below.
- They are informative to field linguists, providing an expectation pattern for certain grammatical features, and should be part of the preparation of any field project in which they are relevant.
- They are informative to historical linguists, helping them understand contact-induced developments of members of a family that are part of a linguistic area.
- They serve a purpose of their own: they can tell us something about the interaction between geography, human behavior, and language or communication strategies.

1.1 A brief history of the field

The history of the notion of a linguistic area can be traced back to the early 20th century. Linguistics had emerged from the 19th century as an independent field, with one particularly powerful tool: the study of regular sound correspondences to identify genealogical relations between languages. Yet more and more observations were made about similarities among languages that could not be attributed to genealogical relatedness as data on previously undescribed languages and newly discovered language families accumulated.

¹A separate issue is how we should go about establishing linguistic areas or areal skewing of linguistic features, e.g. top-down versus bottom-up (Muysken et al. 2014, Campbell 2017) linguistics-first or starting from other disciplines (Stolz 2006, Van Gijn 2020). We come back to this issue in Section 2.

In Europe, observations about structural similarities among the distantly related Indo-European languages of the Balkans date back to as early as 1810 (Friedman 2011: 177). At the very beginning of the 20th century, Jan Baudouin de Courtenay (Boduen-de-Kurtenè 1901), with the Balkan languages in mind, proposed an approach to multilateral influences among geographically close varieties that would entail, for instance, the study of societal history. In a similar vein, Kristian Sandfeld (1926: 8) argued for a dedicated field of study and urged linguists to treat the Balkan languages “as one unit and make them into a starting point of a comprehensive study.”

In Northern America, the study of the Native American languages showed the limits of explaining similarities with genealogical relatedness (see e.g. Mithun 2017: 881–882). Franz Boas described in 1911 how linguistic structure can diffuse across genealogical boundaries between languages (Boas 1911: 47–53). Later, Boas (1929: 6) described the North Pacific coast as a geographical area where genealogical classifications are not helpful in explaining grammatical similarities among neighboring languages.

An interest in the effect of geography on linguistic structure had also arisen in the Italy-based neolinguistic school of thought, famous for arguing, perhaps less convincingly, against some tenets of the historical-comparative tradition. However, Matteo Bàrtoli observed the diffusion of innovations in space and considered, for instance, the effect of prestige in the adoption of linguistic features (Hall 1946).

In a timely development, Nikolai Trubetzkoy (1928: 18) proposed the concept *Sprachbund* in the first International Congress of Linguistics in 1928. The German term *Sprachbund* was a subconcept of *Sprachgruppe*, an umbrella term intended to remain agnostic about the origin of the similarities among the observed languages. *Sprachbund* contrasted with language family (*Sprachfamilie*), and was effectively worded as the absence of the defining criteria for genealogical relatedness. The formulation also made a strong claim about the languages forming a *Sprachbund*: The languages in such a union greatly resemble each other with regard to syntax and principles of morphological structure, and they share a great number of cultural words, as well as sometimes superficial similarities in sound systems.

Despite a number of discoveries of diffusion of linguistic structures across genealogical borders, the Balkan situation is often argued to have been the model for the delineation of *Sprachbünde* (see, e.g., Friedman 2011: 276). However, while Trubetzkoy (1923) does mention the Balkans as a “shining example” of a *jazykovej sojuz* (‘language union’) in a footnote in an earlier article, this was done in order

to promote his ideas of geographically much larger areas such as the “Mediterranean” and “Ural-Altai” unions, which comprise several entire language families.

The concept of Sprachbund proved to be popular. As early as 1931, Roman Jakobson (1962) used the term to describe what he called “the Eurasian Sprachbund.” This was followed up most notably by Harry Velten (1943: 271), who called the American Northwest a “linguistic area,” in a translation of *Sprachbund*, and by Murray Emeneau (1956), introducing the Indian linguistic area. Worth mentioning is also Hans Kurath’s concept “area linguistics,” figuring in the title of what may be the first textbook on the subject, “Studies in area linguistics” (Kurath 1972). Kurath (1972: 1–23) provides, for instance, methodology and a field work guide for the “area linguist.” While Kurath’s own research dealt with North American dialects of English and he does not dwell on the definition of a linguistic area, he clearly regards, for instance, Emeneau’s (1956) work as illustrative of the field.

Meanwhile, the upcoming field of linguistic typology showed a growing interest in areal linguistics. In a seminal paper, Bell (1978) discusses biases in typological samples that can compromise the independence of sample units. One of the biases Bell suggests is areal, because languages that are spoken in each other’s vicinity may have influenced each other through contact, calling for integration of insights from areal linguistics into typology. Dryer (1989) makes a concrete proposal for areal stratification in typological samples. He divides the world into five large areas (Africa, Eurasia, Australia-New Guinea, North America, and South America). Dryer argues that these areas can be assumed to be independent from each other, but within each area, contact effects can be expected (see Hammarström & Donohue 2014 for critical discussion).

Nichols (1992) takes the idea of macro areas and the role they play in the distribution of linguistic features over the globe one step further. For her, language diversity patterns are the objects of typological inquiry, and historical processes (e.g. migrations, linguistic diversification, areal contact) the explanations for these patterns. In this approach to typology, then, areal patterns (especially at the macro level) are no longer only relevant for sampling purposes, but they become a research outcome that needs to be explained.

In less than a hundred years, research concerning linguistic areas has grown into a considerable field, with hundreds of proposed linguistic areas of various sizes and time depths, distributed over all continents (see Campbell 2017 for a list). It has established itself as an independent branch of contact linguistics, with its own research agenda. In what follows we will give an overview of the major conceptual and methodological components of this fascinating field.

1.2 Key terms and concepts

The study of linguistic areas typically operates on terminology shared by other subfields of linguistics. There are, however, a few concepts that are specific to, or receive a particular interpretation in, the context of areal linguistics. A key term in connection with linguistic areas is *adstrate* or *adstratum*, which refers to a language that has influenced the language of a neighboring population, without there being clear differences in prestige between the two groups. Of Ascoli's classical sociolinguistic settings for language contact (see Tristram 2007: 195–96), adstrate effects, more often than *substrate* or *superstrate* effects, are seen as the primary mechanism in the formation of linguistic areas (cf., however, Section 3.1; regarding the three terms, see also discussion in Chapter 5).

Another concept central to the study of linguistic areas is *convergence*, which can be described as the historical process by which languages become structurally more similar to each other as the result of prolonged contact. What the term means, however, differs depending on whether one focuses on the larger processes of area formation or on an individual instance of contact-induced language change. Joseph (2010), among others, uses convergence to describe the overall process resulting from multiple instances of contact-induced language change, implying mutual changes toward a common outcome. Yet according to this use of the term, there can be several mechanisms contributing to convergence, including substrate and adstrate effects, pidginization (see Chapter 5), and speaker-to-speaker accommodation (see Chapter 2).

Matras & Sakel (2007) use convergence to describe a particular type of contact-induced language change, characteristic of linguistic areas: convergence is a shift in the meaning and functional distribution of inherited linguistic material, contrasting with grammatical and lexical borrowing. Convergence in this sense coincides closely with what Ross (2007) calls *metatypy*, which in his terminology stands in opposition to calquing. Adstrate and convergence, in the former sense, highlight the common ideas that, first, linguistic areas arise in situations of symmetric relationships between the different ethnolinguistic groups, and, second, that the influence is mutual. We come back to this point in Section 3.2. The resulting situation is sometimes characterized as (*mutual*) *isomorphism*, structural uniformity across languages (see Matras 1998).

Another recurring term is *isogloss*, a line on a map that defines an area in which the languages share a particular feature (or, more rarely, in the case of exclusively structural features *isogrammatism* is also used, see Gołąb 1956). Each isogloss, therefore, describes the distribution of a single feature. In the almost non-existent ideal case, several overlapping isoglosses define the boundaries of a

linguistic area. A related but less commonly used concept is *isopleth* (see Van der Auwera 1998), which is a set of lines on a map grouping languages that display the same number of features from a set of shared features, although these do not have to be the same across the languages. Using isopleths, one can show where most features are shared.

Isoglosses often play an important role in determining the *boundaries* of a linguistic area. As mentioned, boundaries are a problematic part of the study of linguistic areas because there is no unambiguous way to determine them. Two basic approaches to boundaries (see Section 2) are to look for *hard boundaries*, generally based on physical geography, or to allow for *fuzzy boundaries*, where the outer limits of a linguistic area are left partly undefined, but coincide more or less with the areas where the number of shared features dwindles below a certain threshold (Muysken et al. 2014).

1.3 Linguistic areas and contact linguistics

The study of linguistic areas relates to contact linguistics in that it tries to infer the historical contact situations and contact processes that have given rise to a linguistic area from synchronic linguistic data (the distribution of linguistic features) as well as, where possible, data from other disciplines, such as cultural and geographical data. Given its character, the study of linguistic areas may tell us something about the long-term results of sustained contact between groups of people in a particular area. Thus the study of linguistic areas fills two gaps within contact linguistics: it registers results of long-term contact that are generally unavailable to subdisciplines in contact linguistics that target the individual or a specific language community, and it registers the geographical extents of specific features, thus including a spatial dimension. This latter point is shared with the study of dialect areas, but not with more individually- or societally-oriented approaches to contact linguistics. Table 1 indicates the position of areal linguistics within contact linguistics in terms of scale (both temporal and spatial),² adapted and expanded from Muysken (2008). Areal linguistics concerns the two bottom rows. Therefore it has its own research questions, methods, and data sources.

Linguistic area studies share with other subdisciplines of contact linguistics an interest in establishing what elements of language are especially sensitive to contact situations. The systematic study of areal features in linguistic areas may reveal not only tendencies in what features are prone to contact-induced change,

²Attention in areal linguistics has been heavily tilted toward the spatial scale, largely in the absence of historical data. An interesting exception is Dedio et al. (2019), which traces the development of area formation for the British Isles over a period of about 1300 years.

