SEMANTIC MAP

To appear in New Encyclopaedia of Czech Online (Ed. Petr Karlik)

A semantic map is a graphical representation of functions of linguistic constructions, unto which linguistic forms from one or different languages can be mapped. There exist two main types of semantic maps. The first type is a network that consists of nodes and links between them. The nodes are different functions, which serve as *tertia comparationis* for cross-linguistic comparison of words, constructions or grammatical categories. Since these maps were historically first, they are also called 'classical' (van der Auwera 2013) or 'first-generation maps' (Sansò 2009). Yet another name is 'connectivity maps' (van der Auwera 2013). The second type of maps displays data points, i.e. examples from a corpus or experimental stimuli. The relationships between the data points are represented by distances (or, inversely, proximities). These are probabilistic, or statistical semantic maps, which are also called 'proximity maps' (van der Auwera 2013) or 'second-generation maps' (Sansò 2009).

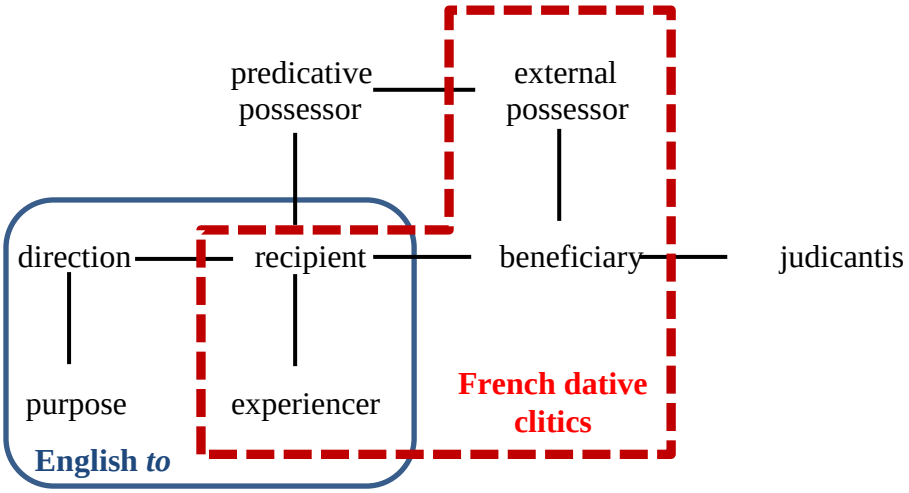


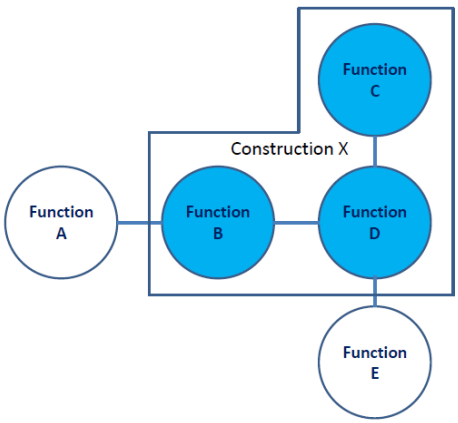
Figure 1. Semantic map of dative functions from Haspelmath (2003) with the areas covered by English *to* and French dative clitics.

In classical semantic maps, the nodes are linked if the corresponding functions co-occur at least in one linguistic construction. Consider an example in Figure 1. This is a map of dative functions from Haspelmath (2003), with the English preposition *to* and the French

dative clitics (*me*, *te*, *lui*, *nous*, *vous*, *leur*) mapped onto the graph. One can see that the areas covered by the English preposition *to* and the French dative clitics partly overlap.

The procedure of creation of classical semantic maps is as follows. The first step is to identify the functions, which will act as nodes of the network. According to Haspelmath (2003), a function should be put on a map if there are at least two languages that differ with regard to this function. The semantic map approach is agnostic as far as the polysemy and monosemy distinction (or, in other terms, ambiguity and vagueness) is concerned. In principle, any function can be an element of a semantic map. Next, a spatial configuration is sought such that satisfies the Semantic Map Connectivity Hypothesis, which was formulated by Croft (2001: 96). If two given functions are expressed by one form in at least one language, the corresponding nodes should be connected. This principle is also known as the Contiguity, or Adjacency Principle. Consider the imaginary examples in Figure 2. The figure on the left represents a semantic map with functions A, B, C, D and E. Construction X covers a connected region on that map (functions B, C and D). The Connectivity Principle is observed. In contrast, Construction Y in the figure on the right has functions A, B and C. C is not connected directly with A or B. Therefore, the Connectivity Principle is violated.

