Volker Gast* and Maria Koptjevskaja-Tamm*

Patterns of persistence and diffusibility in the European lexicon

<https://doi.org/10.1515/lingty-2021-2086>

Received June 15, 2019; accepted October 17, 2020; published online October 13, 2021

Abstract: This article investigates to what extent the semantics and the phonological forms of lexical items are genealogically inherited or acquired through language contact. We focus on patterns of colexification (the encoding of two concepts with the same word) as an aspect of lexical-semantic organization. We test two pairs of hypotheses. The first pair concerns the genealogical stability (persistence) and susceptibility to contact-induced change (diffusibility) of colexification patterns and phonological matter in the 40 most genealogically stable elements of the 100-items Swadesh list, which we call “nuclear vocabulary”. We hypothesize that colexification patterns are (a) less persistent, and (b) more diffusible, than the phonological form of nuclear vocabulary. The second pair of hypotheses concerns degrees of diffusibility in two different sections of the lexicon – “core vocabulary” (all 100 elements of the Swadesh list) and its complement (“non-core/peripheral vocabulary”). We hypothesize that the colexification patterns associated with core vocabulary are (a) more persistent, and (b) less diffusible, than colexification patterns associated with peripheral vocabulary. The four hypotheses are tested using the lexical-semantic data from the CLICS database and independently determined phonological dissimilarity measures. The hypothesis that colexification patterns are less persistent than the phonological matter of nuclear vocabulary receives clear support. The hypothesis that colexification patterns are more diffusible than phonological matter receives some support, but a significant difference can only be observed for unrelated languages. The hypothesis that colexification patterns involving core vocabulary are more genealogically stable than colexification patterns at the periphery of the lexicon cannot be confirmed, but the data seem to indicate a higher degree of diffusibility for colexification patterns at the periphery of the lexicon. While we

***Corresponding authors:** Volker Gast [*fɔl.ke 'gast*], Friedrich-Schiller-Universität Jena, Jena, Germany; and Maria Koptjevskaja-Tamm [*ma'ria kɔp'tʃɛfskaja 'tam*], Stockholm University, Stockholm, Sweden, E-mail: volker.gast@uni-jena.de (V. Gast), tamm@ling.su.se (M. Koptjevskaja-Tamm)

Open Access. © 2021 Volker Gast and Maria Koptjevskaja-Tamm, published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.

regard the results of our study as valid, we emphasize the tentativeness of our conclusions and point out some limitations as well as desiderata for future research to enable a better understanding of the genealogical versus areal distribution of linguistic features.

Keywords: areal semantics; colexification; contact-induced language change; core and peripheral vocabulary; lexical typology; semantic change; semantic typology

1 Introduction

One of the central challenges of comparative linguistics is to determine why languages are the way they are. Possible explanations have been succinctly summarized by Comrie (1989: 201) as follows:

“If we observe similarities between two languages, then there are, in principle, four reasons why these similarities may exist. First, they could be due to chance. Secondly, they could stem from the fact that the two languages are genetically related [...]. Thirdly, the two languages could be in areal contact [...]. Fourthly, the property could be a language universal ...”

Similarities between languages may obtain at various levels of linguistic organization, and different types of similarities may be due to different types of reasons. For example, if two languages exhibit phonological similarities between corresponding core vocabulary items such as French *oreille* and Romanian *ureche* ‘ear’, this similarity is likely due to genealogical relatedness. If two languages use the same type of formulaic expression for a conventionalized speech act, e.g. French *s’il vous plait* and Dutch *alstublieft* ‘please’, lit. ‘if it pleases you’, this similarity is likely due to language contact. If two languages have a nasal and an open vowel in their words for ‘mother’, that might be due to a universal tendency. The present article is concerned with similarities due to genealogical relatedness or language contact, and the relationship between these similarities.

Genealogical relatedness is standardly diagnosed through similarities in the phonological make-up of core vocabulary (e.g. Greenhill et al. 2010, 2017; Hock and Joseph 1996: 257; Nichols 2003; Thomason 2001: 71–72). This practice rests on the assumption that certain concepts are universally lexicalized, and the words expressing them are not easily replaced during language evolution, even though they may differ in their relative diachronic stability (e.g. Holman et al. 2008; Pagel et al. 2007; Tadmor et al. 2010). In other words, core vocabulary is assumed to exhibit a high degree of “persistence” (Swadesh 1955), and a low degree of “borrowability” (see for instance Carling et al. 2019; Tadmor et al. 2010).

There is no comparable standard diagnostic for degrees of language contact, and areal linguistics has traditionally depended on less universally applicable criteria. The more idiosyncratic and specific a shared feature is, the more reliably it is taken to reflect language contact. Evidence for a linguistic area often comes in the form of lists of similarities among geographically contiguous languages that are not found, or found rarely, outside of the area, and that cannot be explained in terms of genealogical relatedness. Typically, these are properties that were first noticed by experts working on the languages in question (e.g. Sandfeld 1930 on the Balkans, Kirchhoff 1943 on Mesoamerica, Emeneau 1956 and Masica 1976 on South Asia, Jakobson 1931, Décsy 1973 and Haarmann 1976 on the Baltic area, etc.). More often than not, such features do not relate to linguistic “matter”, but to “patterns” (Matras and Sakel 2007). For example, the Mesoamerican linguistic area has been characterized in terms of possessive constructions, a specific (vigesimal) numeral system and shared modes of presentation for specific concepts (such as ‘bird-stone’ for ‘egg’), among other features (see for instance Campbell 2013: 301; Gast 2007: 174).

Generalizations concerning borrowability and “diffusibility” (Matisoff 2001; Wichmann and Holman 2009) have mostly been formulated with respect to levels of the language system. For instance, it is well known that the lexicon is particularly susceptible to contact-induced change (cf. the Borrowing Scale proposed by Thomason and Kaufman 1988: 74–75, modified in Thomason 2001: 70–71; see also the surveys provided by Matras and Sakel 2007 and Kuteva 2017). By contrast, “[r]elatively little has been suggested on the potential ordering within pattern replication and on direct transfer of phonetics/phonology” (Koptjevskaja-Tamm 2011: 573). The pervasiveness of lexical calques is taken for granted in research on language contact and linguistic areas, but is also often dismissed as something trivial. Experts in particular linguistic areas are usually well aware of the shared lexico-semantic patterns, but this knowledge is often tacit, and, surprisingly, there have been very few attempts to use such patterns as areality indicators (with the two main exceptions being Mesoamerica, see Smith-Stark 1994 and Brown 2011, and Ethio-Eritrea, see Hayward 1991, 2000. Souag this volume and Liljegren this volume add two other areas: Sahel, the transitional zone between the Sahara to the north and the Sudan savanna to the south; and the Hindu Kush mountain range distributed through Afghanistan, Pakistan and India).

Thanks to the rapid development of areal typology, whose concern is “the study of patterns in the areal distribution of typologically relevant features of languages” (Dahl 2001: 1956), we are gaining new insights into the genealogically versus areally determined skewings in the distribution of linguistic features. We know, for instance, that evidentiality is easily diffused via language contact (Aikhenvald 2018), whereas gender and ergativity seem to be more resistant to

areal convergence (Nichols 2003), even though instances of transfer in these domains are also documented (Mithun 2005; Stolz 2012). Moreover, the development of large-scale linguistic databases, with WALS (Dryer and Haspelmath 2013)¹ as a prominent representative and Glottobank² as a currently growing project, has provided us with easily available and rich data on numerous properties across the world's languages, and has been instrumental for this programme. Such databases have increasingly been used as sources of information on the genealogical versus areal relations among languages (see for instance Dediu and Cysouw 2013; Murawaki and Yamauchi 2018; Wichmann 2015).

In this article we intend to determine the degrees to which linguistic features reflect genealogical relatedness or areal contact, in a sample of European languages. We compare the traditional diagnostic for genealogical relatedness – the phonological matter of nuclear vocabulary – to patterns in the form-meaning mapping in the lexicon; more specifically, to patterns of colexification (François 2008). We assume that patterns of colexification may emerge as a result of “pattern replication” (Matras and Sakel 2007), covering processes traditionally called “loan translation” (Haugen 1950) and “lexical calquing” (Ross 2001, 2007), as well as “loan meaning extension, [...] whereby a polysemy pattern of a donor language word is copied into the recipient language” (Haspelmath 2009: 39). While lexical patterns have been stated to be susceptible to contact-induced change (e.g. Koptjevskaja-Tamm and Liljegren 2017), this tendency has not so far been shown in a quantitative way, as far as we know. A first objective of this article then is to propose a method measuring degrees of persistence and diffusibility, and to determine such degrees in colexification patterns, in comparison to the phonological make-up of nuclear vocabulary. A second objective is to determine to what extent colexification patterns vary in their degrees of persistence and diffusibility. The expectation is that diffusibility is particularly high at the periphery of the lexicon, i.e. for concepts that are more specific and less frequent than core vocabulary, and acquired later in life.