Table 1: Level of scales in contact linguistics (adapted from Muysken 2008)

Time depth	Space	Sources	Scenarios
Real time	Bilingual individual(s)	Conversation analysis, psycholinguistic experiments	Situational strategies
20–200 years	Bilingual community	Fieldwork	Specific contact scenarios
Ca. 200–1000 years	Geographical region	Comparative data, historical sources	Global contact scenarios
Deep time	Larger areas of the world	Typology, genetics, archaeology	No/vague contact scenarios

but also which foreign linguistic elements become part of languages beyond incidental use (a research goal shared with societal studies such as e.g. the study of contact varieties). Linguistic area studies, moreover, have the potential to go one step further: they can isolate those features that persist in an area over long periods of time, sometimes thousands of years. This latter point is connected to the term ‘areal stability’, discussed in Nichols (1992) and Nichols (2003). Just as language families can be consistent with respect to a particular linguistic feature in that all or most of its members have it, so can areas. The more consistent an area is with respect to a feature, the more areally stable that feature is.

As mentioned, the spatial aspect of language contact is shared with the study of dialect areas. This shared objective yields some obvious desiderata for cross-pollination between linguistic area research and dialect area research, especially when it comes to integrating geographical models into the study of linguistic areas, where it is an underdeveloped aspect. The contrast between dialect areas and linguistic areas that is most interesting to contact linguistics more generally is the fact that the languages studied in dialect research are very closely related, whereas in linguistic areas they are by default unrelated or very distantly related. This is interesting because of the persistent claim that closely related languages allow for many more cross-overs from one language to another (e.g. Weinreich 1953, Moravcsik 1975, Winford 2005, see Section 4.1 for more discussion).

The overarching goal that drives areal linguistics is to uncover non-accidental signals of similarity that cluster geographically, which can, moreover, be shown to be due to contact between the speakers of languages spoken in that area. A third goal (which is not always pursued) is to establish the type of contact that existed between the speakers of the different languages.

2 Approaches

The main objectives of the study of linguistic areas given above translate into three basic steps a researcher takes to make the case for a linguistic area:

1. Determine the area of interest;
2. Establish distribution patterns of linguistic characteristics in that area;
3. Establish that the distribution patterns are the result of contact.

A fourth step would involve establishing the type of social scenario that gave rise to the linguistic area. This step is often not taken in linguistic area research, but from the perspective of this book, it is a crucial step that allows us to connect linguistic areas to the broader field of contact linguistics. In order to give that question the space it requires in the context of this book, we postpone its discussion to the next section, and focus here on steps 1–3.

2.1 Determining an area of interest

Making the case for a linguistic area starts with an expectation or suspicion that a particular area may be of interest from the perspective of areal diffusion. This expectation may be based on linguistic or extra-linguistic indicators.

2.1.1 Extra-linguistic indicators

Stolz (2006) lists three non-linguistic sources that are potential entry points which may lead to the postulation of a linguistic area:³ geography, cultural history, and observed communicative practices. Each of these source types has played a role in giving rise to linguistic area hypotheses.

A geography-based approach is proposed in Ranacher et al. (2017), who develop an algorithm that proposes random potential diffusion areas based on the

³Stolz believes that none of these three is sufficient by itself to delimit a linguistic area, but one can still regard them as potential starting points (see Van Gijn 2020.)

dense river network of the Amazon, where rivers are generally considered to function as pathways, facilitating contact. In a second step, each proposed area is tested for its likelihood of being an actual diffusion area based on the languages showing non-accidental linguistic similarities. After evaluating all the proposals, the best (i.e. most likely) diffusion areas are selected.

Examples of cultural history-first approaches can be found in the North American tradition of the 1970s, where areal linguistics was profoundly influenced by the anthropological tradition of studying culture areas (Mithun 2017). In a late example of this tradition, Bereznak (1995) argues that the Pueblo region in the American Southwest, a proposed culture area on the basis of ethnographic data, should be regarded as a linguistic area as well.

Direct observations of sociolinguistic communicative dynamics that are responsible for the rise of a linguistic area are not common. The example that comes closest to a communication-first account of a contact zone is arguably Gumperz & Wilson (1971), who explicitly start out with observations about the sociolinguistic dynamics of the Kupwar village (see Subsections 3.1.2 & 3.2.2), and then investigate what effect these have had on the three main languages spoken there.

An approach that combines several types of non-linguistic information to establish areas of interest is proposed by Bickel & Nichols (2006), who introduce so-called *Predictive Areality Theory* (PAT). PAT incorporates information from a range of disciplines such as genetics, archaeology, ecology, demography, and topography (but crucially not linguistics) to establish areas where interethnic contacts can be presumed to have taken place. In a second step, the linguistic area hypothesis is tested by looking at linguistic data. The test consists of showing that the languages within the proposed area are significantly more similar to each other than to languages outside the area. The role of geography (topography) in PAT is interesting in that it sets boundaries for the area (e.g. mountains, seas, oceans), but this does not mean that they expect isoglosses to coincide with those natural boundaries. In fact, they expect spill-over both ways: languages that are similar to the areal profile but are in fact spoken outside it and, contrarily, dissimilar languages that are spoken inside the area. The authors say that this is to be expected as language groups migrate into and out of the areas with some regularity.

2.1.2 Linguistic indicators

Suspensions of linguistic areas also often start with observations of certain linguistic patterns. These can be individual observations by areal specialists of con-

spicuous forms or constructions that are found in different unrelated languages. These observed shared constructions are rare or very specific, so that they warrant further investigation of the area. This is in fact how the idea of the Balkan Sprachbund first developed. Another type of observation that may lead to further inspection of an area is based on global or large areal surveys of distributions of linguistic features, which may highlight certain areas that behave differently from the rest of the survey. Depending on the density of the sample, this procedure is particularly suitable for finding macro areas (see, e.g., Nichols 1992).

2.2 Establishing feature distribution patterns

Once an area of interest has been established, it can be subjected to closer linguistic scrutiny. Again, there are essentially two ways to do this, which have been termed *bottom-up* and *top-down* (Muysken 2008, Muysken et al. 2014).

In a bottom-up approach, one starts with a few observations (see above) and from there one starts investigating the languages more thoroughly, thus collecting a catalog of areally distributed features. The top-down approach essentially refers to what has been termed “areal typology” (Dahl 2001), where the languages of an area are coded for a set of features for which there is no particular *a priori* suspicion of areal diffusion (see Subsection 2.1.1 for *Predictive Areality Theory*).

There are advantages and disadvantages to both approaches. The bottom-up approach is more likely to yield very specific, detailed constructions and patterns, but it is less systematic and more prone to being influenced by the subjectivity of the researcher. The top-down approach may miss some of the more specific patterns that may be crucial to establishing a linguistic area (see, e.g., Friedman 2011: 278–279), but it is less susceptible to cherry-picking, with its subjective overtones, and better suited for quantification (see Muysken et al. 2014).

2.3 Establishing contact-induced origin

Nonrandom areal patterning of a feature may be regarded as a result in itself, yet, as mentioned in Section 1.3, such discoveries are typically followed by an attempt to establish the contact-induced origin of the phenomenon. This is needed mainly to eliminate common inheritance and universal pressures as an explanation.

An idealized procedure to determine the contact-induced origin of a feature is described in Thomason (2010). The steps to be taken are: 1) look at the languages as a whole: if structural interference of some kind has occurred, it is highly unlikely to be an isolated instance; 2) identify the source language and show that the

contact was sufficiently intense; 3) identify shared structural features in the proposed source language and in the receiving language; 4) prove that the feature did not exist in the receiving language prior to the proposed contact, if necessary, by examining the languages related to the recipient language; and 5) prove that the proposed feature was present in the source language before it came into contact with the recipient language.

From the perspective of linguistic areas, the two first steps prove to be particularly problematic. Step 1), while completely logical for establishing individual contact phenomena, leads to circularity with linguistic areas: the presence of other shared features is typically what encourages the examination of a suspected feature in the first place. Step 2) is problematic, since determining the source language is often difficult in situations of areal diffusion. For instance, in the Balkans, most of the classical features must be regarded as innovations vis-à-vis the earlier attested forms of all contributing languages (Lindstedt 2018).

In reality, only under exceptional circumstances can more than two of the five research steps mentioned above be satisfactorily observed in a typical research setting, and other criteria have been used to make a case for a linguistic area. Campbell (1985: 31–36) distinguishes historicist from circumstantialist approaches. The former include a broad range of non-linguistic historical data to back up a claim for a linguistic area, whereas the latter operate only on synchronic typological variables and therefore risk missing alternative explanations, such as “undiscovered genetic relationships, universals, onomatopoeia, parallel or independent development, sheer chance, etc.”

Distinguishing between inherited and contact-induced features has been at the core of several debates on linguistic areas and areality. For instance, both Sherzer (1976) and Dixon (2002) were criticized for either overlooking, ignoring, or misrepresenting historical-comparative evidence in their arguments for contact-induced areal phenomena (for the criticism, see Campbell 1985, Evans 2005). Conversely, using typological variables to argue for a genealogical relationship is equally problematic as they are prone to diffusing across genealogical boundaries, as Donohue et al. (2008) demonstrate regarding an alternative family tree of Island Melanesian Papuan languages, which has been proposed by Dunn et al. (2007).

The question of chance or independent parallel development in producing areal patterning is complex. Bickel & Nichols (2006) mention a “sufficiently low [cross-linguistic] frequency” of a variable for it to be used as a diagnostic criterion for an area hypothesis. To exemplify this, they mention variables with the cross-linguistic frequency of one-in-four, arguing that with such a high frequency, in any set of, say, 200 variables, five such random bundlings can be

expected among a handful of unrelated languages. What is implied is that, by cherry-picking the variables, any random set of languages can be said to form a linguistic area. Yet in most descriptions of smaller areas, the analysis is finer-grained and does not rely solely on abstract typological variables. In addition, there may be universal pressures that may give rise to independently developed similarities between languages, making it particularly important to take into account universal preferences in language as well (Bickel 2017).

Finally, in establishing the contact-induced origin of a feature, one needs to weigh its independence from the other area-defining features. For instance, the hypothesized European linguistic area, Standard Average European, as well as the Balkans, is characterized by the reduction of case inflection and the rise of definite and indefinite articles. It appears that there is a cross-linguistically observable inverse correlation between the number of cases and there being an article system, implying an interdependence (Sinnemäki & Wahlström 2018). However, such a statistical universal may only weaken a contact hypothesis relative to the strength of the correlation, and only a strict universal or logical dependency, where the presence of feature X can fully predict the presence of feature Y and vice versa, can reduce the diagnostic value of one of these features to zero for a contact phenomenon.

