Natalia Levshina*

European analytic causatives as a comparative concept: Evidence from a parallel corpus of film subtitles

DOI 10.1515/flin-2015-0017
Submitted November 13, 2014; Revision invited March 9, 2015;
Revision received March 19, 2015; accepted June 21, 2015

Abstract: This study investigates formal and functional variation in analytic causatives (ACs) in eighteen European languages from the Indo-European and Uralic language families. Employing the comparative concept approach, the paper presents a probabilistic semantic map of the main functions of ACs on the basis of a multilingual parallel corpus of film subtitles. This method enables us to detect common dimensions of semantic variation in ACs and to pinpoint cross-linguistic commonalities in the form–meaning mapping. The paper also presents three case studies, which test previous hypotheses about the grammaticalization clines in Romance and Germanic and facts of language contact between German and Slavic languages. The role of language contact is further explored in quantitative analyses that compare how the languages "carve up" the semantic space of causation. The results of this comparison suggest that frequently occurring semantically vague ACs may be regarded as a feature of Standard Average European.

Keywords: analytic causative, comparative concept, semantic map, parallel corpus, multidimensional scaling

1 Aims and methods

This study explores the functions of analytic, or periphrastic,[1] causatives (ACs) in eighteen European languages from different language families (Indo-European and Uralic) and genera (Germanic, Romance and Slavic; Finnic and

1 These terms are often used interchangeably, although there are some exceptions, e.g. Dixon (2000), whose restrictive use of the term "periphrastic causative" is discussed in Section 2.

Corresponding author: Natalia Levshina, Gonsbachblick 33, 55122 Mainz, Germany,
E-mail: natalevs@gmail.com

Ugric). ACs are constructions that consist of two verbal predicates and several nominal slots. As an illustration, consider the following sentence with an AC:

(1) *The magician made the rabbit disappear*.

where *the magician* is the causer, *the rabbit* is the causee, *made* is the causative auxiliary, which represents – in a very abstract way – the causing event, and *disappear* is the effected predicate, which specifies the caused, or effected event. A detailed definition of ACs as a comparative concept for cross-linguistic comparison will be provided in Section 2.

ACs in European languages have been described in great detail. Most studies, however, focus on one language, e.g. contemporary British English (Gilquin 2010), Dutch (Verhagen and Kemmer 1997), Estonian (Tamm 2012), Finnish (Leino 2012), French (Achard 1993), German (Nedjalkov 1976), Late Old English (Lowrey 2013) and Swedish (Rawoens 2012). Some studies compare ACs or causative auxiliaries from one genus, e.g. Romance (Soares da Silva 2012) and Slavic (von Waldenfels 2012b), and a few compare ACs across different genera and families (e.g. Degand 1996; von Waldenfels 2012a). However, a large-scale comparative study of European ACs has not been performed yet, to the best of our knowledge. The present paper aims to fill in this gap, contributing to the typology of European causative constructions.

In order to be able to compare any constructions cross-linguistically, one has to establish the main semantic dimensions and functions that are expressed by the constructions and which will serve as *tertia comparationis*. One of the most difficult problems in the cross-linguistic comparison of causative constructions is the great number of potentially relevant semantic distinctions that have been suggested in the literature; causation may be direct or indirect; manipulative or directive; factitive or permissive; volitional, inducive, affective or physical, etc. Many of these categories overlap. It is therefore difficult, if not impossible, to anticipate and operationalize all potentially relevant distinctions for the purpose of constructing an etic grid for cross-linguistic semantic comparison (see Evans 2010). Instead, this paper follows the statistical bottom-up approach proposed by Wälchli and Cysouw (2012) and employs Multidimensional Scaling (MDS) to create a probabilistic token-based semantic map of ACs on the basis of a multilingual parallel corpus, where semantic situations of causation serve as low-level comparative concepts. In this approach, the relevant semantic distinctions can be represented as dimensions in Euclidean space, and semantic categories can be identified as clusters of instances of causative events. This does not mean that the old distinctions are not relevant: we simply avoid making any *a priori* assumptions.

The probabilistic semantic map approach will also help us find out whether some ACs in Germanic and Romance fit the grammaticalization clines proposed in the literature. For example, Lamiroy (2011) proposes two clines representing the degree of grammaticalization for a number of Romance and Germanic languages, based on a comparison of different grammatical phenomena. The cline for Romance languages looks as follows:

(2) French > Italian > Spanish

This means that French is the most grammaticalized language and Spanish is the least grammaticalized one. However, there has been an alternative hypothesis regarding the degree of grammaticalization of causative auxiliaries in Romance. According to Soares da Silva (2012), Italian *fare* is the most grammaticalized causative auxiliary, followed by its French, Spanish and Portuguese cognates, as shown in (3):

(3) Italian > French > Spanish > Portuguese

This order is determined by the degree of semantic bleaching of the causative auxiliary and syntactic integration of the predicates in the corresponding ACs.

In Germanic, the proposed cline looks as follows:

(4) English > Dutch > German

This continuum is also known as the so-called Germanic sandwich, which puts Dutch between English and German with regard to a range of synchronic phenomena, from phonology to word order (van Haeringen 1956). This order finds support in previous contrastive work on verbs of letting in Germanic (Levshina in press b). In particular, German *lassen*, followed by Dutch *laten*, is more frequently used as a lexical verb than its English, Danish, Norwegian and Swedish counterparts. The present study tests if the functions of the ACs in Romance and Germanic differ in a way that resembles these clines and whether these clines can be extended by including more languages.

In addition, the probabilistic semantic map method is employed to test the predictions based on facts of extensive language contact between Germanic and Slavic. The Slavic ACs with verbs of giving have been investigated in detail by von Waldenfels (2012b, 2015). In particular, he observes that the Slavic languages in the West (e.g. Czech) and South West (e.g. Slovenian) have developed more general causative functions in comparison with the East Slavic and South Slavic languages (e.g. Russian and Bulgarian, respectively). It is hypothesized

that the German causative auxiliary *lassen* provides an important model in that respect (von Waldenfels 2015: 125). For example, all ACs with verbs of giving in the Slavic languages where German has played an important role express factitive causation (i.e. "X makes or has Y do Z"), similar to *lassen* (von Waldenfels 2015: 118). The present study puts this hypothesis to the test, comparing the ACs with verbs of giving in five Slavic languages (Bulgarian, Czech, Polish, Russian and Slovenian) and investigating to what extent their range of functions overlaps with that of *lassen*.

To explore the role of language contact on a larger scale, this paper investigates the differences and similarities in the way the ACs in the eighteen languages "carve up" the semantic space of causation. First, the type–token ratios of ACs in the eighteen languages are compared in order to identify which languages are "lumpers" (i.e. those that have relatively few formally distinct ACs to express various causative meanings) and which are "splitters" (i.e. those that have many different ACs to express the causative meanings), based on the data from the parallel corpus. Second, another MDS analysis is performed to explore which of the languages are similar to one another with regard to the "division of labour" between their ACs (i.e. how the languages "carve up" the semantic space of causation). This glottometric approach enables us to examine whether languages cluster according to their genealogical relationships, or whether there are any patterns that can be best explained by language contact. The results allow us to formulate a hypothesis about frequent and semantically vague ACs as a distinctive feature of Standard Average European (SAE), which could be added to the list of SAE features proposed in the literature (Haspelmath 2001).

All analyses in the paper are based on a parallel corpus of film subtitles in eighteen languages. These data have not been used before in functionally oriented typology and areal linguistics, to the best of our knowledge. Film subtitles provide an attractive alternative to other parallel corpora, such as translations of the Bible, fiction, legal documents and European Parliament proceedings, because the language of film subtitles is very similar to spontaneous informal conversations (Levshina In press a). Moreover, subtitles are freely downloadable from online repositories and are available for many typologically diverse languages.

The structure of the paper is as follows. Section 2 proposes a definition of ACs as a comparative concept (Haspelmath 2010), which will be used for identification of instances of ACs in the data. Section 3 describes the multilingual parallel corpus of film subtitles and the procedure of data extraction. A probabilistic semantic map is presented in Section 4, followed by a set of small case studies, which compare particular ACs in Romance, Germanic and Slavic. Section 5 investigates areal patterns in the form–meaning mapping of the

European ACs by using a quantitative index of "lumping" and "splitting". Section 6 provides the glottometric analysis of the languages with regard to their ACs and discusses the role of genealogy and language contact. Finally, Section 7 offers a summary of the results.

2 ACs as a comparative concept

The present paper follows the comparative concepts approach, as recently formulated by Haspelmath (2010); this approach is not rooted in any particular linguistic theory. Comparative concepts are created by linguists for the purpose of language comparison. Such concepts do not have to correspond to any language-specific categories and do not have to be part of the linguistic system of any language speaker;[2] the only criterion is their usefulness for cross-linguistic comparison.

An ANALYTIC CAUSATIVE as a comparative concept can be defined in the following way:

> ANALYTIC CAUSATIVE (AC)
> FUNCTION: An AC designates a causative event, which involves a causing event (or state) and a caused event (or state), and their participants, most importantly, the causer and the causee. The causer initiates or is responsible for the causing event, whereas the causee is the entity that brings about the caused event (state). There can be also other participants involved (such as the affectee, i.e. the final affected entity, cf. Kemmer and Verhagen 1994). The causing event is underspecified.
>
> FORM: An AC is a construction that consists of two VERBS and their arguments. One VERB (V1) represents in an abstract way the causing event, whereas the other VERB (V2) represents the caused event. The order of the predicates may vary. The clauses should be closely integrated: at least some arguments of V2 should be grammatically dependent on V1.[3]

ACs as they are defined here include both so-called complex predicates and periphrastic causatives, following Dixon's (2000) classification. An example of a complex predicate is French *faire* + Vinf (and the corresponding constructions in Italian and Spanish), where V1 is immediately followed by V2. The causee can

2 This does not mean that one cannot successfully use a language-specific construction as a comparative concept as well (van der Auwera and Sahoo In press).