We start in Section 2 with providing some theoretical background on colexification in the context of language contact. Section 3 contains an overview of the data used for the study. In Section 4, two pairs of hypotheses are tested: firstly the hypotheses that in the domain of nuclear vocabulary, colexification patterns show (a) a higher degree of diffusibility, and (b) a lower degree of persistence, than phonological matter; and secondly the hypotheses that degrees of (a) diffusibility and (b) persistence vary across sections of the lexicon (core vs. periphery). Section 5 contains some discussion as well as concluding remarks.

1 <https://wals.info>.

2 <https://glottobank.org>.

2 Theoretical background and two pairs of hypotheses

The term “colexification” was coined by François (2008: 170): “A given language is said to COLEXIFY two functionally distinct senses if, and only if, it can associate them with the same lexical form”. In addition to such cases of “strict” colexification (e.g. TOE and FINGER in Latvian *pirksts*) which is defined on the basis of identity of forms in synchrony, François (2008: Section 3.3) introduces the notion of “loose” colexification, which covers relatedness of forms encoding two concepts from a diachronic point of view as well as cases of partial identity of forms, e.g. in derivation or compounding (e.g. Germ. *Haupt* means ‘head’ as well as ‘main’, but it has the latter function only in compounds such as *Hauptbahnhof* ‘main station’). Whenever we use the term “colexification” in the present article, we mean “strict colexification”.

It should be noted that we will not address the general question of whether language-independent concepts even exist. We use translation equivalence in word lists (as recorded in the CLICS³ database, see Section 3) as a *tertium comparationis*. Possibilities and drawbacks associated with the use of lexical-semantic databases such as CLICS³ in the field of lexical typology have been discussed by Gast and Koptjevskaja-Tamm (2018).

Colexification patterns obviously vary across languages. For example, many Slavic languages (strictly) colexify the concepts MONTH and MOON³ (e.g. Czech *měsíc*), arguably as a result of metonymic extension, while other languages differentiate these concepts in synchrony, even though there is often a diachronic link (e.g. in English *month*, *moon*, German *Monat*, *Mond*). The assumption that languages either colexify or differentiate two concepts obviously implies a fair amount of simplification, as concepts may be encoded by more than one lexical item. For instance, Russian has both *luna* and *mesjac* for MOON. Only one of these words – *mesjac* – also means MONTH, and the question arises whether Russian counts as colexifying or not with respect to this pair of concepts. We treated concepts as being colexified in a given language if the data contained any element with both meanings, so Russian is classified as colexifying with respect to the pair <MONTH, MOON>, as the database contains *mesjac*, which is linked to both concepts.

Colexification patterns have mostly been studied from the point of view of lexical typology, focusing on the question of universal patterns of semantic relatedness (e.g. François 2008). More recently, the areal factor has come into the focus of attention, i.e. the question of the degree to which colexification patterns

3 We will represent concepts in small caps.

are susceptible to diffusion. The current project is embedded within the larger context of studies crossing the domains of areal linguistics and lexical typology (see for instance Juvonen and Koptjevskaja-Tamm 2016; Koptjevskaja-Tamm and Liljegren 2017; Urban 2012). In Gast and Koptjevskaja-Tamm (2018), we used the CLICS² database (List et al. 2018a, 2018b) to identify areal patterns in colexification on a global scale. In addition to some well-known areal clusters, such as <FIRE, TREE> in parts of North-East Australia and Papua New Guinea (cf. Schapper et al. 2016), we identified some new patterns, e.g. the colexification of MOUNTAIN and STONE in Central Africa, and of LEAF and EAR in East Africa as well as parts of North America. Given that we controlled for genealogical relatedness, such patterns have most likely resulted from language contact, some of them reflecting features of the natural habitat where the relevant languages are spoken. For example, the <MOUNTAIN, STONE> colexification seems to be common in arid areas with rocky mountains, where mountains actually resemble (large) stones.

In this article we zoom in on one of the regions with the highest density of data points in CLICS,³ i.e. Europe. Our focus is not on the associations of specific regions or areas with specific colexification patterns (as in Gast and Koptjevskaja-Tamm 2018), but on degrees of persistence and diffusibility of colexification patterns, in comparison to nuclear vocabulary matter, and relative to sections of the lexicon. We distinguish three sections of the lexicon, a nucleus, a core, and a periphery. We operationalize the notion of “core vocabulary” as words denoting concepts contained in the 100-item Swadesh list (Swadesh 1950, 1952, 1955; for a way of structuring the lexicon in terms of concepts, see Carling et al. 2019 in the context of borrowability). We use the term “nuclear vocabulary” for the 40 items of the Swadesh list with the highest frequencies of attestation in the data of the Automated Similarity Judgment Programme (ASJP)⁴ (Wichmann et al. 2018). These are the concepts with at least 4,000 data points in the ASJP database (version 18, comprising a total of 7,221 languages; see also Section 3), and they have been claimed to be particularly robust indicators of genealogical relatedness (Holman et al. 2008). We do not, of course, assume that core vocabulary consists of 100 items only; we take it that the 100-items of the Swadesh list form a representative sample of core vocabulary, and that the 40 “nuclear” items are particularly central elements within the core, as is reflected in their high degree of persistence. The concepts constituting nuclear and core vocabulary thus operationalized are listed in (1) and (2) (nuclear vocabulary is a subset of core vocabulary).

⁴ <http://asjp.clld.org>.

(1) Concepts constituting nuclear vocabulary

BLOOD, BONE, BREAST, COME, DIE, DOG, DRINK, EAR, EYE, FIRE, FISH, FULL, HAND, HEAR, HORN, KNEE, LEAF, LIVER, LOUSE, MOUNTAIN, NAME, NEW, NIGHT, NOSE, ONE, PATH, PERSON, SEE, SKIN, STAR, STONE, SUN, TONGUE, TOOTH, TREE, TWO, WATER

(2) Additional concepts constituting core vocabulary

ALL, ASH, BARK, BELLY, BIG, BIRD, BITE, BLACK, BURN, CLAW, CLOUD, COLD, DRY, EARTH, EAT, EGG, FEATHER, FLESH, FLY, FOOT, GIVE, GOOD, GREASE, GREEN, HAIR, HEAD, HEART, HOT, KILL, KNOW, LIE, LONG, MAN, MANY, MOON, MOUTH, NECK, NOT, RAIN, RED, ROOT, ROUND, SAND, SAY, SEED, SIT, SLEEP, SMALL, SMOKE, STAND, SWIM, TAIL, THAT, THIS, WALK, WHAT, WHITE, WHO, WOMAN, YELLOW

Colexification patterns may exhibit different areal and genealogical distributions. For example, the colexification of LANGUAGE and TONGUE is strongly genealogically conditioned. It is widespread in Romance (Fr. *langue*, Rom. *limbă*, Span. *lengua*), and Slavic (Cz. *jazyk*, etc.), going back to the proto-languages (Lat. *lingua*, Proto-Slavic *ęzyкъ). For Belorussian and Ukrainian CLICS³ lists *mová* for LANGUAGE though, which is also linked to the concept SPEECH. Latvian patterns with the Slavic languages in colexifying LANGUAGE and TONGUE (*mēle*), while Lithuanian colexifies LANGUAGE with SPEECH (*kalbà*). The colexification is not found in major Germanic languages. In the group of Celtic languages, only Irish colexifies LANGUAGE and TONGUE (*teanga*), according to CLICS³ data. Finnic languages (as well as Hungarian) seem to have this colexification (e.g. Finnish *kiele*), whereas the Saami languages do not seem to have it. The genealogical and areal distribution of the <LANGUAGE, TONGUE> colexification is shown in Figure 1. Languages exhibiting the pattern are black, languages not showing it are grey. The fact that black and grey languages form “blocks” within the dendrogram illustrates the close fit with genealogical relations. The right-hand side of Figure 1 shows the areal distribution of the pattern, which is scattered across Europe.

An example of a more areally distributed colexification is the one of BOIL (OF LIQUID) and COOK. Figure 2 shows its distribution. This pattern is scattered across the phylogenetic tree, while it forms a cluster in geographical terms. As the map in Figure 2 shows, the pattern is primarily found in the centre of Europe, specifically in Continental West Germanic and Continental Scandinavian languages (German *kochen*, Dutch *koken*, Yiddish *koxn*, Danish *koge*, Swedish *koka*). According to the CLICS³ data, the West Slavic languages Czech (*vařit*) and Polish also colexify these concepts, though Polish *warzyć* is specialized for ‘brewing (beer)’ in the contemporary language. Croatian *kuhati* has both of these meanings, too. Among the Baltic languages, the colexification is found in Latvian (*vārīt*) and Lithuanian (*virti*). Finally, Vlax Romani colexifies these concepts in the root *kirav-*.

Figure 1: Genealogical and areal distribution of the colexification <LANGUAGE, TONGUE>.

It is precisely such areal versus genealogical distributions that we are interested in. We test two pairs of hypotheses.