3 Processes and patterns

As with several other concepts in linguistics, linguistic areas can be interpreted either as mainly describing a particular, synchronically observable situation or the process that gives rise to it, depending on one's point of view. This section assesses the processes of area formation and the resulting patterns separately – in other words, it addresses the question of what happens within linguistic areas and how it happens. Both processes and patterns will be illustrated through the case examples of three relatively different proposed linguistic areas: Vaupés, Kupwar, and the Balkans.

3.1 Processes

Given the inferential character of areal linguistics, there is usually no certainty about what processes give rise to linguistic areas. Matras (2011) suggests that contact-induced language change goes through the same cycle as internal language change:

1. There is a spontaneous innovation in discourse

2. that is propagated in society, and
3. becomes part of the system that first-language learners acquire.

Nevertheless, how this process takes shape in multilingual societies is special in a number of ways, especially in the first step.

Matras (2011) claims that bilingual speakers do not separate the language systems in their head neatly. Selections out of a repertoire of constructions are made according to the social setting. More abstract form–meaning mappings (constructions) are less subject to strict selection constraints than are more concrete and tightly organized form–meaning mappings such as word forms, making it more likely for the former to be generalized across linguistic contexts. Furthermore, at the societal level, successful propagation of these individual innovations requires a particular societal structure in which there are lax norms when it comes to language use, that is, speakers are allowed to vary in their repertoire (Matras 2011: 157). At the same time, contact-induced language change arises in situations where the speech communities make an effort to maintain their respective languages (see Chapter 4), that is, in situations with strict norms (see the Vaupés below). In the latter situation, it is more likely that word forms are seen as emblematic of identity than are abstract organizational principles of constructions (Aikhenvald 2002), leading to convergence at more abstract levels.

In a few cases, we can base at least part of our understanding of the sociolinguistic circumstances that have led to linguistic areas on direct observation (although it must be stressed that there are no cases known to us where we can observe the entire process from discourse innovation to becoming part of the language system – it remains a puzzle with a good number of the pieces missing). Given the importance of the connection between linguistic areas and other subfields of contact linguistics, we will dwell somewhat longer on this topic, briefly discussing three cases in which we are fortunate to have direct observations of the sociolinguistic circumstances responsible for the areal convergence observed in the languages involved: the Vaupés (Brazil/Colombia border area), Kupwar (west-central India), and the Balkans (eastern Europe). These three cases are contrastive in that they illustrate different sociolinguistic circumstances that have all led to convergence between the languages involved. The Vaupés shows the effects of institutionalized exogamy and intensive exchange coupled with strict policies against linguistic mixing, leading to (mostly) unilateral diffusion of abstract structure; Kupwar shows the effects of male-driven labor-related bilingualism, leading to multilateral (though uneven) diffusion of both form and abstract structure; finally, the Balkans, apart from sharing some aspects with the other

two, in addition shows some evidence of language shifts, and displays shared innovations and a preference for analytical constructions.

3.1.1 The Vaupés

The Vaupés linguistic area (e.g., Aikhenvald 2002, 2003, Epps 2007) is situated along the Vaupés River basin on both sides of the Colombian-Brazilian border. Several ethnolinguistic groups are in contact that belong to one of three linguistic families: Tucanoan, Arawakan, and Nadahup (or Makú). Between the Tariana (Arawakan) and several Tucanoan groups, there is obligatory exogamy, which necessarily creates a situation where children are connected to two different languages. The Nadahup, traditionally semi-nomadic hunter-gatherers (unlike the agriculturalist Tariana and Tucanoan groups), have long-standing and active socio-economic relations with the Tucanoans of the area, but do not participate in intermarriage. Epps (2007: 268) describes the relationship of the Hup (the Nadahup group most involved in the contact zone) with the Tucanoan groups of the Vaupés as a patron-client relationship, in which the Hup are temporary servants and laborers, and engage in trade. The Nadahup family shows a pattern of involvement in the contact situation that clearly matches the geographical distance from the Vaupés: Hup > Yuhup > Dâw > Nadëb. Contact between the Tariana and the Nadahup groups is less common, creating a situation whereby two different groups (Tariana and Hup) are both oriented toward the same third group.

What makes the Vaupés particularly fascinating are the strict language policy against language mixing (going against Matras' general characterizations mentioned above) and the intricate rules for choosing a language in a particular context (Aikhenvald 2003). As a rule, one speaks the language of one's mother to one's mother and her relatives, and one's father's language to one's father, his relatives, and to one's siblings. Further, etiquette says it is polite to speak the language of one's guest, or of the majority of people present. Finally, power relations are expressed through language choice, for instance, if a speaker speaks his father's tongue in a situation where this is considered impolite, this is interpreted as an assertion of one's status. Language mixing (using phonetic material from more than one language) is considered a sign of linguistic incompetence. Even though the relations between the Hup-speakers and the Tucano were different than those of the Tariana, language loyalty in the sense of a strong resistance against language mixing was also part of the communicative ethics of the Hup-speakers (Epps 2007). This situation of multilingual interaction has been in place for at least a century (Epps 2007), probably longer (Aikhenvald 2002: 24).

As we will see below in Section 3.2, this situation results in mutual influence between the languages involved, though not in terms of particular lexemes or morphemes, but in terms of abstract structural-linguistic organization.

3.1.2 Kupwar

Kupwar (Gumperz & Wilson 1971) is a small village in west-central India where four languages from two families come together: Kannada and Telugu (both Dravidian), and Urdu and Marathi (both Indo-Aryan). Partly coinciding with these language groups, there are two major social groups: (1) Landowners and cultivators: Kannada-speaking *Jains* and Urdu-speaking Muslims; (2) *Lingayats*: Kannada-speaking craftsmen, Marathi-speaking untouchables and landless laborers.⁴

Bilingualism or multilingualism in Kupwar is the norm for local men, and probably goes back several centuries. Marathi has a special position as the main language of communication in inter-group communication. Marathi is also the dominant language in the geographical surroundings of Kupwar as well as the dominant literary language. Inter-group communication mainly takes place in work environments; private environments are mostly separated along linguistic and social lines. Gumperz & Wilson (1971) argue that the existence of clear niches for each language in the private sphere is the main reason that the languages involved persisted without succumbing to a more dominant one.

The association of each language with kinship and social closeness is also why people use different languages in the workplace when members of one of the socially dominant groups address people from other groups. Addressing them in Kannada or Urdu would imply that the laborers are socially close to the landowners. Since Marathi does not carry this connotation (it is not the home language of a large group of people) it is the preferred language for inter-group communication.

The situation of constant code-switching has led to some significant changes in the varieties of the languages involved, resulting in a high degree of intertranslatability. Gumperz & Wilson (1971: 155) go so far as to claim that the varieties used in multi-group situations share “a single syntactic surface structure.”

3.1.3 The Balkans

The main varieties contributing to the Balkan linguistic area are Albanian, Greek, Balkan Romance (Romanian, Aromanian, Megleno-Romanian), and Balkan Slavic

⁴There also is a small group of Telugu-speaking rope makers, but they do not surface in the linguistic analysis in Gumperz & Wilson (1971).

(Bulgarian, Macedonian, transitional varieties spoken in Serbia). The formation of the main shared features of the Balkan linguistic area, the so-called Balkanisms, are estimated to have taken place very roughly between the mid-first and mid-second millennium. Friedman & Joseph (2014) describe the language contact in the Balkans as intense, intimate, and sustained. Lindstedt (2000: 239) further lists the main components of the sociolinguistic situation in the Balkans after the Avar invasion in the 6th century and until as late as the 19th century and the demise of the Ottoman empire: 1) Speakers of different languages live close together, often in the same village. 2) There is no single dominant lingua franca. 3) Speakers of each language have sufficient access to the other languages they need. 4) Native languages are important symbols of group identity.

Most researchers seem to agree on the role that this type of balanced language contact played in the formation of the Balkan linguistic area. Some traces of the presumed historical multilingualism can be still observed, despite a drastic ethno-linguistic reshaping of the region during the 19th and 20th centuries and the creation of nation states with one dominant ethnic group, religion, and language (Makartsev & Wahlström 2016: 4–10). The evidence for the earlier community-level multilingual settings is often indirect, yet it must be assumed that, despite the wide-spread multilingualism, several social factors contributed to its level and intensity. Adult males are likely to have been more multilingual than females and children due to their higher mobility, for instance, as traders and seasonal workers. Also, exogamy between different linguistic communities but within the same religious group is likely to have produced gender- and group-specific patterns of multilingualism (see, e.g., Morozova & Rusakov 2018).

The lack of a lingua franca did not mean that all the contributing languages of the Balkan linguistic area enjoyed the same level of prestige. It has been suggested that occupying a mid-position on the prestige scale would explain why some varieties display more shared areal features. According to Lindstedt (2018), the varieties in the middle of the scale, Macedonian in particular, would have been more frequently exposed to the different types of interference from the other languages: there would be significant numbers of both L2 and L1 speakers for that variety. L2 speakers experience interference from their native languages, whereas L1 speakers also try to accommodate their own language, using structures that they know have analogues in the other languages (see also Joseph 2010: 624–628).

It is crucial to note that, despite the duration and intensity of language contacts and widespread multilingualism, the speech communities of the Balkans successfully maintained their linguistic identity for centuries (see, e.g., Joseph 2010: 625). A social norm that holds languages as important group symbols, in combination

with another norm that values multilingualism,⁵ has been an important factor contributing to the linguistic convergence. Yet at least in some cases the areal convergence may also have been sped up by language shifts. One such example is the South Slavic and Aromanian substrate in some varieties of Albanian (Desnickaja 1968: 258–260). Macedonian is also argued to have an Aromanian (Balkan Romance) substrate, having contributed significantly to its “Balkanization” (Gołąb 1997). Yet crucially for the Balkans, these proposed substrata cover only a small, although central, geographical portion of the linguistic area.

3.2 Patterns

The patterns in the focus of the study of linguistic areas range from articulatory commonalities through shared semantic concepts to even beyond the traditional scope of linguistics, such as gesturing (for the latter, see Enfield 2005: 196). Yet some general tendencies regarding the areas of interest can be observed. Trubetzkoy’s definition of linguistic areas (see Subsection 1.1), contrasting the areas with language families, emphasizes shared syntactic and morphological patterns, while also mentioning shared cultural words, as well as occasional, superficial similarities in sound systems. This early prediction is reflected in the choices made in later contributions to the field: morphosyntactic patterns dominate the analyses.