3 VERB can be defined, following the well-known prototype approach by Croft (1991), as a representative of the word class that typically expresses predication (vs. reference and modification) and action (vs. objects and states). The concept WORD is notoriously difficult to define, but, at least for the languages in question, it can be defined as an autonomous meaningful unit, which represents formally "a segment string that cannot be interrupted by a free form without changing its meaning" (Haspelmath 2010: 666).

be expressed by a noun phrase (NP) or prepositional phrase (PP) after V2 or by a clitic pronoun before V1, but can never appear between V1 and V2. In contrast, in periphrastic causatives, such as English *make* + Vinf and *cause* + *to* Vinf or Portuguese *fazer* + Vinf, the causee can come between V1 and V2. Dixon's (2000) classification closely corresponds to the distinction between monoclausal and biclausal ACs, which rests on several criteria, such as passivization or the ability of V1 to have its own arguments. Many authors regard only monoclausal constructions as "proper" ACs (see Kulikov 2001: 887). In our view, an all-encompassing definition of ACs presents some theoretical and descriptive advantages over approaches making a distinction between monoclausal and biclausal causatives. According to Kulikov (2001: 887), and in line with our own experience, this distinction is not unproblematic, in that causative constructions often display degrees of syntactic fusion making it difficult to neatly distinguish between monoclausal and biclausal causative constructions. The present approach covers the entire continuum, thus avoiding the necessity to make arbitrary decisions about a cut-off point. From a descriptive perspective, it enables us to include historically related constructions in different languages that exhibit different levels of syntactic integration. Consider the above-mentioned French AC *faire* + Vinf and the corresponding constructions in Italian and Spanish, where V1 is immediately followed by V2. These constructions are more closely integrated than the Portuguese construction with *fazer* + (X) + Vinf, where the causee can be positioned between V1 and V2, and the Romanian causative *a face* + *să* Vsubj, where V2 is expressed by a finite form in the subjunctive.[4] However, it would be unwise to describe the French, Italian and Spanish constructions and exclude the Portuguese and Romanian constructions from analysis, since all these constructions are closely related.

It should be noted, however, that traditional classifications may still prove relevant. If there are systematic functional differences between mono- and biclausal ACs, or any other types of ACs known in the literature, these differences will find their expression as distinct clusters or dimensions of variation on the probabilistic semantic map.

Further examples of ACs from different languages are provided in Table 1, which focuses on the expression of the causing event (V1), the caused event (V2) and the type of linking element between two verbs (Linking element). This table

4 The finite forms are the only available option in the languages of the Balkan linguistic area, represented in our sample by Bulgarian and Romanian. This is a characteristic feature of the Balkan Sprachbund, where infinitive forms are either non-existent or demoted (Friedman 2006: 665). Although the infinitive exists in Romanian, it has a more restricted range of functions in comparison with the other Romance languages.

Table 1: Formal variation in AC constructions in European languages.

V1	Linking element	V2
verb of making or doing (Fr. *faire* 'make, do' + Vinf, Du. *doen* 'do' + Vinf)	zero (It. *fare* 'make, do' + Vinf)	uninflected infinitive (Hun. *hagy* 'let' + Vinf ending in *-ni*, Est. *laskma* 'let' + *–da*-infinitive)
verb of letting or leaving (Sp. *dejar* 'let' + Vinf)	preposition (Por. *forçar* 'force' + *a* 'to' Vinf)	
verb of coercion (Rus. *vynuždat'* 'force, compel' + Vinf)	infinitive marker historically related to a preposition of location or goal (En. *cause* + *to* Vinf, Ger. *erlauben* 'allow' + *zu* Vinf)	inflected infinitive with personal endings in agreement with the causee (e.g. Por. *fazer eles correr**em*** 'make them run')
verb of caused motion (Ger. *bringen* 'bring' + *zum* Vinf, Fin. *panna* 'put' + Vill)	a combination of the previous two (Nor. *få* 'get' + *til* 'to' + *å* 'at, to' Vinf)	inflected infinitive in a case form expressing goal or destination (Fin. *panna* 'put' + Vill)
verb of possession (having or getting) (En. have + Vinf, get + Ving)	complementizer (Bul. *davam* 'give' + *da* 'that' + Vfin)	present or past participle (En. have + Ved, get + Ving)
verb of giving (Cz. dávat[5] 'give' + Vinf)	modal particle (Sl. *puščati* 'let' + *naj* + Vfin)	finite form, subjunctive or indicative (Rom. *a face* 'make' + *să* Vsubj, Bul. *karam* 'drive' + *da* 'that' + Vind)

illustrates the diversity of constructions that correspond to our definition of ACs as a comparative concept.

Since the definition of ACs given above is very broad, the reader may wonder which constructions are *not* ACs. Examples of constructions that do not meet the criteria are the following:

– Constructions that represent the causing event specifically, not in an abstract way. An example is the construction *order* + *to*-Vinf, where the causer makes a particular speech act (order), as in (5). A similar case is observed in (6), which is an example of the English *into*-causative:

(5) *The general ordered the troops to shoot.*

(6) *He talked her into eloping with him.*

5 Slavic verbs are given in the imperfective form, when it is available.

– Constructions where the arguments of V1 and V2 do not overlap, e.g. Russian *(s)delat' tak, čtoby* … 'to do (so) that', as in (7), or Spanish *dejar* 'let' + *que* 'that' Clause, as in (8):

(7) *Ja sdelal tak, čtoby on byl postojanno rjadom.*
I did so COMP he was always near
'I made it so that he was always near.'

(8) *Déjà que ella camin-e entre nosotros…*
let.IMP.2SG COMP she walk-SBJV.3SG among us
'Let her walk among us…'

– Constructions where the caused event is not expressed by a verb, e.g. *force* + NP, as in (9):

(9) *I need to know how to force their cooperation.*

– Adhortative uses of ACs of letting in Germanic (e.g. *let us go!*). These constructions will not be regarded as instances of ACs because the causative meaning is no longer perceived, at least in our data. The same holds for the Hungarian imperative particle *hadd* derived from *hagy* 'let', where the original permissive meaning can no longer be perceived either (Péteri 2012: 445).

Although all these constructions are semantically and formally related to ACs, they will be excluded from our study.

3 The data

3.1 The parallel corpus

The data used in this study represent a part of a self-compiled corpus of film subtitles called ParTy (a Parallel corpus for Typology).[6] The files with subtitles were downloaded from different online repositories. To the extent possible, the influence of the original language was neutralized by selecting films in different original languages. The films were the following:

6 The corpus is available at http://www.natalialevshina.com/corpus.html.

- *Avatar*: epic science fiction, USA, 2009
- *Das Leben der Anderen* (*The Lives of Others*): drama, Germany, 2006
- *El laberinto del fauno* (*Pan's Labyrinth*): dark fantasy. Mexico/Spain, 2006
- *La vita è bella* (*Life Is Beautiful*): tragicomedy/drama, Italy, 1997
- *Le fabuleux destin d'Amélie Poulain* (*Amélie*): romantic comedy, France, 2001
- *The Tourist*: romantic thriller, USA, 2010
- *Twilight*: vampire romance film, USA, 2008

Each of the films was represented by subtitles in eighteen languages. The total size of the subcorpus was approximately 670,000 words. The choice of films was motivated by the availability of subtitles in all selected languages. The files were in the .srt format and contained the captions and the times when each caption should appear on, and disappear from, the screen. These files were transformed into XML format and sentence-aligned with the help of software "subalign" (Tiedemann 2012).

3.2 Identification of ACs in the corpus

Instances of ACs were identified semi-automatically in every language in the sample. Since there exists no standard list of ACs for all languages under investigation, the procedure was as follows. First, we retrieved the contexts with those ACs that have been described in the literature or were available from personal knowledge. Next, their equivalents were searched for in all other languages. If these equivalents met the criteria of ACs given in Section 2, the list of ACs was updated, and the equivalents of these new ACs were searched in all other languages, until no new ACs were found.

The resulting list of ACs found in the corpus is presented in Appendix 1. The total number of AC tokens was 1,253. The total number of multilingual aligned sentences where at least one language contained an instance of an AC was 392. Each of these sets of multilingual sentences express one underlying semantic situation of causation. These situations serve as low-level implicit comparative concepts, in addition to the explicitly defined comparative concept of an AC.