One pair of hypotheses relates to a comparison of phonological matter and colexification patterns. The first hypothesis, the Hypothesis of High Diffusibility of Colexification Patterns, is formulated in H1 (note that here and in the following, “high” and “low” are used in a relative sense, meaning “higher/lower than, in comparison to phonological form”).

H1: Hypothesis of High Diffusibility of Colexification Patterns

Degrees of similarity between languages in terms of colexification patterns involving nuclear vocabulary reflect language contact to a greater extent than

Figure 2: Genealogical and areal distribution of the colexification <BOIL (OF LIQUID), COOK (SOMETHING)>.

degrees of similarity in terms of the phonological form of nuclear vocabulary items do.

Put differently, the hypothesis in H1 says that colexification patterns are more diffusible than nuclear vocabulary matter (phonological make-up).

While we expect colexification patterns to reflect language contact, we expect the opposite tendency for genealogical relatedness. We thus formulate the Hypothesis of Low Persistence of Colexification Patterns in H2.

H2: Hypothesis of Low Persistence of Colexification Patterns

Degrees of similarity between languages in terms of colexification patterns involving nuclear vocabulary reflect genealogical relatedness to a lesser

extent than degrees of similarity in terms of the phonological form of nuclear vocabulary items do.

We use nuclear vocabulary, i.e. the 40 most frequently recorded items of the Swadesh list, to test the hypotheses in H1 and H2 because the phonological data that we use for our comparison (Jäger 2018, see Section 4) is also based on this set.

The hypotheses in H1 and H2 are motivated by considerations concerning mechanisms of diffusion. Specific types of contact-induced semantic change in the lexicon concern the transfer of a “use pattern” (Heine and Kuteva 2003, 2005) or “routine” (Gast and van der Auwera 2012), i.e. aspects of linguistic behaviour that speakers may perceive as rather minor, and often natural, innovations in the target language, and that can therefore be expected to be relatively unconstrained in terms of transfer. By contrast, contact-induced change in the phonological matter of nuclear vocabulary is known to be more unlikely, as the transfer of phonological segments typically proceeds via lexical borrowings, which are known to be rare in the nucleus or core of the lexicon (see for instance Hock and Joseph 1996: 257).

The second pair of hypotheses tested in this study concerns degrees of persistence and diffusibility across sections of the lexicon. More specifically, we hypothesize that colexifications involving core vocabulary exhibit higher degrees of persistence, and lower degrees of diffusibility, than colexifications involving peripheral lexical items. The corresponding “differential” hypotheses are formulated in H3 and H4.

H3: Hypothesis of Differential Diffusibility of Colexification Patterns

Colexifications involving core vocabulary are less susceptible to diffusion than colexifications involving peripheral vocabulary.

H4: Hypothesis of Differential Persistence of Colexification Patterns

Colexifications involving core vocabulary are more persistent than colexifications involving peripheral vocabulary.

The hypotheses in H3 and H4 are motivated by the assumption that highly frequent words, e.g. words denoting concepts belonging to the personal sphere (body parts), are primarily transmitted through first language acquisition (at an early age), and hence, tend to be genealogically inherited and less affected by

language contact, whereas other, less frequent concepts, which are acquired later in life, may be more susceptible to contact-induced change, for instance because they are acquired under conditions of multilingualism (see for example Gast 2017 and Szeto et al. 2019 on the relationship between language acquisition and language contact).⁵

It is important to note that the hypotheses concerning persistence and diffusibility are not equivalent. Diffusibility relates to the acquisition of new items or patterns through language contact, while persistence is a function of the (un)likelihood for a given item or pattern to be lost in language transmitted through first language acquisition. It is perfectly possible for a language to acquire a new colexification pattern while at the same time retaining existing patterns.

3 The data⁶

3.1 The language sample

The Database of Cross-Linguistic Colexifications (CLICS³, Rzymski et al. 2020) contains information about colexification patterns in 3,156 language varieties, based on 30 individual word lists.⁷ The data is areally biased, with a certain underrepresentation of North America and Africa. Our study focuses on European languages, an area for which the CLICS³ data is relatively dense. We investigate a sample of languages located within a rectangle ranging from 27°W 75°N (in the North-Western corner) to 38°E 32°N (in the South-Eastern corner). This area covers the region from Icelandic to Russian longitudinally, and from Inari Saami to Modern Greek latitudinally. Our European sub-sample comprises 45 (contemporary) languages, most of them coming from the Indo-European and Uralic families, as well as Turkish and Basque. One of the European languages represented in CLICS³ – Livvi – was excluded for reasons to become apparent below (it is not contained in the data used by Jäger 2018, see Section 4). The sample contains only a selection of the languages actually represented in CLICS³ because in many cases the amount of data available was too sparse for the purposes of this study (see Section 4). The sample is shown in Figure 3. National languages are located at their

⁵ Note that this assumption is not, of course, beyond doubt. We take it as a starting point and motivation for our hypotheses. As will be seen in Section 4, the hypothesis in H3 is not confirmed by our data, and the support for the hypothesis in H4 is not very strong.

⁶ The Supplementary Materials for this chapter can be found here: <https://doi.org/10.5281/zenodo.4459478>. Glottocodes for all languages mentioned can be found in the Supplementary Materials.

⁷ See <https://clics.clld.org/contributions> for the wordlists (accessed on 24 May, 2020).

predictors for degrees of similarity in terms of phonological make-up and colexification patterns in the languages of the sample (see Section 4).

3.2 The European languages as a contact network

The hypotheses tested in this study imply a comparison of degrees of genealogical inheritance and susceptibility to contact-induced change. As we pursue a quantitative approach, we need to operationalize both genealogical relatedness and language contact in some quantifiable way. We will use three alternative operationalizations of contact intensity. A very natural, and easy to implement, operationalization is geographical distance. An alternative way to capture language contact is by representing speech communities in the form of a network or graph. Given a language contact graph, we can either make a binary distinction between neighbours and non-neighbours, or we can measure distances between languages within the graph in terms of length of the shortest path.

In order to approximately capture language contact scenarios, we transformed the languages of our sample into a network, in such a way that neighbours in the network are likely language contact partners. We distinguished between major (national) languages and minor languages, as their geographical locations are represented differently. National languages are represented as both points (the coordinates of the weighted population means) and polygons (the borders of the countries where they are spoken). Minor languages are only represented as points (their coordinates). The contact network of the languages of our sample was created in five steps:

- All languages whose coordinates lie within a distance of 600 km from each other were connected. This distance was chosen because, with the exception of the Northwestern outliers Faroese and Icelandic, it yields a network in which all languages are connected to their closest neighbours in all cardinal directions.¹⁰
- All major languages were connected with their immediate geographical neighbours. Languages were considered immediate neighbours if the minimum distance between their polygons is smaller than 50 km.
- Minor languages were connected to the major languages of the countries where their coordinates are located.
- Minor languages were connected to major languages if the coordinates of the minor languages are no more than 100 km away from the polygon of the

10 The minimum distance in contact networks is a function of the density of data points; see Gast and Koptjevskaja-Tamm (2018) on different distances underlying language contact networks.

country where the major language is spoken, at the point of minimum distance.

- Finally, we considered the following seas as “contact bridges”: the Norwegian Sea, the Northern Atlantic, the Irish Sea, the North Sea, the English Channel, the Bay of Biscay, the Baltic Sea, the Western Mediterranean Sea (i.e., the Balearic Sea and the Tyrrhenian Sea), and the Adriatic Sea and Ionian Sea. Pairs of countries with access to the same sea from this list were treated as neighbours.

The resulting language contact network is shown in Figure 4.

3.3 The colexification data

The CLICS³ data was downloaded from the CLICS GitHub repository¹¹ and processed as follows:¹²

Figure 4: The sample as a contact network.

¹¹ <https://github.com/clics>.

¹² We omit some details, such as the mapping of language codes; see the file `py/01_prepare_data.py` in the Supplementary Materials for details.

- Forms with glosses and metadata were extracted from the CLICS³ database.¹³
- Concepts involving disjunction were removed (e.g. *PATH.OR.ROAD*).
- All “independent” colexifications were identified, i.e., pairs of concepts $\langle C_1, C_2 \rangle$ such that there is some form F in some language L that denotes C_1 and C_2 while not denoting any other concept in the database. This is intended to filter out “dependent” colexifications, i.e. colexifications that result only through intermediate concepts. For example, the colexification of *PENIS* and *REAR*, as observed in French *queue*, is (probably) mediated by *TAIL*.
- For any pair of concepts $\langle C_1, C_2 \rangle$ and any language L , if L has a form denoting C_1 and C_2 , and if C_1 and C_2 form an independent colexification, they were regarded as being colexified in L .
- We also added negative evidence. For any pair of concepts $\langle C_1, C_2 \rangle$ and any language L , if the database contains forms for C_1 and C_2 in L , and if there is no form denoting both C_1 and C_2 , C_1 and C_2 were treated as not being colexified in L .