The superficiality of similarities pertaining to sound systems in Trubetzkoy’s (1928) formulation must not be interpreted as non-pervasiveness. These similarities may contribute to particular contrasts in the respective phonological systems of the constituent languages of a linguistic areas, although they are not describable as systematic sound correspondences as with genealogically related varieties. For instance, in the Indian subcontinent, most languages contrast a set of retroflex consonants with dentals (Emeneau 1956: 7). In the Balkans, in several varieties there is a stressed schwa-like (mid-to-high central) vowel (Asenova 2002: 28–30). Also, the preference for tonality in Southeast Asia results only partially from common inheritance, as Vietnamese demonstrates. It became, unlike the related Khmer, a tone language as a result of language contact (Enfield 2005: 191).⁶

The fact that lexicon, like phonology, has not received as much attention in the study of linguistic areas as morphosyntax is most likely because lexical borrowing can take place even in the most superficial language contacts, not subsuming

⁵This is evidenced, for instance, by the ubiquity of the proverb “languages are wealth” in the Balkans (Friedman 2012: 163).

⁶See Subsection 4.2 for potential substrate explanations.

sustained multilingualism. Loanwords do not, therefore, add much to an area hypothesis. Calquing, on the other hand, presupposes somewhat better knowledge of the source language, and lexical semantic convergence has been studied in connection to areality (see, e.g., Asenova 2002: 48–61 for shared proverbs and patterns of polysemy or colexification in the Balkans; Aikhenvald 2009 for polysemy in New Guinea and Australian languages). A novel approach to colexification and areality is proposed by Gast & Koptjevskaja-Tamm (2018). The authors automatically detect areal clustering of colexification patterns from lexical databases. The areally clustering patterns found (such as “fire”–“tree” or “mountain”–“stone”) are then controlled for genealogical relatedness to establish convergence areas.

Yet the existence of a shared structural innovation does not mean that its functions are uniform. This disparity may result from different levels of integration of the feature into a variety, or, in other terminology, its level of grammaticalization. All languages of the Balkan linguistic area display pronominal doubling of direct objects (DOs), yet in some varieties doubling is triggered by all definite direct objects, whereas in others its use depends on less frequent pragmatic contexts. Friedman (2008) establishes a cline from more to less grammaticalized use of the doubling of DOs: the further one moves away from the assumed geographical center of convergence, the less grammatical the feature becomes.

A further question involves the diachrony of the contact-induced convergence. From the perspective of a single language contributing to a linguistic area, a shared feature may be an innovation, but it could also represent the retention of a feature that would have been lost without the contact. Janhunen (2005: 117) presents a matrix of potential diachronic pathways of convergence, see Table 2.

Table 2: Diachronic pathways of convergence (Janhunen 2005: 117)

	Positive	Negative
Active	introduction of a new feature	loss of an old feature
Passive	retention of a feature	non-introduction of a feature

While a positive + active development could be a borrowed feature or a shared innovation, a negative + active development is illustrated by the loss of a feature or category that is not supported by the contacting languages. A positive + passive development, on the other hand, means the retention of a feature due to contact that would be lost otherwise. For instance, the Balkan Romance languages are the only Romance varieties retaining some case inflection, modeled according to a shared Balkan case system (Wahlström 2015: 107). The intuitively most difficult

scenario to grasp, a negative + passive development, would entail a situation in which the diffusion of an innovation within a dialect continuum is prevented in a variety, because the innovation is not supported by its contacting languages. However, while logically possible and credible as scenarios, retention, loss, and especially prevented adoption of a feature can be extremely hard to prove.

In order to show the relation between processes and potential resulting patterns, we briefly zoom in on the patterns found in the three case studies discussed in the previous section (Vaupés, Kupwar, the Balkans).

3.2.1 Vaupés: Unilateral diffusion of abstract patterns

The specific sociolinguistic policies of the Vaupés area have led to numerous shared abstract patterns with very little borrowing of forms from one language to another, across the entire spectrum of language structure (ranging from phonemic contrasts to morphosyntax and discourse structure, see Aikhenvald (2002) for an elaborate overview). Importantly, form (matter) borrowing and changes in word formation processes are hardly attested, both in Tariana and Hup. The contact-induced changes in Hup and Tariana can in many cases be established by comparing their structures to those of related languages that are not (or are less) involved in the exchange system of the Vaupés.

Because of the fact that both the Tariana and Hup speakers are oriented toward the Tucanoans, they have become similar to each other in spite of the lack of intensive contact between them. In other words: the Vaupés linguistic area arose through unilateral, partly parallel influence from Tucanoan into Hup and Tariana. Table 3 shows the contact-induced changes that are shared by (at least) Hup and Tariana.

3.2.2 Kupwar: Multilateral (but uneven) diffusion of matter and patterns

In their description of the contact-induced changes in the varieties involved in the contact situations, Gumperz & Wilson (1971) discuss several areas of grammar where the Kupwar languages, though maintaining their lexical differences, have converged. Changes can be established relatively clearly because all three languages are also spoken in areas that fall outside the Kupwar exchange system proper. Convergence seems mainly to have affected local Urdu, and Kannada to a lesser extent. Their observations are summarized in Table 4. Note the importance of negative categories: these illustrate Janhunen's (2005) negative + active pathway to areal convergence – the loss of an old feature – discussed in Section 3.2. For a more detailed account, the reader is referred to Gumperz & Wilson (1971).

Table 3: Contact-induced changes in the Vaupés languages shared by Hup and Tariana

Grammatical pattern	Affected languages
Nasal spread	Hup, Yuhup, Tariana
Pitch-accent	Hup, Yuhup, Dâw, Tariana
No phonemic oral-nasal contrast	Hup, Tariana
Root-initial allophone [dj] of /j/	Hup, Tariana
Allophone [r] of d in intervocalic position	Hup, Tariana
Mixed nominal classification ^a	Hup, Tariana
Base-five numeral system ^b	Hup, Tariana
Animacy-based split plurality	Hup, Tariana
Elaborate evidential system	Hup, Tariana
Remoteness distinctions in the past	Hup, Yuhup, Tariana
Productive verb-verb compounds	Hup, Tariana
Accusative alignment	Hup, Tariana
Multi-functional object marker	Hup, Tariana
Differential non-subject marking	Hup, Tariana
Morphological passive derivation	Hup, Tariana
Verb-final constituent order	Hup, Tariana
Affixed negation	Hup, Tariana
Inherently negative verb stems	Hup, Tariana

^aInanimates are classified according to shape, animates according to gender. In Hup, this classification system is used for a subset of nouns.

^bThis system is based on adding fingers and toes to five, and also includes a calqued term for number four meaning ‘having a brother’.

Apart from the more abstract patterns displayed in Table 4, there is also some borrowing of linguistic forms between the languages. Gumperz & Wilson (1971: 161) note that, although all elements can be borrowed, content words are borrowed more frequently than function words (and within the group of function words there is a sub-hierarchy: adverbs > conjunction markers > postpositions > other), and function words are in turn more frequently borrowed than morphologically bound material (where derivation < inflection).

Gumperz & Wilson (1971) conclude that the linguistic patterns that are the result of contact reflect the social and regional situation. Locally, the Kannada speakers form the majority and are the economically dominant group. Local

Table 4: Contact-induced changes in the Kupwar languages

Grammatical pattern	Affected languages	Source languages
Semantically-based gender system	Urdu, Marathi	Kannada
Single agreement in auxiliary constructions	Urdu	Kannada
Complement follows V+AUX complex	Urdu	Kannada, Marathi
Person-based agreement on past auxiliaries	Urdu	Kannada, Marathi
Mono-exponential future markers	Urdu	Kannada
No tense-based alignment split	Urdu	Kannada
No NP-internal agreement	Urdu	Kannada
Underived modifying elements in NP ^a	Kannada	Urdu, Marathi
Dative-marked human objects	Kannada	Urdu/Marathi
No special past non-finite construction for compound verbs ^b	Urdu	Kannada, Marathi
Use of a copula in equative constructions	Kannada	Urdu, Marathi
Order communication verb - quotation	Kannada	Urdu, Marathi
Clause-final question word marks Y/N question	Kannada, Urdu	Marathi
Dative/oblique case to mark purpose clause	Kannada	Urdu, Marathi
Genitive on verbs creates modifiers	Urdu	Kannada
Clusivity	Urdu	Kannada

^aKannada has lost the obligatory nominalizing suffixes on pronominally and adjectivally used possessives and demonstratives, like Hindi-Urdu and non-local Marathi.

^bLocal Urdu used to have an exceptional construction for compound verbs in past non-finite contexts, but has extended the use of so-called past non-finite verb forms to include compound verbs, following Kannada and Marathi.

Marathi, in spite of its speakers belonging to a subordinate social group, is protected by the fact that it is the dominant language in the larger region. This leaves local Urdu as the language that has undergone the most changes. The authors also suggest that the resulting intertranslatability is driven by cognitive economy, reducing the need to learn different systems. A further important component, as mentioned, is the fact that each language has its own clear niche in which it is used.

3.2.3 The Balkans: Shared innovations and preference for analytical constructions

The grammatical features shared by the Balkan languages are characteristically constructions consisting of uninflected function words as well as of inflected words that have grammaticalized, for instance, into auxiliaries and articles. In diachronic terms, these so-called Balkanisms represent increasing analyticism, since they replace or compete with strategies relying on inflection. Moreover, the grammatical Balkanisms typically do not give away their origin, but seem to be shared innovations. Judging this is, nevertheless, problematic, since only Greek and Slavic are attested throughout the formation of the linguistic area, whereas the first written sources for Albanian and Balkan Romance surface only by the mid-second millennium, already displaying the grammatical hallmarks of the linguistic area. On the other hand, Balkan Romance, as well as Slavic, has members of the same branch of Indo-European outside the geographical area with a better attested history, allowing for comparison.

Table 5 displays some of the most commonly cited areal features of the Balkans. While shared, for the most part, by all main contributing languages or a number of their varieties, the majority of these are also present to some extent in the varieties of Romani spoken in the Balkans.