Figure 1 shows the frequencies of AC tokens for each language. On average, the Romance and Germanic languages have more AC tokens than the Slavic and Finno-Ugric languages. Within the Romance genus, French and Italian have more AC instances than Spanish, Portuguese and Romanian. In the Germanic genus, the differences are less outspoken, with English having the highest number of ACs and Swedish the lowest. Among the Slavic languages, Czech has the highest token frequency of ACs, followed by Bulgarian at a distance,

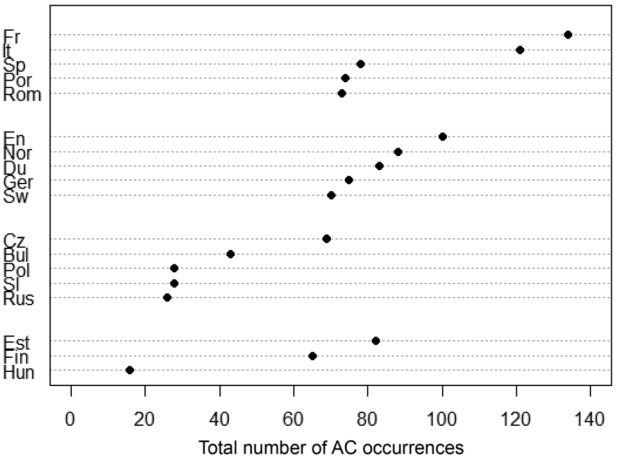

Figure 1: Frequencies of AC tokens in the data set.

whereas Polish, Slovenian and Russian have relatively low frequencies. Finally, Estonian and Finnish have moderate AC token frequencies, whereas Hungarian has the lowest number of occurrences in the corpus.[7]

3.3 Coding of the AC instances

For a finer-grained analysis of the constructions, all instances of ACs in all languages were coded for the following features:

- V1, e.g. *faire*, *make* or *lassen*.[8]
- Encoding of the caused event: e.g. bare infinitive, prepositional infinitive, inflected infinitive or finite form.
- Transitive or intransitive AC (Kemmer and Verhagen 1994); this distinction corresponds to the presence or absence of the affectee, i.e. the final affected entity, which is defined as an NP marked as a direct object. ACs with clausal complements (e.g. *Make them believe it's me*) were treated as a separate type.

7 Interestingly, Uiboaed et al. (2013) demonstrate that ACs, and analytic verbal constructions in general, are more frequently used in western Estonian dialects, which also had strong contacts with Swedish and are more malleable than the conservative eastern dialects. Unfortunately, the present study does not take into account intralinguistic variation.

8 Reflexive/middle voice forms were treated as the same V1 as the corresponding non-reflexive forms (i.e. French *faire* and *se faire* were considered one V1). In Slavic, perfective and imperfective verb form pairs were coded as the same V1 if they had the same lexical meaning and were represented under one entry in the dictionaries, e.g. Russian *zastavit'* 'to make, force (PFV)' and *zastavljat'* 'to make, force (IPFV)' were treated as the same V1.

- The presence or absence of coreferentiality of the causer with another participant of the causation event, e.g. German *Lass dich nicht erwischen!* 'Don't be fooled'.
- Active or passive form of V1, e.g. *They forced me to stay* vs. *I was forced to stay*.

As an illustration, consider example (10):

(10) English, *The Tourist*
And get the office to send me an interpreter.

This example has the following formal features: the V1 is *get*; the caused event is expressed by a *to*-infinitive (*to send*); the construction is transitive (*an interpreter* is the affectee). There is no coreferentiality of the causer (i.e. the addressee of the utterance) with any other participants. Finally, the form of V1 is active.

4 Probabilistic semantic map of European ACs

4.1 Statistical procedure

The data matrix contained eighteen columns for each of the languages and 392 rows for the aligned sentences containing at least one AC in one language. The cells contained the instances of ACs coded for all features described above, e.g. *get_toVinf_Transitive_NoCoreferentiality_Active*. If no AC was found in a translation, the cell contained "NA". The hapax legomena (i.e. rows where only one language contained an AC) were discarded in order to ensure the reliability of comparisons. As a result, about 16.1% of all AC tokens were removed. Note, however, that the hapax legomena were taken into account in the analyses presented in Sections 5 and 6.

Next, a cross-linguistic similarity score was computed for every possible pair of rows in the data matrix. The similarity scores were calculated on the basis of the overlap of the ACs' formal features in each pair across the languages. The principle is similar to the approach in Wälchli and Cysouw (2012): if all formal features were identical, the rows sum similarity score increased by 1; if all formal features were dissimilar, the sum similarity score remained the same. After that, the sum similarity score of the rows was divided by the number of comparisons possible (missing values disregarded). When only some formal features overlapped, as in *get_toVinf_Intransitive_NoCoreferentiality_Active* and *get_toVinf_Transitive_Coreferentiality_Active*, different weights were assigned to different features with the help of a special R script written by the author for

that purpose. The resulting matrices with similarity scores between the rows in the original matrix were then analysed with the help of the *smacof* package (de Leeuw and Mair 2009), which performs MDS by using an iterative majorization approach.

In principle, the data allow us to create an infinite number of possible probabilistic semantic maps, if we take into account different formal features and/or give them different weights. We have tried out a dozen different weight combinations. Due to space limitations, we will only discuss the one that led to results that could best be interpreted from a semantic perspective. The weights for this solution were as follows: 0.7 for an overlapping V1, and 0.1 for each overlap of the following three features: the form of V2, transitivity and coreferentiality. The active/passive distinction has not proven to be helpful due to a very low frequency of passive ACs. The stress (the standard goodness-of-fit measure in MDS) of that two-dimensional model was 0.109, which is a reasonable value. Adding more dimensions did not seriously improve the quality of the representation, judging from a diagnostic scree plot, nor did it yield any new interpretable dimensions of semantic variation.

4.2 A probabilistic semantic map: main dimensions and senses of ACs

Figure 2 demonstrates the two-dimensional MDS map that will be discussed in the remaining part of Section 4. In probabilistic semantic maps, the similarity between

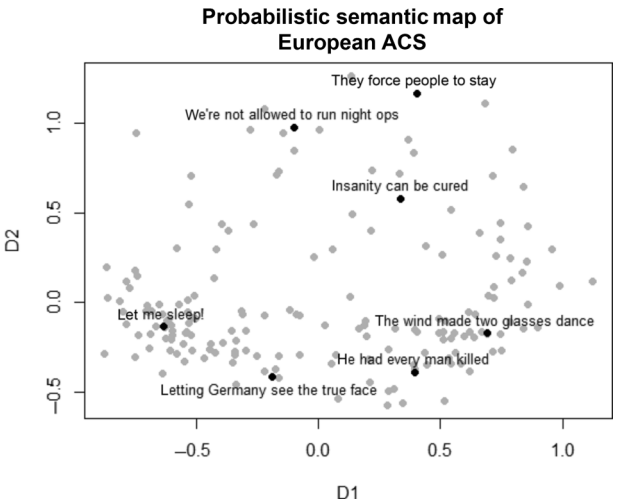

Figure 2: A two-dimensional probabilistic semantic map of ACs in eighteen European languages with selected examples. The text labels correspond to examples (11)–(17) in the text.

instances of linguistic constructions is represented by the distances between them in Euclidean space. Since the similarity scores are based on the cross-linguistic similarity of formal features, the closer two points are on the map, the more often the corresponding causative situations were encoded by the same or highly similar constructions in the eighteen languages. This means that closely located points represent situations that are encoded as similar by the speakers (here, the creators of subtitles) in most languages, whereas points located at a distance represent causative events that are encoded differently in most languages.

When interpreting an MDS map, one should try to interpret the dimensions of the map and of the clusters of points. The horizontal dimension corresponds to the distinction between the constructions "letting X do Y" (left) and "making or having X do Y" (right), each represented by a dense cluster with several points in between. The interpretation of the vertical dimension depends on the section of the plot. The bottom segment on the left contains many examples of letting as non-interference or cessation of impingement, in Talmy's (2000: Ch. 7) terms, as in (11):

(11) English, *The Lives of Others*
Please, ***let*** *me* ***sleep****!*

The upper left section of the map contains examples of permission or authorization, when the causer is "a socially acknowledged authority with the moral power to forbid/permit" (Soares da Silva 2007: 174), as in (12):

(12) English, *Avatar*
We're not ***allowed to run*** *night ops, orders*.

The bottom right sector of the map contains "normal", non-coercive making, as in (13), whereas the top right sector contains a few instances of coercive making, where the causer overcomes the causee's resistance, as in (14):

(13) English, *Amélie*
At that moment, on a restaurant terrace nearby the wind magically ***made*** *two glasses* ***dance*** *unseen on a tablecloth*.

(14) English, *Life Is Beautiful*
What do you think, they ***force*** *people* ***to stay***?

Coercive causation and permission both express less direct causation than simple making and letting, respectively, since the former two normally incorporate an agentive causee, whose presence implies that causation is indirect

(Haiman 1983: 784).[9] Accordingly, the vertical dimension can also be interpreted in terms of the more and less direct causation, or of the degree of conceptual integration of the causing and caused events. It can also be interpreted as the continuum of mono- and biclausal ACs because, as we will see in Section 4.3, the conceptual integration of the causing and caused events appears to correlate with the formal integration of V1 and V2.

In addition, there are two small clusters with distinct semantics: curative causation, as in (15), and middle voice, as in (16):

(15) English, *The Tourist*
Once, he bragged to Alexander that he ***had*** *every man* ***killed*** *his wife had slept with before she met him.*

(16) Czech, *Avatar*
Uvidíme, jestli se tvé bláznovství ***dá*** ***vyléčit.***
see.PFV.1PL if REFL your insanity give.PFV.3SG cure.INF.PFV
'And we will see if your insanity can be cured.'

The area in the middle between the clusters of simple letting and making contains a few situations that can be construed as either letting or making causation, e.g. those representing showing and telling, as in (17). Such examples, however, do not form a coherent cluster.

(17) English, *The Lives of Others*
To ***letting*** *all of Germany* ***see*** *the true face of the GDR!*

4.3 Form–meaning mapping: general tendencies

This section describes the general tendencies in the distribution of ACs and their formal features on the semantic map presented earlier. The languages exhibit great variation in the way their ACs populate the semantic space. Compare, for example, Figures 3 and 4, which demonstrate the distribution of ACs in English and Hungarian. For convenience, only V1 and the form of V2 are displayed.