A data point is thus a quadruple of the form $\langle C_1, C_2, L, CL \rangle$, where C_1 and C_2 are concepts, L is a language, and CL is a binary variable indicating whether or not C_1 and C_2 are colexified in L (i.e., *TRUE* vs. *FALSE*). For example, the data frame contains the quadruple $\langle \text{FUR, HAIR, Italian, TRUE} \rangle$ because *FUR* and *HAIR* are colexified in Italian (*pelle*), and it contains the quadruple $\langle \text{FUR, HAIR, English, FALSE} \rangle$ because English does not have a word that means both *FUR* and *HAIR*. Note that the data contains a high number of gaps (NAs, i.e. missing data points) for the variable CL , as in many cases a form was not retrievable for one of the concepts.

4 Testing the hypotheses

4.1 Association between distance matrices

We tested the four hypotheses under consideration with methods based on distance measures. The logic of our approach can be summarized as follows: We determined various types of distances – degrees of difference – between languages. Specifically, we can determine to what extent the nuclear vocabulary of two languages is (dis)similar in terms of its phonological make-up, and given the CLICS³ data, we can determine to what extent languages differ in their colexification patterns. These “internal” differences between languages can be

¹³ The data is stored in the *sqlite*-file *clics.sqlite*; *SQLite* is a library for relational databases (<http://www.sqlite.org>).

correlated with “external” differences, such as the location of the languages in geographical space, in a contact graph (see Figure 4), or relative to phylogenetic relationships.

As an operationalization of phonological distance, we used the distance matrix provided by Jäger (2018) that is based on correspondences between sound classes in approx. 7,000 lists of 40 items of nuclear vocabulary contained in the database of the Automated Similarity Judgment Program (ASJP,¹⁴ Jäger 2018 used version 17; the current version is 19).¹⁵ In this matrix, the score of dissimilarity between pairs of languages reflects the overall phonological dissimilarity between the words in the list. Degrees of similarity between two words w_1 and w_2 are a function of the degrees of similarity between the (aligned) sounds contained in w_1 and w_2 . For example, /æ/ is more similar to /a/ than it is to /i/ because there are more correspondences between /æ/ and /a/ in cognates (e.g. Engl. *man*, Germ. *Mann*) than there are correspondences between /æ/ and /i/. Degrees of similarity between sound classes were computed on the basis of the ASJP data (see Jäger 2018 for details). The matrix is visualized in the form of a heatmap in Figure 5 for our sample languages.¹⁶ Each cell represents the dissimilarity value for the languages in the respective column and row. Languages with a maximum degree of distance have a value of ‘1’ and are white in the diagram, languages with identical sound sequences in nuclear vocabulary have a value of ‘0’ and are black. This value is only found on the diagonal, where languages are compared with themselves. The dissimilarity matrix shown in Figure 5 will be called the “PHON-matrix” in the following.

(Dis)similarities between languages in terms of colexification patterns can be determined using the CLICS³ data. As was pointed out in Section 3, the data is binary, i.e. for each pair of concepts C_1 and C_2 , and for each language L , the database tells us whether C_1 and C_2 are colexified or not. For illustration, consider Table 1, which shows the data for four colexifications in three languages.

It is obvious that – looking at the four colexification patterns in Table 1 only – English and German are more similar to each other than either language is to Spanish. English and German share three of the four colexification patterns in Table 1 and differ in one. English shares one pattern with Spanish, while there is no pattern exhibited by both German and Spanish. Such intuitions need to be

¹⁴ <https://asjp.clld.org/>. The ASJP data is based on the Swadesh list(s) (Swadesh 1950, 1952, 1955).

¹⁵ We used Jäger’s (2018) list based on sound correspondences, rather than the list on cognate clustering, because the former list represents genealogical relationships more accurately in the data of our sample.

¹⁶ The heatmap was produced with the function `levelplot()` of the package `lattice` for R (Sarkar 2008).

Table 1: Four colexifications in English, German and Spanish.

	<EVENING, NIGHT>	<FINGERNAIL, NAIL>	<HUSBAND, MAN>	<LANGUAGE, TONGUE>
English	<evening, night> 0	<nail, nail> 1	<husband, man> 0	<language, tongue> 0
German	<Abend, Nacht> 0	<Nagel, Nagel> 1	<Mann, Mann> 1	<Sprache, Zunge> 0
Spanish	<noche, noche> 1	<uña, clavo> 0	<esposo, hombre> 0	<lengua, lengua> 1

transformed into numbers for a quantitative comparison. Given that the amount of information available varies across the language pairs, because the data has been taken from different sources,¹⁷ and given that the various colexifications exhibit different skewings between cases of TRUE and FALSE, we used a statistic that compares the observed overlap between two vectors (rows in Table 1) to the overlap that would be expected by chance. This statistic, Cohen’s κ , is commonly used as a measure of inter-rater reliability (Cohen 1960). It proved more robust for our dataset than other methods of similarity or difference, such as Euclidean or cosine distance, for the reasons mentioned above. The exact computation of the values of this matrix is described in Appendix A. The colexification distance matrix was rescaled to values between 0 (most similar) and 1 (most dissimilar).

In order to test the Hypothesis of High Diffusibility and the Hypothesis of Low Persistence, we only used colexification patterns that involved at least one item of nuclear vocabulary. We call these patterns “nuclear colexifications”. Given that we need a certain critical mass of both TRUE and FALSE observations within the sample languages, we only made use of colexification patterns that are attested in at least three sample languages. This led to a sample of 25 colexification patterns, listed in (3).

- (3) <ALONE, ONE>, <ARM, HAND>, <ARRIVE, COME>, <BLOOD, MEAT>, <BONE, LEG>, <BREAST, CHEST>, <CAMPFIRE, FIRE>, <DAY (NOT NIGHT), SUN>, <DIE, DIE (FROM ACCIDENT)>, <EVENING, NIGHT>, <FEEL (TACTUALLY), HEAR>, <FUR, SKIN>, <HAND, PALM OF HAND>, <HEAR, LISTEN>, <HEAR, UNDERSTAND>, <LANGUAGE, TONGUE>, <LEAF, LETTER>, <LEATHER, SKIN>, <LOOK, SEE>, <MAN, PERSON>, <MOUNTAIN, STONE>, <NEW, NEWS>, <NEW, YOUNG>, <PATH, ROAD>, <TREE, WOOD>

17 For the comparison of nuclear vocabulary we only used data from the NorthEuraLex (Dellert et al. 2019, <http://northeuralex.org/>) and the IDS (Key and Comrie 2015, <https://ids.clld.org/>). Note that these sources are not internally homogeneous, either, as they, in turn, rely on a broad range of word lists and dictionaries.

Figure 5: Heatmap visualizing Jäger's (2018) distance matrix reflecting similarity in sound classes between items of nuclear vocabulary (the PHON-matrix).

As pointed out in Section 3.1, we only used data from 45 languages. These are the languages for which the database contains information on at least 22 of the 25 nuclear colexifications. A dissimilarity matrix for these languages, based on the 25 (nuclear) colexification patterns listed in (3), is shown in Figure 6. We will refer to this matrix as the “COLEX-matrix” in the following.

The first two hypotheses to be tested say that colexification patterns exhibit a higher degree of diffusibility, and a lower degree of persistence, than the phonological form of lexical items, in the domain of nuclear vocabulary. We can test these hypotheses by comparing the two distance matrices shown in Figures 5 and 6 to matrices capturing our operationalizations of contact intensity and of phylogenetic distance. As mentioned in Section 3, one way of operationalizing contact intensity is by using geographical distances. An alternative way is by using the contact network shown in Figure 4. This contact network can be transformed into a

Figure 6: Heatmap visualizing shared colexification patterns between 25 European languages (the COLEX-matrix).

distance matrix in two ways. First, we can transform it into a binary matrix showing whether or not two languages are neighbours in the network; and second, we can determine the shortest paths between any pair of languages, which gives us values between 1 (from one neighbour to the next) and 5 (the largest distance found in the network, e.g. between Icelandic and Turkish). After rescaling the data we thus have values between 0.2 and 1. The matrices operationalizing contact intensity are shown in Figure 7 (the first three matrices from the left; the arrangement of languages in rows and columns is the same as in Figures 5 and 6). In the following, these matrices will be called `CONTACT.GEO` (log-transformed geographical distance), `CONTACT.PATH` (shortest path in the neighbourhood graph), and `CONTACT.NB` (neighbourhood in the neighbourhood graph).