However, not all these features form clear-cut, perfectly overlapping isoglosses, delineating the linguistic Balkans. Their integration into the linguistic systems of the respective languages can be assessed along three axes:

1. The Balkan feature may only be a minor pattern in competition with other structures. For instance, while the majority of Greek dialects display the genitive-dative merger, when marking recipients the genitive-dative is in competition with a prepositional construction, pertaining to the other Balkan tendency of favoring analytical constructions.
2. While a certain morphosyntactic pattern may exist, its level of integration into the grammar may vary. Object reduplication is found to some

Table 5: Balkan morphosyntactic features present in the majority of the languages (adapted from Lindstedt 2000)

Grammatical pattern
Enclitic articles
Object reduplication
Prepositions instead of cases
Dative–possessive merger
Goal–location merger
<i>Relativum generale</i>
Aux (+ comp) + finite verb
<i>Volo</i> future
Past future as conditional
<i>Habeo</i> perfect
Evidentials
Analytic comparison

extent in all languages, but its conditioning criteria vary from being obligatory with, for instance, definite and specific objects to being pragmatically conditioned toward the periphery of the Balkan linguistic area (Friedman 2008).

3. The dialectal spread of a feature varies. A Balkan feature may be present in a number of spoken varieties, often in closer proximity to the areas of more intense contacts, but absent in the prestige varieties and the standard language, typically described by reference grammars.

Shared phonological features of the Balkan linguistic area are typically seen as more tenuous than the morphosyntactic. According to Friedman (2011: 278), this is not because there is no convergence on the level of sound systems, but that it is often highly localized. This local convergence attests to intense contacts, but with no all-encompassing, unified results. The best-known candidate for a phonological Balkan feature is a stressed schwa-like (mid-to-high central) vowel, famously lacking in Greek and Macedonian. However, a closer scrutiny reveals that modern Macedonian dialects, not displaying the phoneme, have lost it as a result of phoneme mergers (Koneski 1965: 46–48). Since the phonological convergence of the Balkan languages did not involve all varieties, or, since its results were not necessarily permanent, it is not an area-defining property on par with the shared morphosyntactic features.

4 Factors

In this chapter we introduce factors that have been considered relevant for linguistic area formation. We begin with linguistic factors followed by more numerous non-linguistic factors.

4.1 Linguistic factors

Three types of linguistic factors contributing to area formation have been discussed in the literature. We summarize these as (1) borrowability, (2) typological distance, and (3) complexity.

4.1.1 Borrowability

Linguists interested in the historical development of languages have long noted that not all elements of a language change at the same rate. This is an important insight in historical linguistics where the lexicon is concerned, leading to a list of stable words (Swadesh 1952 and subsequent publications), which proved to be of great importance for establishing relatedness between languages. About the same time as Swadesh introduced his list of stable words, Weinreich (1953) published a similar idea relating to borrowability, making generalizations about the borrowability of linguistic items, including structural elements. Table 6 shows the main conclusions by Weinreich (1953): in terms of formal properties, he concluded that tightly bound forms are less likely to transfer from one language to another than free forms. Likewise, reduced forms and morphemes that show allomorphy are less likely to be transferred. On the content-related side, he deemed obligatory, grammatical, and non-affective forms to be less borrowable than their counterparts.

Table 6: Weinreich’s borrowability generalizations, adapted from Wilkins (1996)

Formal properties			Content-related properties			BORROWABILITY
Boundedness	Form	Variance	Use	Function	Meaning	
tight	reduced	flexive	obligatory	grammatical	non-affective	LOW
free	robust	non-flexive	optional	lexical	affective	↑ HIGH

Weinreich’s seminal work gave rise to many different proposals for borrowing hierarchies (see, e.g., Wilkins 1996 and Curnow 2002: 417–419 for useful

overviews). Although none of these hierarchies has full predictive or implicational power, they still highlight fairly robust tendencies:

1. Content words are generally more easily borrowed than function words.
2. Nouns are generally more easily borrowed than other words.
3. Free elements are generally more easily borrowed than bound elements.
4. Derivational morphology is generally more easily borrowed than inflectional morphology.

These generalizations are, by and large, consistent with many lexical borrowing patterns (see Haspelmath & Tadmor 2009 for the first two generalizations and Gardani et al. 2017 for the last two; see also the lexical borrowing patterns in the Kupwar case described above).

These tendencies, however, apply to the borrowing of formal elements from language A into language B. As we have seen in the three case studies above, contact-induced change in linguistic areas often (perhaps always) involves the borrowing or convergence of abstract patterns. As has been shown by, for instance, Matras & Sakel (2007), borrowing of form (matter in their terminology) is subject to different generalizations than borrowing of patterns. Attempts to come to a borrowing scale for abstract grammatical features have come mainly from typological research, and mainly following the online publication of the World Atlas of Language Structures (Dryer & Haspelmath 2013), in the form of stability measures, which often also include a measure of borrowability. Unfortunately, these proposals, in spite of some promising tendencies across methods (Dediu & Cysouw 2013), have not led to widespread consensus.⁷

4.1.2 Typological distance

It is often assumed that structural similarity between languages facilitates the transfer of linguistic material from one language to another (see, e.g., Bovern 2013 and Seifart 2015 and references therein).⁸ This assumption resonates with

⁷In fact, the methods seem to be more successful in determining stability than borrowability. As Dediu & Cysouw (2013: 18) show, not all of the methods they review actually measure borrowability very clearly, and since the effects of language contact are not uniform across social situations, borrowability is hard to measure from a global typological data set.

⁸Note that an opposite line of reasoning is also sometimes mentioned: elements that are “lacking” in language A are more likely to be transferred from language B, in order to fill a functional gap in language A. Since functional gaps are difficult to define (though see Hale 1975 for an attempt), this issue is rarely systematically pursued.

findings in code-switching research, where repeated observations that speakers will switch at points where the languages involved are structurally similar have led to the postulation of the so-called *equivalence constraint* (Poplack 1978, 1980).⁹ Nevertheless, several scholars (Thomason & Kaufman 1988, Bowerman 2013, Seifart 2015) show or argue that structural similarity plays a role in the background at most, secondary to social factors. The most systematic treatment of this question is arguably Seifart (2015), who looked at affix borrowing in 101 pairs of languages, and related this to typological similarity, based on data from Dryer & Haspelmath (2013). He found no correlation between typological similarity and concludes that “structural-typological similarity plays at best a minor role in constraining or facilitating the borrowability of affixes.” Nevertheless, more detailed views (construction-based) on typological or structural similarity are needed to get a clearer view on this question. Also, structural distance as a factor has mainly been taken into consideration with respect to matter borrowing; how pattern borrowing behaves with respect to this parameter is unclear.

4.1.3 Complexity

A controversial topic in contact linguistics is whether language contact leads to simplification in the languages involved. Following up on earlier work, Trudgill (2004) argues, on the basis of phoneme inventories of Austronesian languages, that certain contact situations lead to simplification. Specifically, he claims that “communities involved in large amounts of language contact, to the extent that this is contact between adolescents and adults who are beyond the critical threshold for language acquisition, are likely to demonstrate linguistic pidginization, including simplification, as a result of imperfect language learning” (Trudgill 2004: 306).¹⁰ This claim finds partial support in a large-scale statistical analysis of the morphological complexity and demographic aspects (Lupyan & Dale 2010) of about 2,000 languages, which found a correlation between speech community size and morphological simplicity. Trudgill argues further that a different type of contact situation can lead to *complexification*: a situation with widespread native bilingualism can lead to the expansion of phoneme inventories to allow for pronunciations of loanwords.

⁹The equivalence constraint has also turned out to be a tendency at best, since several cases of violations have been recorded since its first proposal (Matras 2009: 130)

¹⁰In his phoneme inventory data, such societies tend to have medium-sized phoneme inventories, which minimize complexity in phoneme size on the one hand, such that distinctiveness in the lexicon is maintained.

If Trudgill is right, we can consider complexity as a factor in the sense that complex forms are prime targets of contact-induced change, given the right sociolinguistic circumstances. And indeed, especially if we look at the contact-induced changes in the Kupwar languages discussed in Section 3.2, where much adult second language learning presumably takes place (see Section 3.1), they often seem to involve simplifications (e.g., loss of a tense-based alignment split, loss of multi-exponence, loss of double agreement, loss of intransparencies in the gender system). The Vaupés, on the other hand, where native bilingualism is ubiquitous, seems to involve fewer cases of simplification; indeed they often involve complexification (e.g., creation of allophony, a mixed nominal classification system, remoteness distinctions in the past, elaborate evidential system). Regarding the Balkans, Lindstedt (2018) argues that the bulk of the Balkan contact phenomena are neither clearly simplifying nor complexifying in the sense of Trudgill, but represent a third type that favors explicit analytic grammatical marking that increases direct intertranslatability among the languages.

Nevertheless, Trudgill's claims are controversial (see the debates in the same issue of *Linguistic Typology* in which Trudgill (2004) was published as a target article), and they suffer from two serious shortcomings. The first is that we have little data on historical social situations (see Section 3.1 above), which makes the claim hard to falsify, and the second is that, in spite of a sustained interest in the topic (see e.g. Kusters 2003, Dahl 2004, Miestamo et al. 2008, Baerman et al. 2015), there is no generally agreed-upon theory or even description of complexity.

4.2 Non-linguistic factors

Since non-linguistic factors often have to be reconstructed from the linguistic data in areal linguistics, there is not much firm ground to stand on. This should be kept in mind when reading the following sections.

4.2.1 Speaker-related factors

Very little can be said about speaker-related factors, but we highlight two factors that may play a role in area formation: gender and age.

Gender may play a role in linguistic areas in that, depending on the situation, it may be mainly males or females who promote convergence, for instance when linguistic areas are the result of institutionalized exogamy combined with sex-based post-marital residence patterns. In these cases, it is the mobile group (i.e. the women in patrilocal systems, the men in matrilocal systems) that may drive language change in the societies they are married into (see, e.g., Morozova &

Rusakov 2018, mentioned in Section 3.1). This has been one of the relevant factors in the Vaupés area, especially in Arawakan-Tukanoan contacts (see Aikhenvald 2002, Chacon 2017, Epps 2020, Van Gijn et al. 2022).