The English ACs are much more numerous than the Hungarian ones, in terms of both tokens and types. This may be attributed to the fact that other strategies

9 A terminological remark is due here. Although the more established terms for the types of causation represented by the horizontal dimension are permissive and factitive causation, we prefer to use the labels "letting" and "making", respectively, in order to avoid confusion. Since permission as social sanctioning is a special type of letting, it would be misleading to use the label "permissive causation" to designate the entire domain of letting.

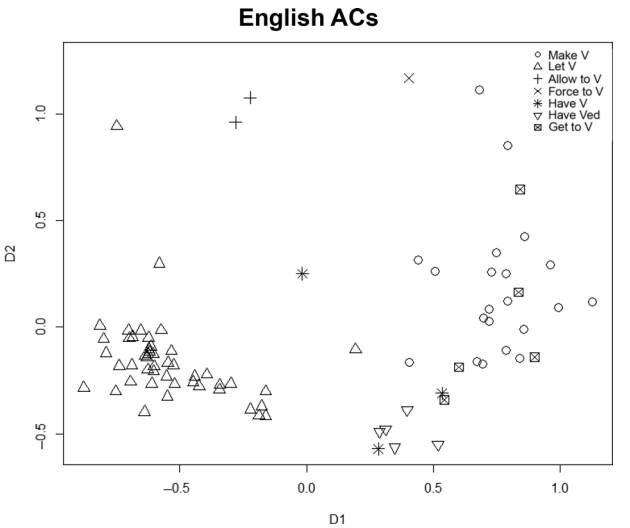

Figure 3: Distribution of ACs in English.

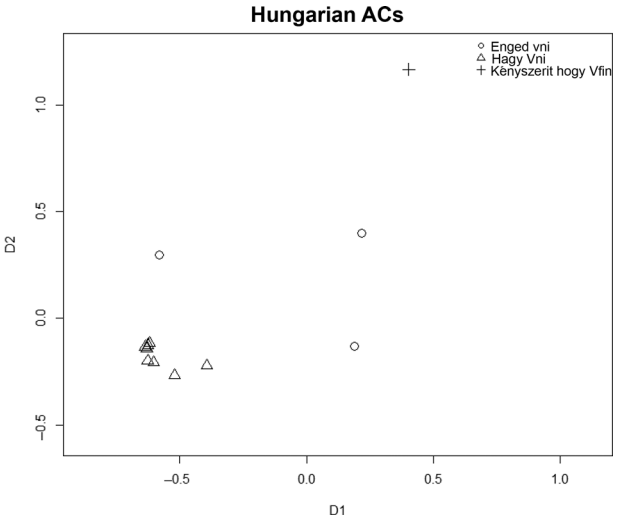

Figure 4: Distribution of ACs in Hungarian.

are available in Hungarian for the expression of causation, most importantly, morphological and lexical causatives, as well as modal verbs, modal suffixes and the imperative particle *hadd*, which was mentioned in Section 2.

Figure 5 displays the same semantic space as in Figure 2, but this time the matrix rows are represented by bubbles. The size of the bubbles corresponds to the number

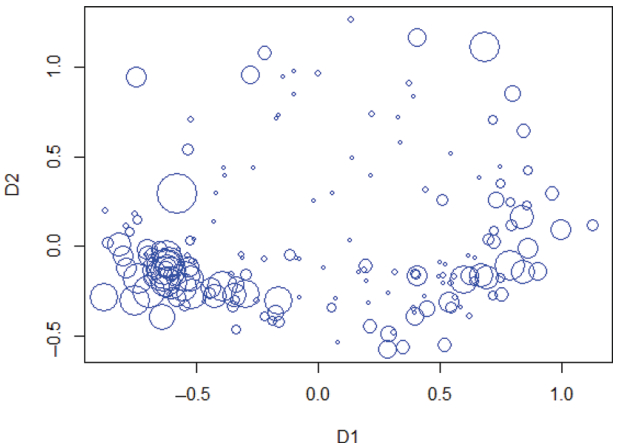

Figure 5: Probabilistic semantic map with multilingual exemplars represented as bubbles. The size of the bubbles reflects the number of ACs in a given matrix row.

of languages where an AC is used. The more languages have an AC in a given row, the larger the bubble. A small bubble means that the subtitle translators selected other linguistic means (other causative constructions or paraphrases) to describe the corresponding situation. It is remarkable that the area of simple (i.e. non-permissive) letting has the highest density of points and these points are represented mostly by large bubbles. Therefore, letting causation should be considered the cross-linguistic semasiological prototype of ACs in the European languages. Simple letting is followed by simple making, which is less densely populated and contains, on average, smaller bubbles. Forceful causation and permission are the least semasiologically salient functions, since they are represented only by a few examples.

It is surprising to find that letting is the most common function of the ACs, as most previous studies of causatives in European and other languages focus on ACs of making. This lower frequency of the ACs of making may be attributed to the fact that they have to "compete" with other causative constructions, such as lexical causatives, as in (18b), or with morphological causatives in the Finno-Ugric languages, as in (19b), which can also be used to express the same meaning as the ACs of making (see (18a) and (19a)).

(18) a. French, *Amélie*
Amandine Poulain aime: (...) ***Faire briller*** *le parquet*
Amandine Poulain likes make shine the parquet
avec des patins.
with ART slippers
'Amandine Poulain likes: (...) polishing the parquet with slippers.'

b. English, *Amélie*
Amandine Poulain likes: (...) **polishing** the parquet...

(19) a. French, *Amélie*
le vent s'engouffrait comme par magie sous une nappe,
the wind REFL.plunged like by magic under a tablecloth,
***faisant danser** les verres...*
making dance the glasses
'The wind crept under a tablecloth, as if by magic, making the glasses dance.'
b. Finnish, *Amélie*
*tuuli **tanssi-tt-i** kahta las-ia*
wind dance-CAUS-PST.3SG two glass-PART.SG
pöytäliin-alla...
tablecloth-AD.SG
'The wind made two glasses dance on the tablecloth.'

Onomasiological variation in the domain of letting is more modest, although it exists. For example, English *let* + *go* is similar semantically to the lexical causative *release*, *let* + *be* has an alternative *leave (alone)* and *let* + *fall* is similar to *drop*.

Let us now explore how specific ACs are mapped onto the semantic space and related to the main functions. ACs in all languages in the sample make some sort of distinction between making and letting. The main types of ACs that express simple letting and are located in the corresponding region on the map involve the following conceptual metaphors, whose metaphorical meaning can be traced etymologically in the semantics of V1 (see Soares da Silva 2007):

- LETTING IS RELEASING and/or LEAVING, e.g. Bulgarian *ostavjam* + *da* Vfin, *puskam* + *da* Vfin, Czech *nechávat* + Vinf or *aby* Vfin, Dutch *laten* + Vinf, French *laisser* + Vinf, Finnish *päästää* + Vill, Hungarian *hagy* + Vinf, Russian *pozvoljat'* + Vinf, Slovenian *puščati* + Vinf or *naj/da* Vfin, Spanish *dejar* + Vinf.
- LETTING IS GIVING, e.g. Estonian *andma* + Vda, Finnish *antaa* + 1st form of the infinitive, Russian *davat'* + Vinf, where the causer is conceptualized as the giver, the causee as the beneficiary and the caused event as the object of transfer.

The V2 in ACs of letting usually have a "default" form (e.g. the bare infinitive in English, the first infinitive in Finnish and *-da*-infinitive in Estonian).[10]

10 An exception is Finnish *päästää* 'let', which is accompanied by the third form of the infinitive in the illative case (Vill).

Simple (i.e. non-coercive) making involves a richer variety of ACs comprising the following conceptual metaphors:

- CAUSING IS MAKING or DOING, e.g. French *faire* + Vinf, Romanian *face* + *să* Vsubj, English *make* + Vinf, Dutch *doen* + Vinf. The caused event is here construed as the product (result) of this activity. The "default" form of V2 is normally used.
- CAUSING IS HAVING, e.g. two English ACs with having verbs *have* + Vinf and *have* + Ved.
- CAUSING IS GETTING, e.g. English *get* + *to* Vinf, Norwegian *få* + *till å* Vinf, Swedish *få* + *att* Vinf, Finnish *saada* + Vill. V2 is marked as the goal or destination.
- CAUSING IS CAUSED MOTION, e.g. BRINGING (German *bringen* + *zum* V), PUTTING (Romanian *a pune* + *să* Vsubj, Estonian *panema* + Vma, Finnish *panna* + Vill) or DRIVING (Bulgarian *karam, nakarvam*, Estonian *ajama*). The caused event is construed as the destination or goal, which manifests itself in the marking of V2. The *-ma*-infinitive, which follows the Estonian auxiliaries of putting and driving, is used with verbs of motion, *inter alia*.
- CAUSING IS GIVING, e.g. Czech *dávat* + Vinf, Estonian *andma* + *Vma*, Slovenian *dati* + Vinf/*da* Vfin.

It appears, then, that (non-coercive) making exhibits more cross-linguistic and intralinguistic variation in terms of different conceptual metaphors.

The distinction between simple letting and permission is quite consistent in most languages. Usually, the languages have special ACs of permission, e.g. French *permettre* + *de* Vinf, *autoriser* + à Vinf, Estonian *lubama* + Vda, Finnish *sallia* + V1st, Hungarian *enged* + Vinf, Norwegain *tillate* + *å* Vinf, German *erlauben* + *zu* Vinf, Czech *dovolovat* + Vinf/Vfin.[11] In the Slavic languages, with the exception of Czech, the verbs of authorization and permission are also widely used in non-authorizing letting senses, so the distinction is somewhat blurred in these languages.