In order to measure phylogenetic relatedness, we used the genealogical information from Glottolog (Hammarström et al. 2019). We first created a

Figure 7: Matrices of contact distance: `CONTACT.GEO`, `CONTACT.PATH` and `CONTACT.NB`, and matrix of phylogenetic distance (`PHYLO`) (from left to right).

dendrogram of the sample languages with this information, see Figures 1 and 2 above. Phylogenetic distance was operationalized as the shortest path between any two nodes in the dendrogram shown in Figures 1 and 2. The branches were assigned weights. In this way the different degrees of granularity exhibited by the various (sub-)families can be taken into account. For example, without weights, the shortest path between Basque and Turkish has the same distance as the one between Czech and Slovak. The branches were weighted in such a way that the distance from any one leaf to the root node is the same. The right-most plot in Figure 7 visualizes the matrix of phylogenetic distance thus created, henceforth the “PHYLO-matrix”.¹⁸ For some of the quantitative analyses, we treated genealogical distance as a categorical variable, grouping languages into “lower-level related” (same branch), “higher-level related” (same family, different branch) and “unrelated”.¹⁹

Our hypotheses can be operationalized in terms of the following questions: To what extent do the matrices containing language-internal information (`PHON` and `COLEX`) correlate with the `CONTACT`-matrices, given the `PHYLO`-matrix? And to what extent do the former matrices (`PHON` and `COLEX`) correlate with the `PHYLO`-matrix, given the `CONTACT`-matrices? In the following, we will focus on the most important results only. More details are provided in Appendix B. All the data and scripts are contained in the Supplementary Materials.

¹⁸ Given that the families represented in the sample differ in terms of time depth, we slightly adjusted the weights using the age estimates of Gray and Atkinson (2003) for the quantitative analysis.

¹⁹ Given the age adjustments mentioned in Note 18, Baltic and Slavic languages stand in a “higher-level” relation to each other, as do Goidelic and Brythonic languages within the Celtic branch.

4.2 The hypotheses of high diffusibility and low persistence

Our data show a clear (negative) interaction between genealogical distance and contact distance as predictors of phonological distance:²⁰ for closely related languages the correlation between contact distance and phonological distance is stronger than for more remotely related, or unrelated, languages (see for instance Epps et al. 2013 on language contact among related languages). Figure 8 visualizes the data for geographical distance as an operationalization of contact intensity (on the *x*-axis). The plots in the top row show phonological distance, the plots in the bottom row show colexification distance (on the *y*-axes), with each dot representing one pair of languages. The three columns correspond to lower-level related, higher-level related and unrelated pairs of languages, from left to right. The slope of the regression line is steepest for closely related languages (in the left-most plot) in the top row showing phonological distances, while it is steepest for unrelated languages (in the right-most plot) in the bottom row showing colexification distances. This suggests that the correlation between geographic distance and phonological distance may be strongest for more closely related languages, while the correlation between geographic distance and colexification distance may be strongest for unrelated languages. The plots also show substantial differences in the variances (scatter of the data points along the *y*-axis). Phonological distances between higher-level related or unrelated languages are consistently rather high, with little variance (the top-centre and top-right plots), while more variance can be observed in phonological distances between closely related languages (the top-left plot), and in colexification distances across levels of phylogenetic distance (all plots in the bottom row). The high amount of variance in colexification distances, in comparison to phonological distances, is unsurprising, given the relatively high amount of noise in the CLICS³ data.

The statistical analysis of distance data as used in the present study is non-trivial because it does not satisfy the independence assumption underlying most relevant statistical test procedures.²¹ We therefore analysed the data by language (using hierarchical regression modelling, see for instance Austin et al. 2001) and based our conclusions on comparisons of paired sets of correlation values thus

20 The interaction was confirmed with distance-based Redundancy Analysis, for which we used the *vegan*-package for R (Oksanen et al. 2020).

21 Measurements are statistically independent of each other if no information about any measurement can be gained from any other measurements. This is not the case in a distance matrix, where distances between pairs of languages can (partly) be inferred from other sets of distances. Moreover, the specific genealogical groupings of the languages have to be taken into account (in addition to the genealogical distances, which are derived from these groupings).

Figure 8: Phonological distances (top) and colexification distances (bottom), plotted against geographical distance (x-axis), for three groups of phylogenetic relationships (lower-level related, higher-level related and unrelated, from left to right). Each dot corresponds to one pair of languages.

obtained (again, using hierarchical regression modelling). Since the focus of this article is on the results, not on the methodology, we have chosen not to present all the details of the statistical analyses here. As mentioned above, more information about the methods is given in Appendix B, and the Supplementary Materials contain all the data and scripts. In the following discussion we use standardized regression coefficients (corresponding to the slopes of the regression lines), for the sake of simplicity called “Beta coefficients”, as indicators of effect size, and as approximations to correlation strength (see Rodgers and Nicewander 1988: 62).²²

In a first step, correlations between the variables of interest were determined separately for each language. Consider Figure 9 for illustration. The plot at the top shows phonological distances (y-axis) in relation to geographical distances (x-axis) for Russian. The regression lines for the various genealogical groups (lower-level related, higher-level related, unrelated) are rather flat. This is different in the plot at the bottom, which shows colexification distances in relation to geographical distance. The red line, corresponding to unrelated languages, is particularly steep, indicating a relatively strong correlation between geographical distance and colexification distance for languages that are not related to Russian. In terms of colexifications, Russian is comparatively similar to the (geographically

²² Beta coefficients are “standardized” insofar as the underlying variables are transformed in such a way that they have a mean of 0 and a variance of 1, thus enabling comparison across different units of measurement and taking differences in variance into account.

close) Finnic languages, and very different from Turkish and Basque. While we have not inspected the data in detail, Figure 9 suggests that Russian and the Finnic languages share a substantial number of colexification patterns as a result of language contact.

By fitting hierarchical regression models to each language separately, for each correlation type, we obtained twelve sets of 45 Beta coefficients (a correlation type can be represented as, for instance, (COLEX ~ CONTACT.GEO)|PHYLO, for the correlation between colexification distance and geographical distance, controlling for

Figure 9: Phonological distances (top) and colexification distances (bottom) in relation to geographical distance (for Russian).

phylogenetic distance).²³ This allowed us to compare pairs of correlation types (represented as paired sets of Beta coefficients). Technically, this was done, again, using hierarchical regression modelling (see Appendix B). For the sake of simplicity, in the following we report the mean values of the Beta coefficients determined for all sample languages as well as confidence intervals around these values, and we indicate approximate *p*-values indicating global degrees of significance for differences between the paired sets of Beta coefficients.

The plot on the left in Figure 8 compares correlations between phylogenetic distance and (i) colexification distance (black), and (ii) phonological distance (grey), for language pairs at a mid-high geographical distance, for the three operationalizations of contact intensity (“mid-high” distances are those located in the third quantile of geographical distances). The dots show the mean values of the Beta coefficients for the 45 sample languages, with a 95%-confidence interval. These values can be interpreted as showing degrees of persistence. The horizontal lines in the plots represent pairwise comparisons: solid lines indicate significant differences between paired sets of Beta coefficients, dashed lines show the absence of a significant difference. The Beta coefficients for phonological distance and phylogenetic distance are significantly higher than those for colexification distance and phylogenetic distance, for all operationalizations of contact intensity ($p < 0.01$).²⁴ Regardless of how contact intensity is measured, the phonological matter of core vocabulary is correlated more strongly with phylogenetic distance – i.e. it is more persistent – than colexification patterns of the relevant lexical items.

The plot on the right compares correlations between the three operationalizations of contact intensity and (i) phonological distance (black), and (ii) colexification distance (grey), for unrelated languages. The values shown in this plot can be interpreted as reflecting degrees of diffusibility. For phonological distance, the values tend towards 0, and the confidence interval crosses the zero line if CONTACT.NB is used as a predictor matrix. There is thus no significant correlation between phonological distance and contact distance measured in terms of neighbourhood in the contact graph. The correlations are consistently stronger for colexification patterns, even though a statistically significant difference between sets of Beta coefficients can only be observed for two operationalizations of contact intensity, i.e. PATH ($p = 0.03$) and GEO ($p < 0.01$). It is important to mention, however,

²³ There are twelve correlation types because there are two response variables (PHON and COLEX) and two predictors (CONTACT and GEO), with three different operationalizations of CONTACT (NB, PATH and GEO).

²⁴ This holds for all geographical distances between languages, and all operationalizations of contact distance, with one exception: no significant difference was observed for languages in the upper quantile of geographical distance, under the PATH-operationalization of contact distance.

that significant differences between Beta coefficients for phonological distances and colexification distances could only be observed for unrelated languages (see also Figure 8).

The results of the regression analysis can be used to test our first pair of hypotheses, i.e. the Hypothesis of High Diffusibility and the Hypothesis of Low Persistence. With respect to the latter hypothesis, our analyses show clearly that the phonological matter of nuclear vocabulary is more persistent than colexification patterns involving nuclear vocabulary, under any of the control conditions for contact intensity. The Hypothesis of Low Persistence of Colexification Patterns thus receives clear support from our data, perhaps unsurprisingly so.