Convergence may also be driven by sex-based labor divisions. If, for instance, trade or work migrations contribute to area formation, and are typically male activities (as in the Kupwar case, for instance), males might be the driving force behind convergence. In the case of the Balkans, seasonal work migrations in the Ottoman Empire, called *gurbet*, are considered to be one of the contributing contexts for adult multilingualism (Lindstedt 2018). However, they were predominantly a male activity, with the exception of young, unmarried women (Hristov 2008: 3–5).

Age may be important in that interference from a first language into a second language is more likely if a person learns the second language at a later age. Native, or infant bilingualism, on the other hand, may leave fewer (and different) traces in the languages involved. L2 effects have played a role in the area formation in the Balkans and in the Kupwar case, while early multilingualism seems to have been the standard case for the Tariana in the Vaupés. More generally, it is assumed that adult L2 learning will more often lead to simplification, whereas infant bilingualism is more likely to preserve the intricacies of the linguistic systems involved intact. This latter point is illustrated, for example, by Mithun (2015) for the North American context, where the widespread occurrence of complex bound pronominal systems and large inventories of lexical-like affixes (i.e. affixes with a high semantic load) can partly be attributed to the long-term existence of infant bilingualism and the absence of large-scale adult L2 learning.

4.2.2 Interactional factors

For most linguistic areas, we do not know anything about the conversation practices of the people involved, and so the only clear assumption we can make here is that conversations contain elements from both (or more) languages involved. What exactly these “foreign” elements are depends on the societal norms of language use (see Subsection 4.2.3), but it seems that one of the things that many linguistic areas have in common is the fact that, contrary to “normal” situations of language contact (as for instance portrayed in Thomason & Kaufman 1988), the contact signal in linguistic areas seems to be found in the patterns of grammar.¹¹ It therefore seems reasonable to assume that conversations between speakers

¹¹This is not to say that there can be, or indeed is, no matter borrowing in areal patterns, but most linguistic areas do seem to deviate from the pattern that grammar is affected by contact only after the lexicon has clearly been affected, as argued for in Thomason & Kaufman (1988).

from different ethnolinguistic backgrounds were mixed at more abstract levels of organization (see for this point Matras 2011, discussed in Section 3.1).

4.2.3 (Supra)-societal factors

One reasonable perspective on linguistic areas, present in the view of e.g. Campbell (2006) and Matras (2011), is that they simply represent networks of language communities in contact with each other under similar circumstances, with either one language in contact with several others, or a chain of contacts between geographically neighboring languages. On this view, the question of what societal factors are important for the formation of a contact area is partly a question about what shared or suprasocietal factors are conducive to area formation. Without claiming exhaustivity, the most important (supra)societal factors we see in the literature as facilitating area formation are the following:

4.2.3.1 An incentive for contact

At the very least, there must be a reason for two or more societies to get into contact with each other. There may be many different reasons for contact (trade, expansion/war, broadening the gene pool), which nevertheless can arguably be summarized as increasing the standard of living of at least one of the groups involved; often it is a necessary step for a society to come into contact with another group in order to avoid food shortage (see Nettle 1999, which will be briefly discussed in Section 4.2.4 below).

4.2.3.2 Widespread, intense, and long-term bilingualism or multilingualism

Another necessary societal circumstance is that there is a large portion of people involved in the areal pattern who speak more than one language. The scenario that is often associated with area formation is one of symmetrical, mutual bilingualism (Muysken 2010). In addition, there are other logically possible contact scenarios, resulting in areal convergence. A shared superstrate language could result in areal patterning of linguistic features without horizontal diffusion between the converging varieties. However, much more popular than superstrate explanations are theories about a common substrate language, especially in the past. Yet proving that a particular feature, especially a structural one, results from a substrate is notoriously difficult, and their former popularity may be explained by substrate being perhaps the earliest well-studied type of language contact (see,

e.g., Wahlström 2015: 14–16). Also, it is not clear whether the expected linguistic outcome of extensive adult multilingualism and language shift are necessarily different. Language shift (see Chapter 4) is typically assumed to have been preceded by a period of adult multilingualism. Nevertheless, a number of uncontested features of well-known linguistic areas may well have arisen through the substrate effect rather than by sustained multilingualism. For instance, the retroflex consonants typical of the linguistic area of the Indian subcontinent have been attributed to a Dravidian substrate (Thomason & Kaufman 1988: 39–40).

4.2.3.3 Language loyalty at the societal level

The other side of the coin of multilingualism is that linguistic areas arise in circumstances of predominant language maintenance, i.e. there must be incentives for a group of speakers that is involved in a situation of sustained interethnic contact to maintain their own language. These incentives can be functional-communicative (particular niches exist for the language so that it must be learned by children, e.g. as in the Kupwar case) and/or ideological (as in the Vaupés). This perhaps seems a trivial point to make, but it is a point that is not always made explicit.

4.2.3.4 Lax norms of language use

As was briefly discussed in Section 3.1, in order for innovations to spread through a society, there needs to be some amount of tolerance in the norms of language use. It seems that the type of innovation that can spread through society is determined by these norms of language use. For instance, in the Vaupés, where norms of language use are very strict when it comes to mixing of linguistic matter, permitted variation is limited to more abstract patterns of language use. In the Kupwar situation there seems to be more tolerance for using matter from language B in a language A context. In the Balkans, there seems to be tolerance for the use of innovative, analytical constructions, perhaps related to adult bilingualism or imperfect competence in non-native languages.

4.2.4 Geographical factors

In spite of its obvious connections to geography, geographical factors (beyond the very general “geographical proximity”) are not commonly taken into consideration in areal linguistics. However, if we extend our view to approach the more general topic of migration and expansion, we can bring up a number of suggestions for geographical factors that may influence area formation.

Elevation is often regarded as a barrier to language contact, linguistic diffusion, and language expansion (e.g. Nichols 1992, 1997). We mentioned Predictive Areality Theory (Bickel & Nichols 2006), where mountain ranges are taken into consideration as topographical elements impeding spread, and thus as boundaries to the linguistic area. They find that these mountain ranges are indeed good indicators of the limits of the large contact zone (the Circum-Pacific) under consideration. Elevation differences are also taken into account in Van Gijn (2014) and Van Gijn & Muysken (2020), whose results suggests an important role for societies on the mountain slopes in the diffusion of features across elevation differences in the Andean-Amazonian context.

Bodies of water are another factor that has been contemplated in areal linguistics. They have been considered both barriers and facilitators of contact. It has long been claimed that river systems have functioned as important pathways for language spreads and contact in Amazonia (see e.g. Hornborg 2005, Eriksen 2011). Van Gijn et al. (2017) set out to test the role of rivers in the formation of linguistic areas in lowland South America. The results do not clearly support a facilitating role for the river systems, but this may be due to methodological issues, so this needs to be investigated further.

A case in which water seems to function as a separator of sorts is in the historical development of the languages of the British Isles and those in their immediate surroundings (Dedio et al. 2019). In their study, which focuses on reflexivity marking, the authors show that, over the period between 1200 and 1900 in particular, the languages spoken on the isles have become more similar to each other, and that the differences between the languages on the isles, on the one hand, and the languages on the continent, on the other, have grown. The continental languages, finally, have exhibited no significant change in their similarity. This means that the Channel and North Sea have acted, at least in that time period, as a barrier to contact.

Traveling time has been argued to be a superior measure to straight-line distance in explaining dialectal diversity (Gooskens 2005), see also Chapter 6. While (historical) traveling time can be seen as a proxy measure subsuming topography, it is not clear whether it could be feasibly applied to larger linguistic areas, as the effect of traveling time is typically assumed to reflect distances covered regularly by individuals. However, traveling time could be a useful factor in measuring diversity within more compact linguistic areas.

Ecological circumstances are considered in connection to linguistic diversity in Nettle (1999), who sets up a hypothesis that linguistic diversity is driven by the ecological circumstances. In this view, the behavior of speech communities is heavily influenced by “ecological risk” (the risk of food shortage for a given

community). In areas where resources are scarce and agricultural produce inconsistent, people are forced to forge extensive social bonds with other people, in order to increase the land they can draw food from. High-resource areas with constantly favorable climatic conditions allow smaller social groups to provide everything a society needs. Contact areas, in this view, can be the result of a strategy to reduce ecological risk.

Güldemann (2011), considering large contact areas in Africa, claims that geography has played an important role in that areas tend to spread in a general east-west direction (as opposed to north-south). Building on an idea by Diamond (1997), Güldemann connects this tendency to ecological circumstances in that the general climatic and ecological circumstances remain relatively constant in a east-west direction, whereas these circumstances tend to change more from north to south.

5 Conclusions

Linguistic area research, in spite of a number of fundamental difficulties, has enjoyed increasing attention within linguistics in general, and contact linguistics in particular. This has led to numerous proposals for putative linguistic areas all over the world. This steep increase in data over many different contexts, however, has not led to clear conclusions on the processes that lead to areal patterns of similarity. This is undoubtedly due to the fact that linguistic areas can arise under many different circumstances: areal patterns can arise in situations of symmetrical bilingualism, but also in situations where there is a dominant lingua franca; it can arise in situations where societies cooperate to improve their standard of living, but also in circumstances where one group dominates the other; it can involve infant bilingualism or adult bilingualism; it can involve few groups in a small territory, or many groups in a vast territory, to name but a few parameters on which situations can differ.

In order to press forward, therefore, it is vital in each proposed case of a linguistic area to know as much about the social history of the area as we can. In Campbell's (1985) terms: we need to maximize historicism and minimize circumstantialism. Unfortunately, history, in many cases, is unknown. But rather than ignoring the cases where we do not have a historical basis, we need to reconstruct social history as well as we can by taking recourse to other disciplines, such as human ecology, population genetics, archaeology and ethnology (see Van Gijn 2020). This is a big challenge, because it means that we need to build theories of how we can bring the signals from these different sources together in meaningful ways to create a picture that is maximally complete.

