“Let in translation”: A typological study of the concept of LETTING in a parallel corpus of film subtitles

Introduction

This paper is an onomasiological study of the differences and similarities in the conceptualization of LETTING in twelve European languages. Bringing together Cognitive Semantics, typology and quantitative corpus-based methods, the paper presents a new way of creating bottom-up second-generation semantic maps, which is is based on a hypothesis that conceptually similar situations are coded in a similar way across different languages. Unlike most semantic maps in typology, the approach is exemplar-, or token-based. The aims of the study are twofold. First, using a sample of exemplars of LETTING from a parallel multilingual corpus of film subtitles, I examine how the languages cut the common conceptual space represented by the map. The approach is thus a probabilistic constructionist interpretation of well-known structuralist lexical fields (e.g. Hjelmslev 1959[1957]; Lehrer 1974). Second, following the aggregation methods applied in contemporary typology and socio- and dialectometry, the paper demonstrates how the same matrix of exemplars can be used to establish the distances between the languages with regard to the conceptualization of LETTING events.

Theoretical background and novelty

The concept of LETTING has received significant attention in Cognitive Semantics, mainly in Talmy's force-dynamic framework (Talmy 2000). Letting is considered to be one of the key force-dynamic patterns. As a kind of negative causation, it involves a stronger entity, which does not override the lettee's intrinsic tendency towards rest or motion. There exist a few studies of cross-linguistic and cross-cultural differences in the conceptualization of causation and causality, for instance, Bally and Wierzbicka's typology of 'analytical' and 'phenomenological' languages (Bally 1920; Wierzbicka 1988), which focus, respectively, on the cause-effect relations, or on the properties of events as perceived by a subject. This continuum of analyticity looks as follows:

English > French > German > Russian

Another important work is by Wierzbicka (2002), who pinpoints several unique features of English *let* in comparison with the similar constructions in German and Russian (see also Soares da Silva 2007 for a few peculiarities of the cognates of *let* in European languages). Yet, to the best of my knowledge, all these typological and semantic hypotheses have not been tested quantitatively on a broad selection of languages. In addition, no onomasiological (concept-based) study of this fundamental force-dynamic category has ever been performed.

Data

Multilingual parallel corpora are gaining popularity in typological research (van der Auwera et al. 2005; Cysouw and Wälchli 2007). The study is based on the synchronized subtitles of a few blockbusters (e.g. Avatar, Star Wars and Twilight) in twelve languages from the Germanic, Romance and Slavic language families. The subtitles were taken from www.subscene.com and www.opensubtitles.org. The first step was the selection of the items that represent the constructional field of LETTING with the help of the WordNet, FrameNet and other lexicographic sources. The near-synonyms included *let*, *allow*, *release*, *grant permission, may* (permissive) and other words and constructions. Next, the lexical items were searched for in the English (original) subtitles, and the equivalent contexts in the other 11 languages were identified with the help of the synchronizing information. After a manual check, a sample of 123 multilingual exemplars was selected. The resulting data matrix looked as shown in Table 1.

EN	DE	NL	SV	ES	FR	IT	PT	BG	PL	RU	SL
let	*lassen*	*laten*	*låta*	*dejar*	*laisser*	*lasciare*	*deixar*	*puskam*	*pozwalać*	*pozwolit'*	*dovoliti*
allow	*zulassen*	*laten*	*tillåta*	*permitar*	*permettre*	*permettere*	*deixar*	*pozvoljavam*	*dopuścić*	*dopustit'*	*dovoliti*
...	...	...	...	...	...	...	...	...	...	...	...

Table 1. A segment of the data frame with equivalents of LETTING-exemplars (rows) in 12 languages (columns).

Quantitative analyses

Unlike most semantic map models, which cluster more or less abstract semantic functions as types (e.g. Haspelmath 1997; van der Auwera & Plungian 1998; Cysouw 2010), this paper employs an exemplar-, or token-based approach. The semantic functions emerge in a bottom-up way from clusters of exemplars as contiguous areas on a map.

To investigate how the languages cut the conceptual space, I created a matrix of distances between the exemplars. The distances corresponded to the (dis)similarities between the exemplars. The (dis)similarities were established by comparing the constructions that were used to convey LET in the twelve languages, as suggested in Cysouw (2010). An adjusted version of Gower's (1971) similarity metric for categorical and mixed data was used, which also took into account a partial overlap between the constructions (e.g. Dutch *laten* 'let, make' and *loslaten* 'release, let go'). Missing values were ignored. The algorithm was implemented in a Python script. The resulting distance matrix was next represented in a low-dimensional space with the help of Multidimensional Scaling (cf. Levinson and Meira 2003; Croft and Poole 2008). I used the SMACOF majorization algorithm by de Leeuw and Mair (2009) inplemented in the `smacof` package in R. The dimensionality and overall quality of the solution were tested with the help of scree plots, Shepard plots and other diagnostic methods.