A similar situation is observed with coercive causation: in most languages, one finds special ACs that express this type of causation, e.g. Dutch *dwingen* + *om te* Vinf, English *force* + *to* Vinf, Estonian *sundima* + Vma, Finnish *pakottaa* + Vill, *kiristää* + Vill, Italian *costringere* + *a* Vinf, Norwegian *tvinge* + *til å* Vinf. Many of them have special marking, which is historically related to the

11 In Swedish and Dutch, no special ACs were found in the data – probably, this is due to the low frequency of this semantic type of causation in this corpus.

marking of the location and/or goal and now expresses purpose. Again, the distinction between non-coercive making and coercive causation is blurred in some Slavic languages, since the same ACs are used for both coercive and non-coercive making: Bulgarian *karam/nakarvam* + *da* Vfin, Polish *zmuszać* + Vinf/Vfin and Russian *zastavljat'* + Vinf.

In the special ACs that express permission and coercive causation, V2 predicates tend to be at a greater distance from V1 than those expressing simple letting and making, or at least they are at the same distance. The English ACs provide a good example. The caused events in ACs that designate simple letting and making are expressed by the bare infinitive (*make* + Vinf, *let* + Vinf), whereas the ones in ACs of coercion and permission are expressed by the infinitive with the particle *to* (*force* + *to* Vinf, *allow* + *to* Vinf, *permit* + *to* Vinf). A similar contrast is observed in all other Germanic ACs expressing permission or coercion. The Romance languages (except Romanian) display a similar correspondence, as can be illustrated by the French ACs *forcer* + *à* Vinf vs. *faire* + Vinf; or Italian *permittere* + *di* Vinf vs. *lasciare* + Vinf. In fact, we do not find languages where V2 would be at a smaller distance in ACs of coercion or permission than V2 in ACs of simple letting and making. This may have to do with the fact that coercive causation and permission both express less direct causation, as mentioned in Section 4.2. Therefore, the greater formal distance between V1 and V2 signals, in accordance with Haiman's (1983) principle of iconicity, a weaker degree of conceptual integration of the causing and caused events. On the other hand, our data clearly demonstrate that permission and coercive making are much less frequent than the other types of causation. According to the principle of economy (Haiman 1983; Haspelmath 2008), one could expect rarer constructions to be longer. Which motivation is more important is a question for future research.

4.4 Case studies

The semantic map approach allows for a straightforward cross-linguistic comparison of historically and/or semantically related constructions. This section addresses the hypotheses about specific languages and ACs that were introduced in Section 1. The first case study compares the semantic areas of the ACs with MAKE-auxiliaries in Romance. The second case study focuses on the semantic areas of the ACs with LET-auxiliaries in Germanic. These case studies test the hypotheses about the grammaticalization clines in (2), (3) and (4). The third case study compares the functions of the ACs with verbs of giving in Slavic and their similarity to the functions of the German causative AC with *lassen*.

4.4.1 *ACs with* MAKE*-auxiliaries in Romance*

Section 1 introduced two clines describing the degree of grammaticalization in the Romance languages in general and in the ACs with verbs of making in particular. This case study compares these ACs by mapping the specific constructions onto the semantic space. Figure 6 shows the constructions with French *faire*, Italian *fare*, Portuguese *fazer*, Spanish *hacer* and Romanian *a face*. It is a so-called contour plot, which shows the areas with a high density of exemplars of each AC (the cut-off level of 0.5 is selected for the purposes of visual presentation). The semantic areas of the ACs are located on the map in the following order:

(20) Italian – French – Spanish – Portuguese – Romanian

Semantically, this order corresponds to a decreasing prominence of letting in comparison with making. The map clearly demonstrates that the Italian fare + Vinf covers a significant part of the area that corresponds to letting causation in the left-hand part of the map. Consider the following examples, which illustrate this function of *fare* + Vinf:

(21) Italian, *Avatar*
Stronzate, *fammi* *vedere!*
bullshit make.me see
'Bullshit, let me see that!'

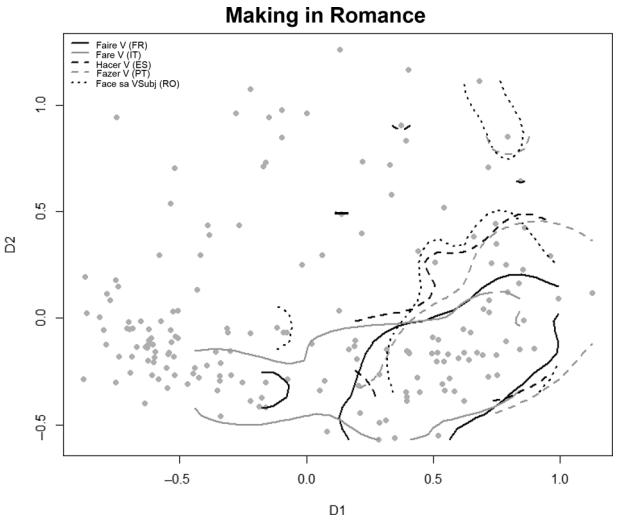

Figure 6: Distribution of ACs with MAKE-verbs in five Romance languages.

(22) Italian, *Life Is Beautiful*
Hanno fermato il treno per far-la salire.
have stopped the train for make-her mount
'They stopped the train to let her get on.'

The order in (20) also corresponds to an increasing strength of causation from Italian to Romanian, which is represented by the vertical dimension. This is not surprising, since the degree of coerciveness and the degree of letting/making neutralization are closely related. If a construction has a strong coercive sense, it is less likely to be used to express letting causation. For example, Portuguese *fazer ver* and Spanish *hacer ver* 'make see' denote stronger causation, something like 'make understand, demonstrate, make consider', in comparison with Italian *fare vedere* 'show, let see', as in (21) (Cerbasi 1997, cited by Soares da Silva 2012: 520–521).

The order in (20) corresponds to the cline in (3) suggested by Soares da Silva (2012), rather than the more general one in (2) proposed by Lamiroy (2011). But do we really observe a cline of grammaticalization, with Italian *fare* as the most grammaticalized causative verb and Romanian *a face* as the least grammaticalized one? The answer seems to be positive. First, the decrease of causation strength from Romanian to Italian can be interpreted as an indication of semantic bleaching of the causative predicate. Second, the Romance ACs exhibit varying degrees of syntactic integration of the causative and effected predicates (Soares da Silva 2012: 529–534), with Italian ACs exhibiting the tightest integration and the Romanian ACs, which can only take a finite clause complement, the loosest. Thus, the degree of independence of the causative auxiliary also increases from Italian to Romanian.

4.4.2 *ACs with LET-auxiliaries in Germanic*

This section compares the ACs of letting in Germanic. The contour plot in Figure 7 shows the areas occupied by each of the ACs. Again, we see some cross-linguistic differences, although they are less clear than with the ACs of making. The German (DE) and Dutch (NL) ACs protrude into the area of making, with German occupying a larger territory. The other three languages only have a few exemplars in that area. Moreover, the German and Dutch ACs are only marginally associated with the function of permission/authorization (top left), whereas the three other ACs do have a few instances in that area.

Thus, the figure provides some evidence in favour of the cline in (4) from English to Dutch and German. In addition, Swedish and Norwegian seem to

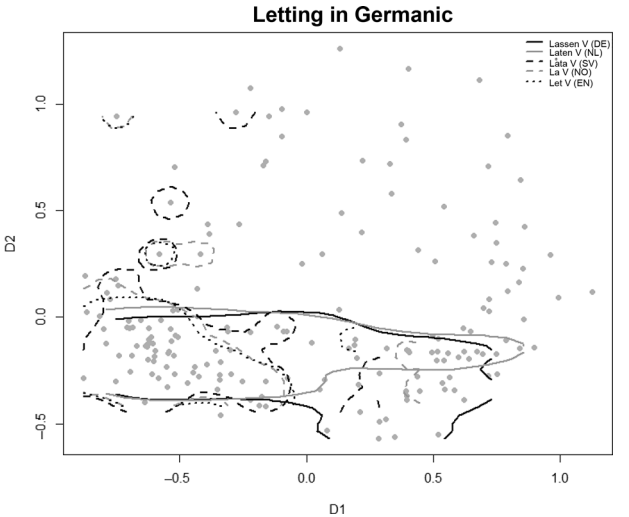

Figure 7: Distribution of ACs with LET-verbs in five Germanic languages.

occupy an intermediate position between English, on the one hand, and Dutch and German, on the other. Can this pattern be interpreted as a cline in grammaticalization of the corresponding verbs, as was the case with the Romance causatives? The question is a difficult one. On the one hand, we know from previous research that English *let* and its Scandinavian cognates have (almost) lost their capacity to be used as lexical verbs, unlike German *lassen* and Dutch *laten* (Levshina In press b). This might serve as evidence of grammaticalization. However, the English and Scandinavian verbs seem to be more semantically specialized than the German and Dutch ones, contrary to what one might expect in a grammaticalization scenario. In fact, there is evidence that making and letting have in fact coexisted in the Germanic ACs of letting since the earliest attestations of the constructions (e.g. Baron 1974; Van der Horst 1998: 62; Rawoens 2013). In addition, the Old English predecessor of the English AC with *let*, *lætan* + Vinf, was used to express both letting and making in Old English (Lowrey 2013), which no longer holds for contemporary English (with the exception of mental causation, e.g. *Let me know* and *Let me see*). Thus, English *let* + Vinf has moved further from the Germanic prototype than its Dutch and German counterparts, becoming less semantically broad. The same, but to a lesser extent, seems to hold for the Swedish and Norwegian constructions, which still retain the function of making. Thus, one can say that English *let* is more grammaticalized than German *lassen* and Dutch *laten*, but the underlying process in the Germanic causative verbs (namely, specialization as an auxiliary

verb) is very different from the ones observed in Romance (namely, semantic bleaching and syntactic integration with the effected predicate).