In addition, in line with the purpose of this book, linguistic area research would profit from being more intimately integrated within contact linguistics, seeking more systematic comparisons with other subdisciplines of language contact. Following ideas presented in, for instance, Niedzielski & Giles (1996) and Muysken (2013), the study of patterns in bilingual language use (see Chapters 2 and 3) provides an interesting comparative perspective with the areal patterns found in linguistic areas. Society-level studies of language shift (Chapter 4), borrowing (especially pattern borrowing), and the emergence of contact languages (Chapter 5) form a potential bridge between language use patterns on the one hand and deep-time areal patterns on the other (see Muysken 2008). Finally, for the areal aspect, we mentioned that geographical models are hardly used in linguistic area research. In this sense, linguistic area research can profit from the rich tradition of geographical modeling that exists in dialectology (Chapter 6).

Acknowledgements

Van Gijn gratefully acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 818854 - SAPPHIRE).

References

- Aikhenvald, Alexandra Y. 2002. *Language contact in Amazonia*. Oxford: Oxford University Press.
- Aikhenvald, Alexandra Y. 2003. Multilingualism and ethnic stereotypes: The Tariana of northwest Amazonia. *Language in Society* 32. 1–21.
- Aikhenvald, Alexandra Y. 2009. 'Eating', 'drinking' and 'smoking': A generic verb and its semantics in Manambu. In John Newman (ed.), *The linguistics of eating and drinking*, 91–108. Amsterdam & Philadelphia: John Benjamins Pub. Co.
- Asenova, Petja. 2002. *Balkansko ezikoznanie: Osnovni problemi na balkanskija ezikov sąjuz*. 2nd edn. Veliko Tärnovo: Faber.
- Baerman, Matthew, Dunstan Brown & Greville G. Corbett (eds.). 2015. *Understanding and measuring morphological complexity*. Oxford: Oxford University Press.
- Bell, Allan. 1978. Language samples. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language, vol. 1: Method and theory*, 123–156. Stanford: Stanford University Press.

- Bereznak, Catherine. 1995. *The Pueblo region as a linguistic area: Diffusion among the indigenous languages of the southwest United States*. Louisiana State University. (Doctoral dissertation).
- Bickel, Balthasar. 2017. Areas and universals. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 40–54. Cambridge: Cambridge University Press.
- Bickel, Balthasar & Johanna Nichols. 2006. Oceania, the Pacific Rim, and the theory of linguistic areas. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, 3–15.
- Boas, Franz. 1911. *Handbook of American Indian languages*, vol. 1. Washington: Smithsonian Institution, Bureau of American Ethnology.
- Boas, Franz. 1929. Classification of American Indian languages. *Language* 5(1). 1–7.
- Boduen-de-Kurtenè, Ivan A. 1901. O směšannom charakterě vsěh jazykov. *Žurnal Ministerstva narodnago prosvěščenija* 337. 12–24.
- Bowern, Claire. 2013. Relatedness as a factor in language contact. *Journal of Language Contact* 6. 411–432. DOI: 10.1163/19552629-00602010.
- Campbell, Lyle. 1985. Papers from the 6th international conference on historical linguistics. In Jacek Fisiak (ed.), *Areal linguistics and its implications for historical linguistics*, 25–56. Amsterdam: Benjamins.
- Campbell, Lyle. 2006. Areal linguistics: A closer scrutiny. In Yaron Matras, April McMahon & Nigel Vincent (eds.), *Linguistic areas: Convergence in historical and typological perspective*, 1–31. Basingstoke: Palgrave MacMillan.
- Campbell, Lyle. 2017. Why is it so hard to define a linguistic area? In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 19–39. Cambridge: Cambridge University Press.
- Chacon, Thiago. 2017. Arawakan and Tukanoan contacts in Northwest Amazonian prehistory. *PAPIA: Revista Brasileira de Estudos Crioulos e Similares* 27. 237–265.
- Curnow, Timothy J. 2002. What language features can be borrowed? In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, 412–436. Oxford: Oxford University Press.
- Dahl, Östen. 2001. Principles of areal typology. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals*, vol. 2, 1456–1470. Berlin: Mouton de Gruyter.
- Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: Benjamins.

- Dedio, Stefan, Peter Ranacher & Paul Widmer. 2019. Evidence for Britain and Ireland as a linguistic area. *Language* 95(3). 498–522. DOI: 10.1353/lan.2019.0054.
- Dediu, Dan & Michael Cysouw. 2013. Some structural aspects of language are more stable than theirs: A comparison of seven methods. *PLoS ONE* 8(1). 1–20. DOI: 10.1371/journal.pone.0055009.
- Desnickaja, Agnija V. 1968. *Albanskij jazyk i ego dialekty*. Leningrad: Nauka.
- Diamond, Jared M. 1997. *Guns, germs, and steel: The fates of human societies*. New York/London: W. W. Norton.
- Dixon, R. M. W. 2002. *Australian languages*. Cambridge: Cambridge University Press.
- Donohue, Mark, Søren Wichmann & Mihai Albu. 2008. Typology, areality, and diffusion. *Oceanic Linguistics* 47(1). 223–232.
- Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13. 257–292.
- Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *WALS online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/>.
- Dunn, Michael, Robert Foley, Stephen Levinson, Ger Reesink & Angela Terrill. 2007. Statistical reasoning in the evaluation of typological diversity in Island Melanesia. *Oceanic Linguistics* 46(2). 388–403.
- Emeneau, Murray Barnson. 1956. India as a linguistic area. *Language* 32(1). 3–16.
- Enfield, Nick J. 2005. Areal linguistics and mainland Southeast Asia. *Annual Review of Anthropology* 34. 181–206.
- Epps, Patience. 2007. The vaupés melting pot: Tucanoan influence on Hup. In Alexandra Aikhenvald & R. M. W. Dixon (eds.), *Grammars in contact*, 267–289. Oxford: Oxford University Press.
- Epps, Patience. 2020. Amazonian linguistic diversity and its sociocultural correlates. In Mily Crevels & Pieter Muysken (eds.), *Language dispersal, diversification, and contact*, 275–290. Oxford: Oxford University Press.
- Eriksen, Love. 2011. *Nature and culture in prehistoric Amazonia using G.I.S. to reconstruct ancient ethnogenetic processes from archaeology, linguistics, geography, and ethnohistory*. Lund University. (Doctoral dissertation).
- Evans, Nicholas. 2005. Australian languages reconsidered: A review of Dixon (2002). *Oceanic Linguistics* 44(1). 242–286.
- Friedman, Victor A. 2008. Balkan object reduplication in areal and dialectological perspective. In Dalina Kallulli & Liliane Tasmowski (eds.), *Balkan object reduplication in areal and dialectological perspective* (Clitic doubling in the Balkan languages), 35–63. Amsterdam: Benjamins.

- Friedman, Victor A. 2011. The Balkan languages and Balkan linguistics. *Annual Review of Anthropology* 40. 275–291.
- Friedman, Victor A. 2012. Languages are wealth: The Sprachbund as linguistic capital. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 38. 163–177.
- Friedman, Victor A. & Brian D. Joseph. 2014. Lessons from Judezmo about the Balkan Sprachbund and contact linguistics. *International Journal of the Sociology of Language* 2014(226). 3–23.
- Gardani, Francesco, Peter Arkadiev & Nino Amiridze (eds.). 2017. *Borrowed morphology*. Berlin, Boston: De Gruyter Mouton.
- Gast, Volker & Maria Koptjevskaja-Tamm. 2018. The areal factor in lexical typology: Some evidence from lexical databases. In Daniël Van Olmen, Tanja Mortelmans & Frank Brisard (eds.), *Aspects of linguistic variation*, 43–81. Berlin: Mouton de Gruyter.
- Gołąb, Zbigniew. 1956. The conception of "isogrammatism". *Biuletyn polskiego towarzystwa językoznawczego* 15. 3–12.
- Gołąb, Zbigniew. 1997. The ethnic background and internal linguistic mechanism of the so-called balkanization of Macedonian. *Balkanistica* 10. 13–19.
- Gooskens, Charlotte. 2005. Traveling time as a predictor of linguistic distance. *Dialectologia et Geolinguistica* 13. 38–62.
- Güldemann, Tom. 2011. Sprachraum and geography: Linguistic macro-areas in Africa. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation: Language mapping*, vol. 2, 561–585. Berlin: Mouton de Gruyter.
- Gumperz, John J. & Robert Wilson. 1971. Convergence and creolization: A case from the Indo-Aryan/Dravidian border in India. In Dell Hymes (ed.), *Pidginization and creolization of languages: Proceedings of a conference held at the University of the West Indies, Mona, Jamaica, April, 1968*, 151–165. Cambridge: Cambridge University Press.
- Hale, Kenneth L. 1975. Gaps in grammar and culture. In M. Dale Kinkade, Kenneth L. Hale & Oswald Werner (eds.), *Linguistics and anthropology: In honor of C. F. Voegelin*, 295–315. Lisse: The Peter de Ridder Press.
- Hall, Rober A. 1946. Bartoli's 'Neolinguistica'. *Language* 22(4). 273–283.
- Hammarström, Harald & Mark Donohue. 2014. Some principles on the use of macro-areas in typological comparison. *Language Dynamics and Change* 4. 167–187.
- Haspelmath, Martin & Uri Tadmor. 2009. *Loanwords in the world's languages: A comparative handbook*. 1st edn. Berlin: De Gruyter Mouton.