The Hypothesis of High Diffusibility receives partial support from our analysis. The correlation between contact distance and colexification distance is significantly stronger than the correlation between contact distance and phonological distance only for pairs of unrelated languages. Moreover, a significant difference can only be observed for two operationalizations of contact intensity (PATH and GEO). We should also bear in mind that the Hypothesis of High Diffusibility was formulated in a comparative way, saying that colexification patterns are more diffusible than the phonological matter of nuclear vocabulary. In absolute terms, the correlations between colexification distance and contact intensity are rather weak even for unrelated languages, with a maximum mean Beta coefficient 0.15 (for CONTACT.GEO). As Figure 10 shows, distances based on colexification patterns – like those based on the

Figure 10: Results of regression analyses (mean Beta coefficients with 95% confidence intervals). The plot on the left shows the values for PHON and PHYLO (black) in comparison to COLEX and PHYLO (grey), for language pairs at a mid-high distance (corresponding to the third quantile of geographical distances), with the three operationalizations of contact intensity as control variables ($p < 0.01$ for all scenarios). The plot on the right shows the mean Beta coefficients for the three CONTACT matrices and the PHON-matrix (black), in comparison to the COLEX-matrix, for unrelated languages ($p > 0.05$ for CONTACT.NB, $p = 0.03$ for PATH, $p < 0.01$ for GEO).

phonological make-up of nuclear vocabulary – primarily reflect genealogical relatedness, not contact intensity.

4.3 The hypotheses of differential diffusibility and persistence

In order to test the two “differential” hypotheses, we compared colexification patterns involving core vocabulary (operationalized as the 100 items of the Swadesh list) to colexification patterns not involving core vocabulary. As the coverage of the data for the latter group was somewhat uneven (the number of missing values varies across colexification patterns), we filtered the data, excluding those colexification patterns for which more than 50% of the data was missing. Moreover, we removed languages for which less than 50% of the data was available. This left us with 281 “peripheral colexification patterns” in 43 languages, which were compared to the 61 “core colexification patterns”. We created two distance matrices for the 43 sample languages, using the method described in Section 4.1, and we analysed the data as described in Section 4.2.

The mean Beta coefficients for the correlations between the *PHYLO*-matrix and the two *COLEX*-matrices (based on core colexifications and peripheral colexifications) are shown on the left-hand side of Figure 11. These values can be interpreted as indicating degrees of persistence. There is no significant difference. By contrast, the plot on the right shows significant differences between the Beta

Figure 11: Results of regression analyses (mean Beta coefficients with 95% confidence intervals). The plot on the left shows the values for *COLEX.CORE* and *PHYLO* (black) in comparison to *COLEX.NON-CORE* and *PHYLO* (grey). The plot on the right shows the values for the two *COLEX*-matrices and the three *CONTACT* matrices (black) ($p > 0.05$ for NB, $p < 0.01$ for PATH, $p < 0.01$ for GEO).

coefficients corresponding to the correlations between contact intensity and distances computed on the basis of the two groups of colexifications for two operationalizations of contact intensity, *PATH* and *GEO* ($p < 0.01$ in both cases). While the Hypothesis of Differential Persistence of Colexification Patterns can be rejected, the Hypothesis of Differential Diffusibility of Colexification Patterns thus receives some support from our data, though the evidence is not compelling. Note also that in absolute terms, the correlations between peripheral colexification patterns and contact intensity are relatively weak.

5 Conclusions

In the present study we set out to test two pairs of hypotheses concerning degrees of persistence and diffusibility of colexification patterns, in comparison to the phonological make-up of vocabulary, and in relation to the sections of the lexicon (core vs. periphery). Three of the hypotheses received at least some support: the Hypothesis of Low Persistence of Colexification Patterns, the Hypothesis of High Diffusibility of Colexification Patterns, and the Hypothesis of Differential Diffusibility of Colexification Patterns. Unsurprisingly perhaps, (nuclear) colexification patterns are less genealogically persistent than the phonological matter of nuclear vocabulary. They seem to be more susceptible to contact-induced change, according to our results, though significant differences between the degrees to which phonological distance and colexification distance correlate with contact intensity were only found for unrelated languages, and only for two of three operationalizations of contact intensity (geographical distance/*CONTACT.GEO* and distance in a contact graph/*CONTACT.PATH*). Moreover, the overall correlations that we observed between colexification distance and contact intensity were rather weak. Similar observations can be made with respect to the Hypothesis of Differential Diffusibility. While our analysis showed a significant difference between the degrees to which core and peripheral colexifications correlate with contact intensity for two of the three operationalizations of contact intensity (*CONTACT.GEO* and *CONTACT.PATH*), the absolute correlations were very weak in both cases. The Hypothesis of Differential Persistence of Colexification Patterns was rejected on the basis of the data used for this study.

With this study we hope to have made a contribution to the programme of a theoretically informed, quantitative synthesis of historical-comparative and areal linguistics. It should have become clear, however, that our results are highly tentative, and that they should be regarded as starting points for more in-depth studies of this type. One of the major challenges of the type of project pursued in this article remains the operationalization of contact intensity. We used three

exploratory measures of contact intensity, two of which were based on a ‘contact graph’. The make-up of the contact graph is obviously influenced by data availability, and it depends on a number of subjective decisions, e.g. what minimum distance for network neighbours is assumed, whether or not languages across a sea should be connected, etc. Another difficulty concerns data availability. The CLICS³ data used for this study is rather noisy, and there is a lot of missing data. This is not intended to criticize this endeavour, to which we have ourselves contributed (Rzyski et al. 2020), but to call for more cooperation between and among specialists and generalists with the objective of generating high-quality resources with broad coverage enabling a more precise understanding of why languages are the way they are.

Acknowledgements: The study was designed and written jointly by both authors. The statistical analyses were carried out by VG. We are indebted to various colleagues for valuable feedback on earlier versions of this article, and talks given about its topic. Comments made by Martin Haspelmath and Christoph Rzyski have been particularly valuable. We also wish to thank two anonymous reviewers as well as the editors of this special issue for their remarks and productive criticism. Any inaccuracies are of course our own responsibility.

Research funding: MKT’s research is supported by grant 2018-01184 from the Swedish Research Council.

Appendix A: Cohen’s κ

We illustrate Cohen’s κ with the data in Table 1. Since in this table, English exhibits one colexification out of four (0.25), and German two (0.5), we can expect 0.125 shared colexifications by chance ($= 0.25 \times 0.5$), and 0.375 ($= 0.75 \times 0.5$) cases of shared differentiation, because three of the four pairs are differentiated in English, and two in German. Altogether we can expect an overlap of 0.5 ($= 0.125 + 0.375$). What we find is an overlap of 0.75, as three of four columns show identical values in English and German. These languages thus share more colexification patterns than would be expected by chance. The κ -value is determined as shown in (1) (p_o is the observed overlap, p_e is the expected overlap).

$$(1) \quad \kappa = \frac{p_o - p_e}{1 - p_e}$$

The κ -value ranges from -1 (no overlap) and to $+1$ (identity). In the example given above, we would get a value of 0.5 for German and English (see (2) below), a value

of -0.5 for English and Spanish, and a value of -1 for German and Spanish. The values were rescaled to a range from 0 to 1.

$$(2) \quad \kappa = \frac{0.75-0.5}{1-0.5} = 0.5$$

Appendix B: Notes on statistical analyses

The statistical analysis of distance data as used in the present study does not satisfy the independence assumption in two ways (see also Section 4.2): first, the distances are not independent of each other because each language contributes to a number of measurements; and second, the genealogical (sub-)groupings can be expected to be associated with specific tendencies, even though phylogenetic distance is used as a predictor. For example, if Romance languages and Saami languages exhibit a specific pattern of (dis)similarity, adding more Romance and Saami languages to the sample would artificially strengthen that effect at the sample level. In order to address the first problem, we analysed correlations between the two predictor matrices (phylogenetic distance and contact distances) and the response matrices (phonological distance and colexification distance) by language. The second problem was addressed by using hierarchical regression modelling (Austin et al. 2001), treating the lowest-level genealogical branch (e.g. Saami, Romance) as a random variable. We thus fitted a separate model for each language, obtaining pairs of measurements such that the two values indicate correlations between a specific predictor (genealogical distance or contact distance) and two alternative response variables (phonological distance vs. colexification distance, distance based on core colexifications vs. distance based on non-core colexifications). We use standardized regression coefficients (“Beta coefficients”) as approximate indicators of correlation strength (see for instance Rodgers and Nicewander 1988: 62), but we also inspected other statistics such as the structure coefficients and, of course, the amounts of variance explained. The plots show the mean Beta coefficients obtained with the regression analyses for the sample languages, with 95% confidence intervals obtained with bootstrapping. Significance values were determined by fitting hierarchical regression models to the sets of paired Beta coefficients, using ‘language’ and ‘branch’ as random variables. We also applied Wilcoxon signed-rank tests (Wilcoxon 1945) for comparison. The two test procedures converged insofar as they delivered p -values in identical ranges ($p > 0.05$, $0.05 > p > 0.01$, $0.01 > p$). This shows that the effect of specific genealogical groupings is in fact negligible. In order to get a better idea of the methods and results, readers are invited to inspect the Supplementary Materials.