4.4.3 *ACs with GIVE-auxiliaries in Slavic*

Finally, we compare constructions with the verbs of giving in the five Slavic languages. As Figure 8 demonstrates, the Russian construction *davat'* + Vinf covers only the semantics of letting, as in (22). This extension from the basic sense of verbs of giving (i.e. transfer of possession) can be found in many languages of the world (Newman 1996: 188–194). In the languages under investigation, we also find a similar extension in the Finnish AC with *antaa* 'give' and the Estonian AC with *andma* 'give'.

(23) Russian, *The Lives of Others*
Daj-te *mne*, *požalujsta*, *pospat'*.
give.IMP.PFV-2PL me.DAT please sleep.INF.PFV
'Please, let me sleep!'

In contrast, Czech *dávat* + Vinf and Slovenian *dati* + Vinf/*da*-Clause do not express letting, but do express curative causation (24), middle voice (25) and giving information, as in *Let me know*:

(24) Slovenian, *The Lives of Others*
Dal *ga* *bom* *nadzorovati*.
give.PASTPART him be.FUT.1SG watch.IPFV.INF
'I'll have him monitored.'

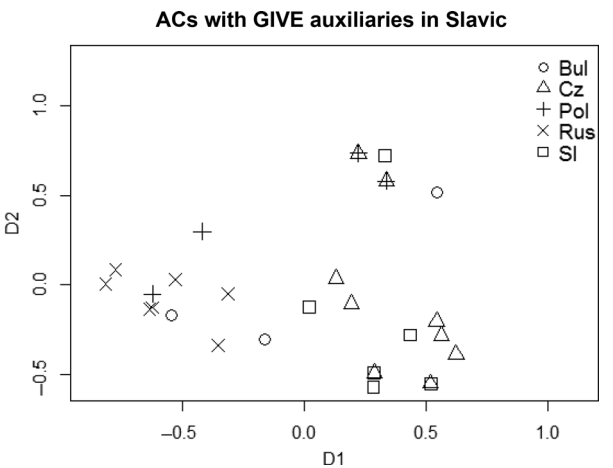

Figure 8: Distribution of ACs with GIVE-verbs in five Slavic languages.

(25) Czech, *Avatar*
Uvidíme, jestli se tvé bláznovství dá vyléčit.
see.PFV.1PL if REFL your insanity give.PFV.3SG cure.INF.PFV
'And we will see if your insanity can be cured.'

Polish *dawać* + Vinf has semantics of letting and middle voice. Bulgarian *davam* + *da* Vfin predominantly expresses letting.

These differences can be explained by language contact: Czech and Slovenian, which, similar to Sorbian and Slovak, have had extended contact with German, are used the most productively to express making (von Waldenfels 2015). The use of GIVE-verbs to express middle voice may be influenced by German as well (cf. *Das lässt sich machen* 'This is doable'). It is interesting, however, that we find no examples of this use in the German subtitles.

4.4.4 Case studies: conclusions

The first case study revealed that the general grammaticalization cline from French to Italian and Spanish does not hold in case of ACs with verbs of making. Instead, we find support for the more specific cline proposed by Soares da Silva (2012), with Italian having the most grammaticalized causative auxiliary. We do find some evidence in favour of the proposed Germanic cline from English to Dutch and German, although the picture is less clear.

A remarkable result is that the distinction between letting and making causation, which corresponds to the first dimension on the map, is blurred in some languages, especially in Dutch, German and Italian. Since these languages belong to the core of the European linguistic area, alongside French (van der Auwera 1998), one may wonder whether this may be an areal property. The similarity of Czech and Slovenian ACs to German, which was discussed in the third case study, lends further support to the importance of language contact in the development of ACs. Sections 5 and 6 explore the areal patterns more systematically.

5 "Lumpers" and "splitters": a hypothesis for SAE

The differences in the distribution of ACs suggest that some languages may be "lumpers" and some may be "splitters" in expressing causation with the help of ACs. "Lumpers" make fewer distinctions between causative situations than "splitters" and therefore should have a smaller number of constructional types.

The types were defined as a combination of V1 with a particular form of the second verb (e.g. *faire* + Vinf or *force* + *to* Vinf). A direct comparison of the number of constructional types was impossible because the number of tokens was very different, from sixteen in Hungarian to 134 in French. A comparison of type–token ratios would be misleading, as well, because the relationship between type and token frequencies is non-linear; namely, an increase in the number of types (new words or constructions) gradually diminishes as the size of a corpus grows. To overcome this obstacle, we retrieved 1,000 random samples with the size of sixteen exemplars (the token frequency of ACs in the Hungarian subcorpus) from all AC tokens in each language, and then computed the average number of types. The results are shown in Figure 9. Note that the Hungarian score is the actual AC-type frequency in the entire Hungarian subcorpus, which was too small to split it into subsamples.

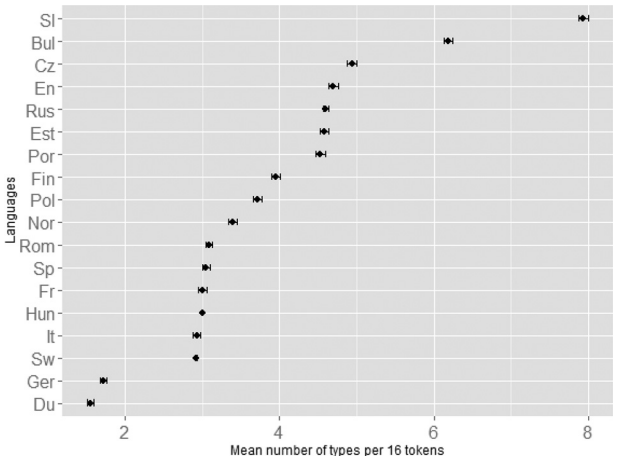

Figure 9: Mean type frequencies of ACs in 1,000 random samples of AC tokens (n = 16), sorted in ascending order, and 95% confidence intervals around the means. For Hungarian, the actual type frequency is given.

The means and the 95% confidence intervals around them show that, on average, the Germanic languages are the greatest "lumpers" and that the Slavic ones are the greatest "splitters". As for the individual languages, the greatest "lumpers" are Dutch and German, followed by Swedish, Italian, Hungarian and French, the languages that – with the exception of Swedish and Hungarian – belong to the so-called Charlemagne Sprachbund, the core of SAE (van der Auwera 1998; Haspelmath 2001). In the Germanic and Romance genera, the more "lumping" languages are located closer to the centre of the

European linguistic area, and the "splitters" are more peripheral geographically. In the Slavic and Finno-Ugric genera, this tendency is less evident. The high-type frequency of Slovenian can be explained by fact that its ACs allows for both non-finite and finite V2 to follow the same V1.[12]

The most important factor that explains these differences seems to be the presence of highly frequent and semantically bleached ACs. In particular, Dutch, German and Italian exhibit a strong neutralization of the distinction between making and letting in the semantics of their most frequent ACs, as shown in Section 4.4.

6 Genealogy or language contact?

As we have seen in the previous sections, there are indications that language contact has played an important role in the development of ACs. How strong is the effect of language contact in comparison with the genealogical relationships between the languages? To answer this question, we computed similarity scores for all pairs of languages. These scores were based on the formal overlap of ACs between all possible pairs of causative situations. For a pair of languages, if a given pair of causative situations had identical ACs (defined as a combination of a specific V1 and a V2 form) in both languages, the sum similarity score increased by 1. In the constructed data in Table 2, this will be the case of Situations 1 and 2. If both situations had dissimilar ACs in each language, as do Situations 2 and 3, the sum similarity score increased by 1 as well. If two situations had identical ACs in one language but different ACs in the other

Table 2: Constructed data to illustrate how similarity scores were computed for pairs of languages.

Situation ID	Language 1	Language 2
1	let_Vinf	lassen_Vinf
2	let_Vinf	lassen_Vinf
3	make_Vinf	bringen_Vinf
4	make_Vinf	lassen_Vinf

12 According to our informant, a native speaker of Slovenian, the functional differences between ACs with different complements may have to do with the aspect and modality. However, more research is needed to understand the factors that constrain this variation.

language, as do Situations 3 and 4, nothing was added to the sum similarity score. The sum similarity score was then divided by the number of possible comparisons. The similarity score serves as a measure of the overlap between the categorization choices of a pair of languages; it shows how often on average two languages "agree" in considering two causative situations similar or different.

After computing the distances by subtracting every similarity score from 1, we performed another MDS. The result can be seen in Figure 10. The closer one language is to another language, the more often these two languages agree on considering causative events to be (dis)similar, or, in other words, the more similarly these languages "carve up" the semantic space. Some patterns seem to reflect the genealogical relationships: there is a distinct Slavic cluster; French and Italian are situated on one left-hand side of the plot; most Germanic languages are in the top area of the map. Yet, the similarities and differences cannot be explained by the genealogical relationships only. Consider the Finno-Ugric languages: Hungarian is close to the Slavic languages, whereas Estonian is closer to German and Finnish is closer to Swedish. In the Slavic cluster, Czech is closer to German than the other Slavic languages. The results are also corroborated by a hierarchical cluster analysis (not shown here).