- Hornborg, Alf. 2005. Ethnogenesis, regional integration, and ecology in prehistoric Amazonia: Toward a system perspective. *Current Anthropology* 46(4). 589–620.
- Hristov, Petko. 2008. Mobilités du travail (gurbet), stratégies sociales et familiales: Une étude de cas dans les balkans centraux. *Balkanologie. Revue d'études pluridisciplinaires* 11(1–2). 1–16.
- Jakobson, Roman. 1962. K karakteristike evrazijskogo jazykovogo sojuza. In Roman Jakobson, Martha Taylor & Rudy Stephen (eds.), *Selected writings*, vol. 1, 144–201. The Hague: Mouton & co.
- Janhunen, Juha. 2005. The role of the Turkic languages in the Amdo Sprachbund. *Studia Turcologica Cracoviensia* 10. 113–122.
- Joseph, Brian D. 2010. Language contact in the Balkans. In Raymond Hickey (ed.), *The handbook of language contact*, 618–633. London: Wiley-Blackwell.
- Koneski, Blaže. 1965. *Istorija na makedonskiot jazik*. Skopje: Kočo Racin.
- Kurath, Hans. 1972. *Studies in area linguistics*. Indiana University Press.
- Kusters, Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflection*. Utrecht. (Doctoral dissertation).
- Lindstedt, Jouko. 2000. Linguistic balkanization: Contact-induced change by mutual reinforcement. In Dicky Gilbers, John Nerbonne & Jos Schaeken (eds.), *Linguistic balkanization: Contact-induced change by mutual reinforcement* (Languages in contact), 231–246. Amsterdam: Rodopi.
- Lindstedt, Jouko. 2018. Diachronic regularities explaining the tendency towards explicit analytic marking in Balkan syntax. In Joseph Brian & Krapova Iliyana (eds.), *Balkan syntax and universal principles of grammar*, 70–84. Berlin: De Gruyter Mouton.
- Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PLoS ONE* 5. 1–10. DOI: 10.1371/journal.pone.0008559.
- Makartsev, Maxim & Max Wahlström (eds.). 2016. *In search of the center and periphery: Linguistic attitudes, minorities, and landscapes in the Central balkans* (Slavica Helsingiensia 49). Helsinki: University of Helsinki.
- Masica, Colin. 2001. The definition and significance of linguistic areas: Methods, pitfalls, and possibilities (with special reference to the validity of South Asia as a linguistic area). In Peri Bhaskararao & Karumuri Venkata Subbarao (eds.), *Tokyo symposium on South Asian languages: Contact, convergence, and typology*, 205–267. New Delhi: Sage Publications.
- Matras, Yaron. 1998. Convergent development, grammaticalization, and the problem of ‘mutual isomorphism’. In Winfried Boeder, Cristoph Schroeder, Karl H. Wagner & Wolfgang Wildgen (eds.), *Sprache in Raum und Zeit: In memoriam Johannes Bechert*, 89–103. Tübingen: Gunter Narr Verlag.

- Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.
- Matras, Yaron. 2011. Explaining convergence and the formation of linguistic areas. In Hieda Osamu, Christa König & Hiroshi Nakagawa (eds.), *Geographical typology and linguistic areas: With special reference to Africa*, 143–160. Amsterdam: John Benjamins.
- Matras, Yaron & Jeanette Sakel. 2007. Investigating the mechanisms of pattern replication in language convergence. *Studies in Language* 31(4). 829–865.
- Miestamo, Matti, Kaius Sinnemäki & Fred Karlsson (eds.). 2008. *Language complexity: Typology, contact, change*. Amsterdam: John Benjamins.
- Mithun, Marianne. 2015. Morphological complexity and language contact in languages indigenous to North America. *Linguistic Discovery* 13(2). 37–59.
- Mithun, Marianne. 2017. Native north american languages. In Raymond Hickey (ed.), *The cambridge handbook of areal linguistics*, 878–933. Cambridge: Cambridge University Press.
- Moravcsik, Edith. 1975. Verb borrowing. *Wiener Linguistische Gazette* 8. 3–30.
- Morozova, Maria & Alexander Rusakov. 2018. Černogorsko-albanskoe jazykovoe pogranič'e: V poiskah "sbalansirovannogo jazykovogo kontakta". *Slověne* 7(2). 258–302. DOI: 10.31168/2305-6754.2018.7.2.10.
- Muysken, Pieter. 2008. Conceptual and methodological issues in areal linguistics. In Pieter Muysken (ed.), *From linguistic areas to areal linguistics*, 1–23. Amsterdam: John Benjamins.
- Muysken, Pieter. 2010. Scenarios for language contact. In Raymond Hickey (ed.), *The handbook of language contact*, 265–281. London: Wiley-Blackwell.
- Muysken, Pieter. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and cognition* 16(4). 709–730.
- Muysken, Pieter, Harald Hammarström, Joshua Birchall, Rik van Gijn, Olga Krasnoukhova & Neele Müller. 2014. Linguistic areas: Bottom-up or top-down? The case of the Guaporé-Mamoré. In Bernard Comrie & Lucia Golluscio (eds.), *Language contact and documentation / Contacto lingüístico y documentación*, 205–238. Berlin: Walter de Gruyter.
- Nettle, Daniel. 1999. *Linguistic diversity*. Oxford & New York: Oxford University Press.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Nichols, Johanna. 1997. Modeling ancient population structures and movement in linguistics. *Annual Review of Anthropology* 26(1). 359–384. DOI: 10.1146/annurev.anthro.26.1.359.

- Nichols, Johanna. 2003. Diversity and stability in language. In Richard D. Janda & Brian D. Joseph (eds.), *Handbook of historical linguistics*, 283–310. London: Blackwell.
- Niedzielski, Nancy & Howard Giles. 1996. Linguistic accommodation. In Hans von Goebel, Peter Nelde, Zdeněk Sary & Wolfgang Wölck (eds.), *Kontaktlinguistik: Ein internationales Handbuch zeitgenössischer Forschung*, vol. 1, 332–342. Berlin: de Gruyter.
- Poplack, Shana. 1978. Dialect acquisition among Puerto Rican bilinguals. *Language in Society* 7. 89–103.
- Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics* 18(7-8). 581–618.
- Ranacher, Peter, Rik van Gijn & Curdin Derungs. 2017. Identifying probable pathways of language diffusion in South America. In Arnold Bregt, Tapani Sarjakoski, Ron van Lammeren & Frans Rip (eds.), *Societal Geo-innovation: Selected Papers of the 20th AGILE Conference on Geographic Information Science, Wageningen, The Netherlands*. Cham, Switzerland: Springer.
- Ross, Malcolm. 2007. Calquing and metatypy. *Journal of Language Contact* 1(1). 116–143.
- Sandfeld, Kristian. 1926. *Balkanfilologien: En oversigt over dens resultater og problemer*. Copenhagen: Københavns Universitet.
- Seifart, Frank. 2015. Does structural-typological similarity affect borrowability? *Language Dynamics and Change* 5(1). 92–113.
- Sherzer, Joel. 1976. *An areal-typological study of American Indian languages north of Mexico*. Amsterdam: North-Holland.
- Sinnemäki, Kaius & Max Wahlström. 2018. *Case loss and the rise of articles: Evidence for a typological tendency*. Paper presented at 40. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft (DGfS). <https://www.dgfs2018.uni-stuttgart.de/arbeitsgruppen/ag-11-specificity-definiteness-and-article-systems-across-languages/dgfs2018-AG11-SinnemakiWahlstrom.pdf>.
- Stolz, Thomas. 2002. No Sprachbund beyond this line: On the age-old discussion of how to define a linguistic area. In Paolo Ramat & Thomas Stolz (eds.), *Mediterranean languages: Papers from the MEDTYP workshop, Tirrenia, June 2000*, 259–281. Bochum: Universitätsverlag Dr. N. Brockmeyer.
- Stolz, Thomas. 2006. All or nothing. In Yaron Matras, April McMahon & Nigel Vincent (eds.), *Linguistic areas: Convergence in historical and typological perspective*, 32–50. Basingstoke: Palgrave MacMillan.
- Swadesh, Morris. 1952. Lexicostatistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society* 96. 452–463.

- Thomason, Sarah G. 2010. Contact explanations in linguistics. In Raymond Hickey (ed.), *The handbook of language contact*, 29–47. London: Wiley-Blackwell.
- Thomason, Sarah G. & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Tristram, Hildegard. 2007. Why don't the English speak Welsh. *Britons in Anglo-Saxon England*. 192–214.
- Trubetzkoy, Nikolai. 1923. Vavilonskaja bašnja i smešenie jazykov. *Evrazijskij vremennik* 3. 107–124.
- Trubetzkoy, Nikolai. 1928. [proposition 16]. In *Actes du premier congrès international de linguistes*, 17–18. Paris.
- Trudgill, Peter. 2004. Linguistic and social typology: The Austronesian migrations and phoneme inventories. *Linguistic Typology* 8. 305–320.
- Van der Auwera, Johan. 1998. Revisiting the Balkan and Meso-American linguistic areas. *Language Sciences* 20(3). 259–270.
- Van Gijn, Rik. 2014. The Andean foothills and adjacent Amazonian fringe. In Loretta O'Connor & Pieter Muysken (eds.), *The native languages of South America: Origins, development, typology*, 102–125. Cambridge: Cambridge University Press.
- Van Gijn, Rik. 2020. Separating layers of information: The anatomy of contact zones. In Norval Smith, Tonjes Veenstra & Enoch O. Aboh (eds.), *Advances in contact linguistics: In honour of Pieter Muysken*, 162–178. Amsterdam: John Benjamins.
- Van Gijn, Rik, Harald Hammarström, Simon van de Kerke, Olga Krasnoukhova & Pieter Muysken. 2017. Linguistic areas, linguistic convergence, and river systems in South America. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 964–996. Cambridge: Cambridge University Press.
- Van Gijn, Rik & Pieter Muysken. 2020. Highland-lowland language interactions. In Adrian Pearce, David Beresford-Jones & Paul Heggarty (eds.), *Rethinking the Andes-Amazonia 'Divide': A cross-disciplinary exploration*, 178–210. London: UCL Press.
- Van Gijn, Rik, Sietze J. Norder, Leonardo Arias, Nicholas Q. Emlen, Matheus Azevedo, Allison Caine, Saskia Dunn, Austin Howard, Nora Julmi, Olga Krasnoukhova, Mark Stoneking & Jurriaan Wiegertjes. 2022. The social lives of isolates (and small language families): the case of the northwest Amazon. *Interface Focus* 13. 20220054. DOI: 10.1098/rsfs.2022.0054.
- Velten, H. V. 1943. The nez perce verb. *The Pacific Northwest Quarterly* 34(3). 271–292.

- Wahlström, Max. 2015. *Loss of case inflection in Bulgarian and Macedonian*. Helsinki: University of Helsinki. (Doctoral dissertation). <https://helda.helsinki.fi/bitstreams/9eb81cec-90d5-4a56-8e46-c36e259f3cac/download>.
- Weinreich, Uriel. 1953. *Languages in contact*. New York: Linguistic circle of New York.
- Wilkins, David P. 1996. Morphology. In Hans Goebel, Peter H. Nelde, Zdenek Stary & Wolfgang Wölck (eds.), *Contact linguistics: An international handbook of contemporary research*, vol. 1, 109–117. Berlin: Walter de Gruyter.
- Winford, Donald. 2005. Contact-induced changes: Classification and processes. *Diachronica* 22. 373–427.