References

- Aikhenvald, Alexandra Y. 2018. *Evidentiality*. Oxford: Oxford University Press.
- Austin, Peter C., Vivek Goel & Carl van Walraven. 2001. An introduction to multilevel regression models. *Canadian Journal of Public Health* 92(2). 150–154.
- Brown, Cecil H. 2011. The role of Nahuatl in the formation of Mesoamerica as a linguistic area. *Language Dynamics and Change* 1. 171–204.
- Campbell, Lyle. 2013. *Historical linguistics. An introduction*, 3rd edn. Edinburgh: Edinburgh University Press.
- Carling, Gerd, Sandra Cronhamn, Robert Farren, Elnur Aliyev & Johan Frid. 2019. The causality of borrowing: Lexical loans in Eurasian languages. *PloS One* 14(10). e0223588.
- Cohen, Jacob. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1). 37–46.
- Comrie, Bernard. 1989. *Language universals and linguistic typology*, 2nd edn. London: Blackwell.
- Dahl, Östen. 2001. Principles of areal typology. In Martin Haspelmath, Ekkehard König, Wolfgang Raible & Wulf Oesterreicher (eds.), *Language universals and language typology: An international handbook*, vol. 2, 1456–1470. Berlin: de Gruyter Mouton.
- Décsy, Gyula. 1973. *Die linguistische Struktur Europas. Vergangenheit–Gegenwart–Zukunft*. Wiesbaden: Otto Harrassowitz.
- Dediu, Dan & Michael Cysouw. 2013. Some structural aspects of language are more stable than others: A comparison of seven methods. *PloS One* 8(1). e55009.
- Dellert, Johannes, Thora Daneyko, Alla Münch, Alina Ladygina, Armin Buch, Natalie Clarius, Ilja Grigorjew, Mohamed Balabel, Hizniye Isabella Boga, Zalina Baysarova, Mühlenbernd Roland, Wahle Johannes & Gerhard Jäger. 2019. NorthEuraLex: A wide-coverage lexical database of Northern Eurasia. *Language Resources and Evaluation* 54. 273–301.
- Dryer, Matthew & Martin Haspelmath. 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: <https://wals.info>.
- Emeneau, Murray Barnson. 1956. *India and historical grammar*. No 5. Annamalainagar: Annamalai University Publications in Linguistics.
- Epps, Patience, John Huehnergard & Pat-El Na'ama. 2013. Introduction: Contact among genetically related languages. *Journal of Language Contact* 6. 209–219.
- François, Alexandre. 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In Martine Vanhove (ed.), *From polysemy to semantic change*, 163–215. Amsterdam: Benjamins.
- Gast, Volker. 2007. From phylogenetic diversity to structural homogeneity: On right-branching constituent order in Mesoamerica. *SKY Journal of Linguistics* 20. 171–202.
- Gast, Volker. 2017. Paradigm change and language contact: A framework of analysis and some speculation about the underlying cognitive processes. *JournaLIPP* 5. 49–70.
- Gast, Volker & Johan van der Auwera. 2012. What is “contact-induced grammaticalization”? Examples from Mesoamerican languages. In Björn Wiemer & Björn Hansen (eds.), *Grammatical replication and grammatical borrowing in language contact*, 381–426. Berlin: 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: de Gruyter Mouton.
- Gray, Russel D. & Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426. 435–349.

- Greenhill, Simon J., Quentin D. Atkinson, Andrew Meade & Russell D. Gray. 2010. The shape and tempo of language evolution. *Proceedings of the Royal Society B* 277. 2443–2450.
- Greenhill, Simon J., Chieh-Hsi Wu, Hua Xia, Michael Dunn, Stephen C. Levinson & Russell D. Gray. 2017. Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences of the United States of America* 114(42). E8822–E8829.
- Haarmann, Harald. 1976. *Aspekte der Arealtypologie: Die Problematik der europäischen Sprachbünde*. Tübingen: Narr.
- Hammarström, Harald, Robert Forkel & Martin Haspelmath. 2019. *Glottolog 3.3*. Jena: Max Planck Institute for the Science of Human History. <http://glottolog.org> (accessed 3 January 2019).
- Haspelmath, Martin. 2009. Lexical borrowing: Concepts and issues. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the world's languages: A comparative handbook*, 35–54. Berlin: Mouton de Gruyter.
- Haugen, Einar. 1950. The analysis of linguistic borrowing. *Language* 26. 210–231.
- Hayward, Richard J. 1991. A propos patterns of lexicalization in the Ethiopian language area. In Daniela Mendel & Ulrike Claudi (eds.), *Ägypten im afroorientalischen Kontext. Special issue of Afrikanistische Arbeitspapiere*, 139–156. Cologne: Institute of African Studies.
- Hayward, Richard J. 2000. Is there a metric for convergence? In Colin Renfrew, April M. S. McMahon & Robert Lawrence Trask (eds.), *Time depth in historical linguistics Vol 2 (Papers in the Prehistory of Languages)*, 621–640. Cambridge: The McDonald Institute for Archaeological Research.
- Heine, Bernd & Tania Kuteva. 2003. On contact-induced grammaticalization. *Studies in Language* 27(3). 529–572.
- Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Hock, Hans Henrich & Brian D. Joseph. 1996. *Language history, language change, and language relationship: An introduction to historical and comparative linguistics*. Berlin: Mouton de Gruyter.
- Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller & Dik Bakker. 2008. Explorations in automated language classification. *Folia Linguistica* 42(2). 331–354.
- Jakobson, Roman. 1931. Über die phonologischen Sprachbünde. *Travaux* 4. 234–240.
- Jäger, Gerhard. 2018. Global-scale phylogenetic linguistic inference from lexical resources. *Scientific Data* 5(2018). 180189.
- Juvonen, Päivi & Maria Koptjevskaja-Tamm (eds.). 2016. *The lexical typology of semantic shifts*. Berlin & New York: de Gruyter Mouton.
- Key, Mary Ritchie & Bernard Comrie (eds.). 2015. *The intercontinental dictionary series*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://ids.clld.org> (accessed 19 April 2020).
- Kirchhoff, Paul. 1943. Mesoamérica, sus límites geográficos, composición étnica y caracteres culturales. *Acta Americana* 1(1). 92–107.
- Koptjevskaja-Tamm, Maria. 2011. Linguistic typology and language contact. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*, 504–533. Oxford: Oxford University Press.
- Koptjevskaja-Tamm, Maria & Henrik Liljegen. 2017. Lexical semantics and areal linguistics. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 204–236. Cambridge: Cambridge University Press.
- Kuteva, Tania. 2017. Contact and borrowing. In Adam Ledgway & Ian Roberts (eds.), *The Cambridge handbook of historical syntax*, 163–186. Cambridge: Cambridge University Press.
- List, Johann-Mattis, Simon J. Greenhill, Cormac Anderson, Thomas Mayer, Tiago Tresoldi & Robert Forkel. 2018a. CLICS2: An improved database of cross-linguistic colexifications assembling lexical data with the help of cross-linguistic data formats. *Linguistic Typology* 22(2). 277–306.