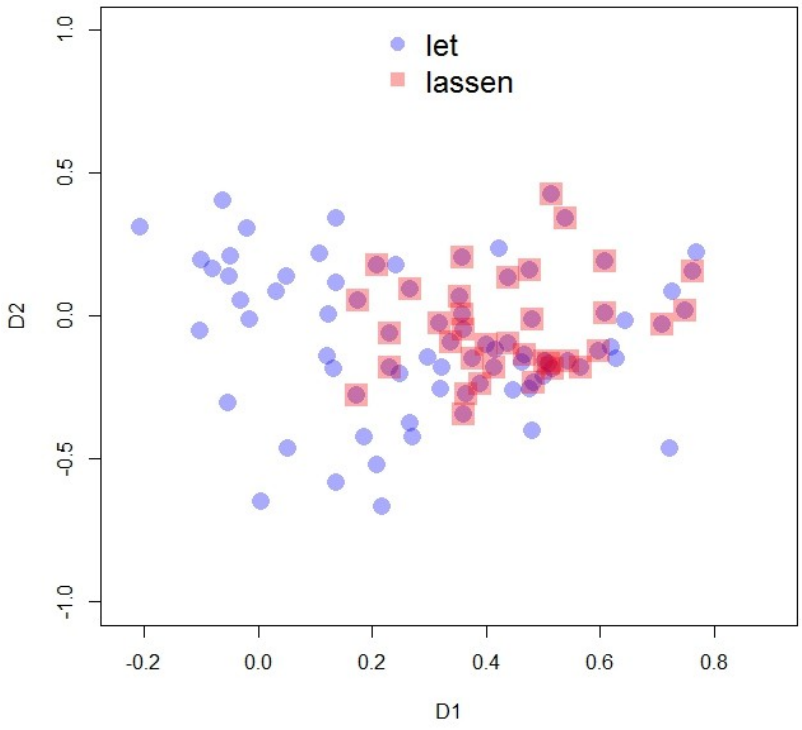

Figure 1. The semantic overlap between English *let* and German *lassen*.

Next, I plotted the language-specific lexemes on the MDS map and compared the ways in which the languages cut the common conceptual space. Figure 1 shows a segment of the semantic space, with English *let* and German *lassen*. The data points are exemplars (rows in Table 1). The symbols show if the exemplars are encoded by *let* or *lassen* in the English and German versions of the subtitles, correspondingly. The central part, where the constructions overlap, contains exemplars that are related to non-impingement scenes, exemplified by such senses as non-interruption (*Let the girl eat her garden burger*) and non-prevention (*Why didn't you just let the van crush me?*), in Wierzbicka's (2002) terms. The senses that are typical of *let*, but uncommon for *lassen* are permission found at the bottom left part of the cloud, and *let* of cooperative dialogue and interaction (*Let me tell you a litle bit about Taylor Durden*) at the top left side. The former is covered in German by *erlauben* and some other constructions, whereas the other is usually expressed paraphrastically.

The same data were used for a series of aggregate analyses of the languages. Several approaches, which treat the equivalent constructions at different levels of granularity, were tested. For instance, at the most coarse-grained level, the languages were compared with regard to the presence or absence of a LETTING expression in the context. Again, Gower's distance and MDS were used to model the (dis)similarities. The resulting three-dimensional solution is shown in Figure 2. It displays genetically conditioned clusters of the languages, although some languages (e.g. French) are very close to the languages from another group, which suggests the presence of areal effects.

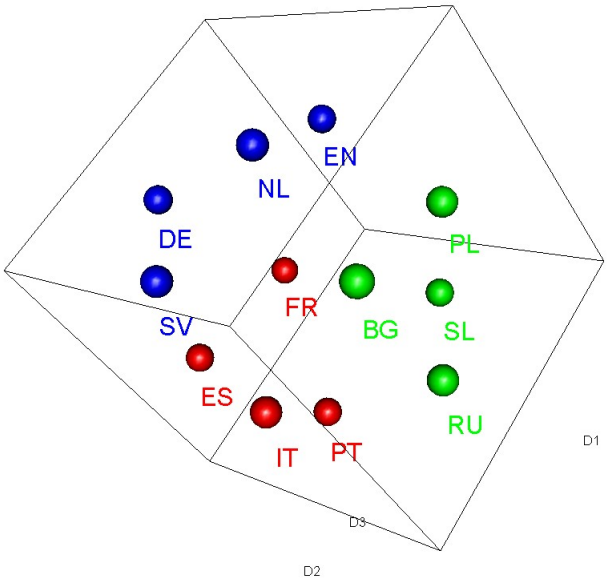

Figure 2. Similarity between the twelve languages visualized in a rotated three-dimensional MDS plot. The distances are based on the presence or absence of a marker of LETTING in each language.