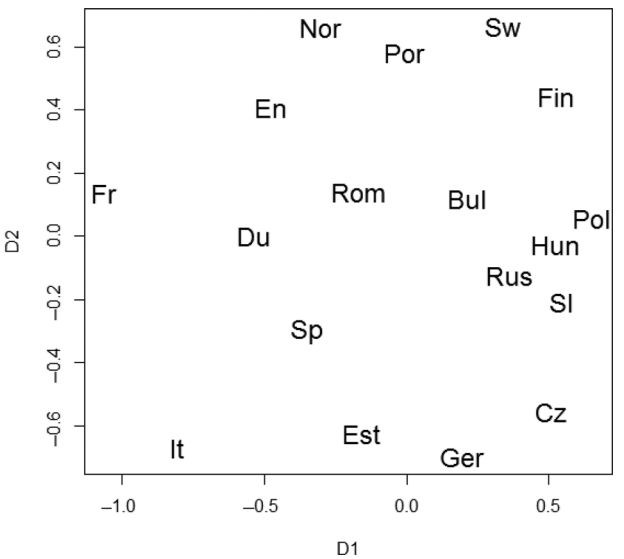

Figure 10: A Multidimensional Scaling plot of eighteen languages, showing how similar the languages are in carving up the semantic space. Stress = 0.078.

The most plausible explanation of these deviations seems to be language contact. This is the case of Czech and German, and of Hungarian and Slavic. The position of Estonian close to German deserves a special mention. Estonian has an AC *laskma* 'let' + Vda, which is distributed very similarly to German *lassen* + Vinf. In particular, it is also quite frequently used in the corpus to express both letting and curative causation (see Tamm 2012). This may be a result of language contact with Germanic. Old Estonian was written mainly by German clergymen. In addition, there was extensive contact with Swedish (Uiboaed et al. 2013: 26).

7 Summary

In this paper, I have proposed a definition of ACs as a comparative concept for cross-linguistic comparison and applied it to a sample of eighteen European languages. Although there were a few borderline cases, in general, the identification of new ACs was straightforward. The definition may have to be modified, of course, when we face new data from more typologically diverse languages.

The bottom-up statistical analyses based on a multilingual parallel corpus have revealed the following. It has been found that there are two main dimensions of variation of ACs in the eighteen languages. The first dimension can be labelled "letting vs. making". The second dimension separates, on the one hand, simple letting from permission or authorization, and, on the other hand, non-coercive from coercive making. This dimension in fact correlates with the degree of conceptual integration of the causing and caused events and formal integration of the predicates, similar to the well-known continuum of mono- and biclausality. The ACs expressing permission and coercive causation tend to have more elaborately marked V2 predicates in comparison with ACs expressing simple letting and making. The more elaborate V2 forms are separated from V1 by additional linking elements, such as particles and prepositions, which suggests a higher degree of formal integration of the predicates. Whether this tendency can be best explained by the principle of iconicity or the principle of economy (see Haspelmath 2008) remains an open question.

Another finding is that simple letting is the most systematically covered by AC tokens in all languages, in comparison with the other semantic regions on the map. Perhaps counterintuitively, this suggests that letting is the prototype of European ACs. The ACs that express making are less onomasiologically salient than letting ACs, since they have to "compete" with lexical and morphological causatives. This may have to do with the relatively low degree of conceptual integration of letting causative events in comparison with events of making

(Givón 1980). In general, the causee of a letting event is more autonomous because the causer fails to override the causee's intrinsic tendency towards some action or state, whereas in making the causer does change the causee's intrinsic tendency (Talmy 2000: Ch. 7). Letting also represents a more formally homogeneous category, whereas making exhibits more intra- and cross-linguistic formal variation. From a semantic perspective, the metaphorical source domains of making are particularly diverse, as well.

Although most languages usually have different ACs for expressing letting and making causation, there are some overlaps. In some Germanic languages, as well as in Italian and French, one can find strong evidence of the neutralization of the distinction between making and letting, since some ACs tend to express both semantic categories, such as German *lassen* + Vinf, Dutch *laten* + Vinf and Italian *fare* + Vinf. The Slavic languages tend to lump together coercive and non-coercive making, on the one hand, and permission and simple letting, on the other hand, although this is more evident in Bulgarian, Polish and Russian than in Czech and Slovenian.

This paper has also tested some grammaticalization clines proposed in the literature for Germanic and Romance. As for Romance, the results corroborate Soares da Silva's (2012) proposal that Italian causative *fare* is the most grammaticalized among the major Romance languages, rather than Lamiroy's (2011) general cline with French as the most grammaticalized language. At the same time, the English, Norwegian and Swedish ACs with verbs of letting are less semantically vague than the corresponding constructions in Dutch and especially German. This supports the idea of the "Germanic sandwich", according to which Dutch is located between English and German, although the evidence is less clear than in the case of Romance. However, the underlying historical processes that have led to these changes are very different. In Romance, we observe different degrees of semantic bleaching and syntactic integration of the causative and effected predicates (the greatest in Italian and the weakest in Romanian). In Germanic, the cline corresponds to the degree of specialization of the verbs of letting as a causative auxiliary (the highest in English and the lowest in German). Whether it is useful to use the umbrella term "grammaticalization" to refer to these very different processes is a question for theoretical debate (cf. Wiemer 2014).

The data reveal substantial variation within the language genera and families, which cannot be explained by the genealogical relationships alone. Such is the similarity between Czech and Slovenian GIVE-ACs and German *lassen* + Vinf in the curative function, or the overlap between Estonian *laskma* 'let' + Vda and German *lassen* + Vinf. There are numerous indications that ACs are very malleable with regard to language contact, which may even override

genealogical similarities between the languages of the same genus or even family, as shown by our glottometric MDS analysis. This supports the predictions based on previous research of ACs in these languages.

Moreover, the average number of constructional types bootstrapped from the data shows that the languages that form the historical core of SAE and belong to the Charlemagne Sprachbund tend to be "lumpers" (i.e. have fewer constructional types), whereas the languages at the periphery are "splitters" (i.e. have more constructional types). This tendency is particularly evident in the Germanic and Romance genera. This seems to be an interesting hypothesis for future investigation. Notably, these quantitative patterns could have been pinpointed only on the basis of corpus frequencies. This demonstrates the importance of corpus evidence for typological investigations and generalizations.

Acknowledgements: The author is very grateful to the anonymous reviewers of this paper for their useful feedback and constructive criticisms and to Hubert Cuyckens for his numerous insightful suggestions. Of course, all remaining imperfections are the author's own responsibility. This study was funded by the Belgian research foundation F.R.S. – FNRS, whose generosity is most gratefully acknowledged. All usual disclaimers apply.

Abbreviations

Descriptive categories

AD = adessive; ART = article; COMP = complementizer; CONJ = conjunction; CNEG = connegative; DAT = dative; FUT = future; IMP = imperative; INF = infinitive; IPFV = imperfective; NEG = negation; PART = partitive; PASTPART = past participle; PFV = perfective; PL = plural; PRS = present; PST = past; REFL = reflexive; SBJV = subjunctive; SG = singular; V1st = (Fin.) 1st infinitive; Vda = (Est.) -da-infinitive; Ved = (En.) past participle; Vfin = verb is a finite form; Vill = (Fin.) 3rd infinitive in illative; Vind = verb in indicative; Vinf = infinitive; Vinfl = (Por.) inflected infinitive; Ving = (En.) present participle; Vma = (Est.) -ma-infinitive; Vsubj = verb in subjunctive

Languages

Bul. = Bulgarian; Cz. = Czech; Du. = Dutch; En. = English; Est. = Estonian; Fin. = Finnish; Fr. = French; Ger. = German; It. = Italian; Hun. = Hungarian; Nor. = Norwegian; Pol. = Polish; Por. = Portuguese; Rom. = Romanian; Rus. = Russian; Sl. = Slovenian; Sp. = Spanish; Sw. = Swedish

Appendix

ACs found in the corpus

Indo-European, Romance

French	Italian	Portuguese	Romanian	Spanish
autoriser à Vinf	*costringere a* Vinf	*dar para* Vinf	*a convinge să* Vsubj	*dejar* Vinf
faire Vinf	*fare* Vinf	*deixar* Vinf	*a face să* Vsubj	*forzar a* Vinf
forcer à Vinf	*lasciare* Vinf	*deixar* Vinfl	*a lăsa să* Vsubj	*hacer* Vinf
laisser Vinf	*permettere di* Vinf	*fazer* Vinf	*a permite să* Vsubj	*obligar a* Vinf
		fazer Vinfl		
obliger à Vinf	*obbligare a* Vinf	*forçar a* Vinf	*a pune să* Vsubj	*permitir* Vinf
permettre de Vinf		*obrigar a* Vinf		
		permitir Vinf		

Indo-European, Germanic

Dutch	English	German	Norwegian	Swedish
doen Vinf	*allow to* Vinf	*bringen zum* Vinf	*få* PastPart	*få att* Vinf
			få til å Vinf	
dwingen om te Vinf	*force to* Vinf	*erlauben zu* Vinf	*la* Vinf	*låta* Vinf
laten Vinf	*get to* Vinf	*lassen* Vinf	*tillate å* Vinf	*tvinga att* Vinf
	have Vinf		*tvinge til å* Vinf	
	have PastPart			
	let Vinf			
	make Vinf			
	permit to Vinf			