- List, Johann-Mattis, Simon J. Greenhill, Corman Anderson, Thomas Mayer, Tiago Tresoldi & Robert Forkel. 2018b. *Database of cross-linguistic colexifications*. Jena: Max Planck Institute for the Science of Human History. <http://clics.clld.org> (accessed 19 March 2019).
- Masica, Colin. 1976. *Defining a linguistic area: South Asia*. Chicago: University of Chicago Press.
- Matisoff, James. 2001. Genetic versus contact relationship: Prosodic diffusibility in South-East Asian languages. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Grammars in contact: A cross-linguistic typology*, 291–327. Oxford: Oxford University Press.
- Matras, Yaron & Jeanette Sakel. 2007. Investigating the mechanisms of pattern replication in language convergence. *Studies in Language* 31. 829–865.
- Mithun, Marianne. 2005. Ergativity and language contact on the Oregon Coast: Alsea, Siuslaw, and Coos. *Berkeley Linguistic Society* 27. 77–95.
- Murawaki, Yugo & Kenji Yamauchi. 2018. A statistical model for the joint inference of vertical stability and horizontal diffusibility of typological features. *Journal of Language Evolution* 3(1). 13–25.
- Nichols, Johanna. 2003. Diversity and stability in language. In Richard D. Janda & Brian D. Joseph (eds.), *The handbook of historical linguistics*, 283–310. Oxford: Blackwell.
- Oksanen, Jari, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, Robert Brian O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs & Helene Wagner. 2020. vegan: Community ecology package. R package version 2.5-7. Available at: <https://CRAN.R-project.org/package=vegan>.
- Pagel, Mark, Quentin D. Atkinson & Andrew Meade. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449(7163). 717–720.
- Rodgers, Joseph Lee & W. Alan Nicewander. 1988. Thirteen ways to look at the correlation coefficient. *The American Statistician* 42(1). 59–66.
- Ross, Malcolm. 2001. Contact-induced change in Oceanic languages in North-West Melanesia. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, 134–166. Oxford: Oxford University Press.
- Ross, Malcolm. 2007. Calquing and metatypy. *Journal of Language Contact* 1(1). 116–143.
- Rzyski, Christoph, Tiago Tresoldi, Simon Greenhill, Mei-Shin Wu, Nathanael Schweikhard, Maria Koptjevskaja-Tamm, Volker Gast, Timotheus Bodt, Abbie Hantgan, Gereon Kaiping, Sophie Chang, Yunfan Lai, Natalia Morozova, Heini Arjava, Nataliia Hübler, Ezequiel Koile, Steven Pepper, Mariann Proos, Briana Van Epps, Ingrid Blanco, Carolin Hundt, Sergei Monakhov, Kristina Pianykh, Sallona Ramesh, Russell Gray, Robert Forkel & Johann-Mattis List. 2020. The database of cross-linguistic colexifications, reproducible analysis of cross-linguistic polysemies. *Scientific Data* 7. 13.
- Sandfeld, Kristian. 1930. *Linguistique balkanique*. Paris: Honoré Champion.
- Schapper, Antoinette, Lila San Roque & Rachel Hendery. 2016. Tree, firewood and fire in the languages of Sahul. In Maria Koptjevskaja-Tamm & Päivi Juvonen (eds.), *Lexico-typological approaches to semantic shifts and motivation patterns in the lexicon*, 355–422. Berlin: de Gruyter Mouton.
- Smith-Stark, Thomas. 1994. Mesoamerican calques. In Carolyn J. MacKay & Verónica Vásques (eds.), *Investigaciones Lingüísticas en Mesoamérica*, 15–50. México: D.F. Universidad Nacional Autónoma de México.
- Stolz, Thomas. 2012. Survival in a niche. On gender-copy in Chamorro (and sundry languages). In Martine Vanhove, Thomas Stolz, Hitomi Otsuka & Aina Urdtze (eds.), *Morphologies in contact*, 93–140. Munich: Akademie-Verlag.
- Swadesh, Morris. 1950. Salish internal relationships. *International Journal of American Linguistics* 16. 157–167.

- Swadesh, Morris. 1952. Lexico-statistic dating of prehistoric ethnic contacts. With special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96. 452–463.
- Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21. 121–137.
- Szeto, Pui Yiu, Stephen Matthews & Virginia Yip. 2019. Bilingual children as “laboratories” for studying contact outcomes: Development of perfective aspect. *Linguistics* 57(3). 693–723.
- Tadmor, Uri, Haspelmath Martin & Taylor Bradley. 2010. Borrowability and the notion of basic vocabulary. *Diachronica* 27(2). 226–246.
- Thomason, Sarah G. 2001. *Language contact*. Edinburgh: Edinburgh University Press.
- Thomason, Sarah G. & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Urban, Matthias. 2012. *Analyzability and semantic associations in referring expressions. A study in comparative lexicology*. PhD Dissertation, Leiden University.
- Wichmann, Søren & Eric Holman. 2009. *Assessing temporal stability for linguistic typological features*. LINCOM Europa: München.
- Wichmann, Søren. 2015. Diachronic stability and typology. In Claire Bower & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 212–224. London: Routledge.
- Wichmann, Søren, Eric W. Holman & Cecil H. Brown (eds.). 2018. The ASJP database (version 18). Available at: <http://asjp.cld.org>.
- Wilcoxon, Frank. 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1(6). 80–83.

Software and packages used

R

R Core Team. 2020. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at: <https://www.R-project.org>.

boot

Canty, Angelo & Brian Ripley. 2020. boot: Bootstrap R (S-Plus) Functions. R package version 1.3–25.

Davison, Anthony C. & David V. Hinkley. 1997. *Bootstrap methods and their applications*. Cambridge: Cambridge University Press.

cluster

Maechler, Martin, Peter Rousseeuw, Anja Struyf, Mia Hubert & Kurt Hornik. 2019. cluster: Cluster Analysis Basics and Extensions. R package version 2.1.0.

data.table

Dowle, Matt & Arun Srinivasan. 2019. data.table: Extension of `data.frame`. R package version 1.12.8. Available at: <https://CRAN.R-project.org/package=data.table>.

ecodist

Goslee, Sarah C. & Dean L. Urban. 2007. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* 22(7). 1–19.

geodist

Padgham, Mark & Michael D. Sumner. 2020. geodist: Fast, dependency-free geodesic distance calculations. R package version 0.0.4. Available at: <https://CRAN.R-project.org/package=geodist>.

geosphere

Hijmans, Robert J. 2019. geosphere: Spherical trigonometry. R package version 1.5-10. Available at: <https://CRAN.R-project.org/package=geosphere>.

ggmap

Kahle, David & Hadley Wickham. 2013. ggmap: Spatial visualization with ggplot2. *The R Journal* 5(1). 144–161.

ggplot2

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.
Kassambara, Alboukadel. 2020. ggpubr: ‘ggplot2’ based publication ready plots. R package version 0.4.0. Available at: <https://CRAN.R-project.org/package=ggpubr>.

ggrepel

Slowikowski, Kamil. 2020. ggrepel: Automatically position non-overlapping text labels with ‘ggplot2’. R package version 0.8.2. Available at: <https://CRAN.R-project.org/package=ggrepel>.

gplots

Warnes, Gregory R., Bolker Ben, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber, Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz & Bill Venables. 2020. gplots: Various R programming tools for plotting data. R package version 3.0.4. Available at: <https://CRAN.R-project.org/package=gplots>.

gridExtra

Auguie, Baptiste. 2017. gridExtra: Miscellaneous functions for “grid” graphics. R package version 2.3. Available at: <https://CRAN.R-project.org/package=gridExtra>.

gtools

Warnes, Gregory R., Bolker Ben & Thomas Lumley. 2020. gtools: Various R programming tools. R package version 3.8.2. Available at: <https://CRAN.R-project.org/package=gtools>.

igraph

Csardi, Gabor & Tamas Nepusz. 2006. The igraph software package for complex network research. *InterJournal, Complex Systems* 1695. Available at: <http://igraph.org>.

irr

Gamer, Matthias, Jim Lemon & Ian Fellows Puspendra Singh. 2019. irr: Various coefficients of interrater reliability and agreement. R package version 0.84.1. Available at: <https://CRAN.R-project.org/package=irr>.

lattice

Sarkar, Deepayan. 2008. *Lattice: Multivariate data visualization with R*. New York: Springer.

lme4

Bates, Douglas, Martin Maechler, Bolker Ben & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

lmerTest

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26.

maps

Original S code by Richard A. Becker, Allan R. Wilks. R version by Ray Brownrigg, Enhancements by. Minka, Thomas P. & Deckmyn Alex. 2018. maps: Draw geographical maps. R package version 3.3.0. Available at: <https://CRAN.R-project.org/package=maps>.

MuMIn

Bartoń, Kamil. 2020. MuMIn: Multi-model inference. R package version 1.43.17. Available at: <https://CRAN.R-project.org/package=MumIn>.

partR2

Stoffel, Martin A., Shinichi Nakagawa & Holger Schielzeth. 2020. partR2: Partitioning R2 in generalized linear mixed models. *bioRxiv*. <https://doi.org/10.1101/2020.07.26.221168>.

raster

Hijmans, Robert J. 2020. raster: Geographic data analysis and modeling. R package version 3.3-7. Available at: <https://CRAN.R-project.org/package=raster>.

rcompanion

Mangiafico, Salvatore. 2020. rcompanion: Functions to support extension education program evaluation. R package version 2.3.26. Available at: <https://CRAN.R-project.org/package=rcompanion>.

rgeos

Bivand, Roger & Colin Rundel. 2020. rgeos: Interface to geometry engine – open source ('GEOS'). R package version 0.5-3. Available at: <https://CRAN.R-project.org/package=rgeos>.

rworldmap

South, Andy. 2011. rworldmap: A new R package for mapping global data. *The R Journal* 3(1). 35–43.

scales

Wickham, Hadley & Dana Seidel. 2020. scales: Scale functions for visualization. R package version 1.1.1. Available at: <https://CRAN.R-project.org/package=scales>.

sjPlot

Lüdecke, Daniel. 2020. sjPlot: Data visualization for statistics in social science. R package version 2.8.6. Available at: <https://CRAN.R-project.org/package=sjPlot>.

standardize

Eager, Christopher D. 2017. standardize: Tools for standardizing variables for regression in R. R package version 0.2.1. Available at: <https://CRAN.R-project.org/package=standardize>.

vegan

Oksanen, Jari, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Solymos Peter, M. Henry H. Stevens, Eduard Szoecs & Helene Wagner. 2020. vegan: Community ecology package. R package version 2.5-7. Available at: <https://CRAN.R-project.org/package=vegan>.

Supplementary Materials: <https://doi.org/10.5281/zenodo.4597526>.