Conclusions

How do the results of the study relate to the two research questions mentioned in the Introduction? First, the semantic maps show that the languages cut the common conceptual space in different ways. English *let* seems to have the broadest semantics in comparison with its equivalents in the other languages. At the opposite end of the spectrum, the Slavic languages, especially Polish and Russian, seem to have the most fragmented space. This suggests that the cause-effect relatonships in those languages are blended with, and sometimes substituted by, other conceptual information. This may be interpreted as evidence in favour of Bally-Wierzbicka's hypothesis about the 'analytical' and 'phenomenological' languages. An important methodological conclusion, which supports Cysouw's (2010) working hypothesis, is that the conceptual space generated with the help of multilingual translations is interpretable from a conceptual point of view. Namely, its areas strongly resemble the

semantic types suggested in previous research.

On the other hand, the results of the agglomerate analyses suggest that the cross-linguistic conceptualization differences can be explained by both genetic and areal factors. With a few exceptions, the Romance, Germanic and Slavic languages are normally clustered together. However, the results seem to depend on the types and granularity of the tertia comparationis (constructions). From the epistemological perspective, this demonstrates that the results of typological clustering models, as well as lectometric analyses, should not be overgeneralized, since the methods and approaches may give prominence to different facets of multidimensional linguistic phenomena.

References

Bally, Ch. 1920. Impressionisme et grammaire. In: *Mélanges d'histoire littéraire et de philology offerts à M. Bernard Bouvier*, 261-279. Geneva: Sonor.

Croft, W. & K. T. Poole. 2008. Inferring universals from grammatical variation: multidimensional scaling for typological analysis. *Theoretical Linguistics* 34, 1-37.

Cysouw, M. 2010. Semantic maps as metrics on meaning. *Linguistic Discovery* 8(1), 70-95.

Cysouw, M. & B. Wälchli. 2007. Parallel texts: Using translational equivalents in linguistic typology. Introduction to the special issue of *Sprachtypologie und Universalienforschung* (STUF) 60(2), 95-99.

Gower, J. C. 1971. A general coefficient of similarity and some of its properties, *Biometrics* 27: 857-874.

Haspelmath, M. 1997. *Indefinite Pronouns*. Oxford: Oxford University Press.

Hjelmslev, L. 1959 [1957]. Pour une sémantique structurale. In L. Hjelmslev (ed.), *Essais de linguistique générale*, 96-112. Copenhagen: Nordisk Sprog-og Kulturforlag.

de Leeuw, J. & P. Mair. 2009. Multidimensional scaling using majorization: The R package smacof. *Journal of Statistical Software* 31(3), 1-3.

Lehrer, Adrienne. 1974. *Semantic Fields and Lexical Structure.* Amsterdam: North Holland.

Levinson, S. C., & S. Meira. 2003. 'Natural concepts' in the spatial topological domain - adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language* 79(3), 485-516.

Soares da Silva, A. 2007. Verbs of letting: Some cognitive and historical aspects. In N. Delbeque & B. Cornillie (eds.), *Cognitive approaches to lexical semantics*, 171-200. Berlin/New York: Mouton de Gryter.

Talmy, L. 2000. *Toward a Cognitive Semantics*. Vol. 1. Cambridge, MA: MIT Press.

van der Auwera, J., E. Schalley & J. Nuyts. 2005. Epistemic possibility in a Slavonic parallel corpus: A pilot study In: P. Karlik & B. Hansen (eds.), *Modalität in slavischen Sprachen. Neue Perspektiven*, 201-217. München: Sagner.

van der Auwera, J. & V. Plungian. 1998. Modality's Semantic Map. *Linguistic Typology* 2: 79-124.

Wierzbicka, A. 1988. *The semantics of grammar.* Amsterdam/Philadelphia: John Benjamins.

Wierzbicka, A. 2002. English causative constructions in an ethnosyntactic perspective: Focusing on LET. In: Nick Enfield (ed.), *Ethnosyntax*. Oxford: Oxford University Press. 162-203.