Indo-European, Slavic

Bulgarian	Czech	Polish	Russian	Slovenian
davam da Vfin	*dávat* Vinf	*dawać* Vinf	*davat'* Vinf	*dati da* Vfin
karam da Vfin	*dovolovat aby* Vfin	*pozwalać* Vinf	*pozvoljat'* Vinf	*dati* Vinf
nakarvam da Vfin	*nechávat aby* Vfin	*zmuszać* Vinf	*razrešat'* Vinf	*dopuščati* Vinf
ostavjam da Vfin	*nechávat* Vinf	*zmuszać żeby* Vfin	*zastavljat'* Vinf	*dovoliti da* Vfin
pozvoljavam da Vfin	*přimět* Vinf	*zmuszać żeby* Vinf		*dovoliti* Vinf
razrešavam da Vfin	*přinucovat aby* Vfin			*puščati da* Vfin
	přinucovat Vinf			*puščati naj* Vfin
				puščati Vinf
				siliti da Vfin
				siliti Vinf

Uralic, Finno-Ugric

Estonian	Finnish	Hungarian
ajama Vma	*antaa* V1st	*ended* Vinf
andma Vda	*kiristää* Vill	*hagy* Vinf
laskma Vda	*päästää* Vill	*kényszerít hogy* Vfin
lubama Vda	*pakottaa* Vill	
panema Vma	*panna* Vill	
sundima Vma	*saada* Vill	
	sallia V1st	

References

Achard, Michel. 1993. Causative structures in French: Word order following *faire*, *laisser*, and *forcer*. In Joshua S. Guenter, Barbara A. Kaiser & Cheryl C. Zoll (eds.), *Proceedings of the nineteenth Annual Meeting of the Berkeley Linguistics Society: General session and parasession of semantic typology and semantic universals*, 1–12. Berkeley, CA: Berkeley Linguistics Society.

Baron, Naomi S. 1974. The structure of English causatives. *Lingua* 33(4). 299–342.

Cerbasi, Donato. 1997. Las construcciones causativas del typo hacer + Infinitivo en español, portugués e italiano. *Lingüístuca Española Actual* 19. 155–171.

Croft, William. 1991. *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press.

Degand, Liesbeth. 1996. Causation in Dutch and French: Interpersonal aspects. In Ruqaya Hasan, Carmel Cloran & David Butt (eds.), *Functional descriptions: Theory in practice*, 207–235. Amsterdam: John Benjamins.

De Leeuw, Jan & Patrick Mair. 2009. Multidimensional Scaling using majorization: SMACOF in R. *Journal of Statistical Software* 31(3). 1–30. http://www.jstatsoft.org/v31/i03/ (accessed 27 March 2015).

Dixon, R. M. W. 2000. A typology of causatives: Form, syntax and meaning. In R. M. W. Dixon & Alexandra Y. Aikhenvald (eds.), *Changing valency: Case studies in transitivity*, 30–83. Cambridge: Cambridge University Press.

Evans, Nickolas. 2010. Semantic typology. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*, 504–533. Oxford: Oxford University Press.

Friedman, Victor A. 2006. Balkans as a linguistic area. In Keith Brown (ed.), *Encyclopedia of language and linguistics*, 2nd edn., 657–672. Oxford: Elsevier.

Gilquin, Gaëtanelle. 2010. *Corpus, cognition and causative constructions*. Amsterdam: John Benjamins.

Givón, Talmy. 1980. The binding hierarchy and the typology of complements. *Studies in Language* 4(3). 333–377.

Haiman, John. 1983. Iconic and economic motivation. *Language* 59(4). 781–819.

Haspelmath, Martin. 2001. The European linguistic area: Standard Average European. In Martin Haspelmath, Ekkehard König, Wolfgang Oesterreicher & Wolfgang Raible (eds.), *Language*

typology and language universals: An international handbook, 1492–1510. Berlin: Mouton de Gruyter.

Haspelmath, Martin. 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19(1). 1–33.

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

Kemmer, Suzanne & Arie Verhagen. 1994. The grammar of causatives and the conceptual structure of events. *Cognitive Linguistics* 5. 115–156.

Kulikov, Leonid I. 2001. Causatives. In Martin Haspelmath, Ekkehard König, Wolfgang Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, 886–898. Berlin: Mouton de Gruyter.

Lamiroy, Béatrice. 2011. Degrés de grammaticalisation à travers les langues de même famille. In *Mémoires de la Société de linguistique de Paris : Nouvelle Série*, Vol. 19, 167–192. Leuven: Peeters.

Leino, Jaakko. 2012. Analytical expressions for permissive causation in Finnish. In Jakko Leino & Ruprecht von Waldenfels (eds.), *Analytical causatives: From 'give' and 'come' to 'let' and 'make'*, 221–246. München: Lincom Europa.

Levshina, Natalia. In press a. Film subtitles as a corpus: An *n*-gram approach. To appear in *Corpora*.

Levshina, Natalia. In press b. Verbs of letting in Germanic and Romance languages: A quantitative investigation based on a parallel corpus of film subtitles. To appear in *Languages in Contrast*.

Lowrey, Brian. 2013. The old English causative *hatan* and its demise. *Token* 2. 23–43. http://www.ujk.edu.pl/token/issues/volume-2/ (accessed 27 March 2015).

Nedjalkov, Vladimir. 1976. *Kausativkonstruktionen*. Tübingen: TBL.

Newman, John. 1996. *Give: A cognitive linguistic study*. Berlin: Mouton de Gruyter.

Péteri, Attila. 2012. The Hungarian imperative particle *hadd*: A contrastive look at the development of analogous markers of sentence mood in some European languages. *Acta Linguistica Hungarica* 59. 439–63.

Rawoens, Gudrun. 2012. Analytical causative constructions in Swedish: An analysis of syntactic and semantic patterns. In Jakko Leino & Ruprecht von Waldenfels (eds.), *Analytical causatives: From 'give' and 'come' to 'let' and 'make'*, 105–124. München: Lincom Europa.

Rawoens, Gudrun. 2013. The Swedish verb *låta* 'let' from a synchronic and diachronic perspective. *Language Sciences* 36. 113–123.

Soares da Silva, Augusto. 2012. Stages of grammaticalization of causative verbs and constructions in Portuguese, Spanish, French and Italian. *Folia Linguistica* 46(2). 513–552.

Soares da Silva, Augusto. 2007. Verbs of letting: Some cognitive and historical aspects. In Nicole Delbecque & Bert Cornillie (eds.), *On interpreting construction schemas: From action and motion to transitivity and causality*, 171–200. Berlin: Mouton de Gruyter.

Talmy, Leonard. 2000. *Toward a cognitive semantics*. Cambridge, MA: MIT Press.

Tamm, Anne. 2012. Let me introduce the Estonian analytical causatives: The permissive and factitive *laskma* 'let, make, have, allow, permit'. In Jakko Leino & Ruprecht von Waldenfels (eds.), *Analytical causatives: From 'give' and 'come' to 'let' and 'make'*, 247–289. München: Lincom Europa.

Tiedemann, Jörg. 2012. Parallel data, tools and interfaces in OPUS. In *Proceedings of the eighth international conference on language resources and evaluation (LREC-2012 Istanbul, 21–27 May)*, 2214–2218.

Uiboaed, Kristel, Cornelius Hasselblatt, Liina Lindström, Kadri Muischnek & John Nerbonne. 2013. Variation of verbal constructions in Estonian dialects. *Literary & Linguistic Computing* 28(1). 42–62.

van der Auwera, Johan. 1998. Conclusion. In Johan van der Auwera (ed.), *Adverbial constructions in the languages of Europe*, 813–36. Berlin: Mouton de Gruyter.

van der Auwera, Johan & Kalyanamalini Sahoo. In press. On comparative concepts and descriptive categories, *such* as they are. To appear in *Acta Linguistica Hafniensia*.

Van der Horst, Joop M. 1998. *Doen* in Old and Early Middle Dutch: A comparative approach. In Ingrid Tieken-Boon de Ostade, Marijke van der Wal & Arjan van Leuvenstijn (eds.), *'Do' in English, Dutch and German: History and present-day variation*, 53–64. Münster: Nodus Publikationen.

van Haeringen, Coenraad B. 1956. *Nederlands tussen Duits en Engels* [Dutch between German and English]. Den Haag: Servire.

Verhagen, Arie & Suzanne Kemmer. 1997. Interaction and causation: Causative constructions in modern standard Dutch. *Journal of Pragmatics* 24. 61–82.

von Waldenfels, Ruprecht. 2012a. Finnish *antaa* and Russian *davat'* 'to give' as causatives: A contrastive analysis. In Jakko Leino & Ruprecht von Waldenfels (eds.), *Analytical causatives: From 'give' and 'come' to 'let' and 'make'*, 187–220. München: Lincom Europa.

von Waldenfels, Ruprecht. 2012b. *The grammaticalization of 'give' + infinitive: A comparative study of Russian, Polish and Czech*. Berlin: De Gruyter Mouton.

von Waldenfels, Ruprecht. 2015. Grammaticalized 'give' in Slavic between drift and contact: Causative, modal, imperative and existential constructions. In Brian Nolan, Gudrun Rawoens & Elke Diedrichsen (eds.), *Causation, permission, and transfer: Argument realisation in GET, TAKE, PUT, GIVE and LET verbs*, 107–127. Amsterdam: John Benjamins.

Wälchli, Bernhard & Michael Cysouw. 2012. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50(3). 671–710.

Wiemer, Björn. 2014. *Quo vadis* grammaticalization theory? Why complex language change is like words. *Folia Linguistica* 48(2). 425–467.