The Connectivity Principle is observed



The Connectivity Principle is violated

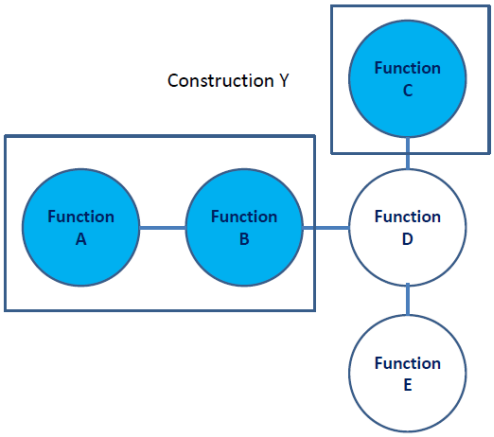


Figure 2. Imaginary semantic map with the Connectivity Principle observed (left) and the Connectivity Principle violated (right).

In order to have a scientific value, the semantic map should be parsimonious. Linking every node with all other nodes makes the semantic map vacuous (Haspelmath 2003). According to Haspelmath (2003), a dozen genealogically diverse languages are enough to obtain a stable solution that does not have to be modified when a new language is added. Usually, the spatial configuration is sought manually, by trial and error. Recently, there have been attempts to employ Graph Theory in order to create these configurations automatically. Although the semantic map inference is computationally intractable, there exist heuristics that can approximate it (Regier *et al.* 2013).

In addition to allowing for a straightforward language comparison, semantic maps can be used for formulation of implicational hierarchies (e.g. Haspelmath 1997). For example, if a dative marker has the functions of an experiencer and a beneficiary, it should also have the function of a recipient, according to the semantic map of the dative functions that was introduced above. One can also use semantic maps to generate diachronic hypotheses about the semantic development of words and constructions: if a construction that had a function A now has a function Q, it should have had at a certain stage the nodes that are located between A and Q. It is also possible to create semantic maps with directional links, which show which function was diachronically first and which emerged later as a result of semantic extension (Narrog 2010). For example, a directional semantic map of the GOAL-RECIPIENT domain, based on the diachronic information provided in Narrog (2010), may look as shown in Figure 3. The map was constructed automatically with the help of the package *igraph* in R. The nodes are case functions. Such maps represent a particular interest for studies of grammaticalization.

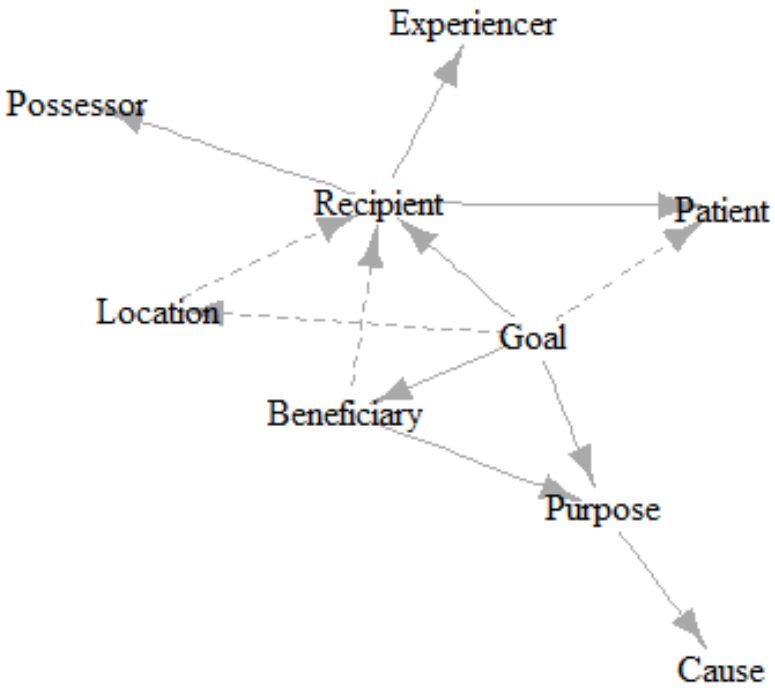


Figure 3. A directional semantic map of the GOAL-RECIPIENT domain based on Narrog (2010).

The majority of classical semantic maps are usually non-hierarchical: they do not display hyponymous and hyperonymous functions. However, there are a few hierarchical semantic maps, e.g. the semantic map of modality by van der Auwera & Plungian (1998) and the semantic map of the concept EMPTINESS in Koptjevskaja-Tamm *et al.* (Forthc.). Consider a simple example from the domain of causation. All causative situations can be subdivided into two types, according to Nedjalkov (1976): factitive and permissive. Factitive causation, in its turn, can be subdivided into direct and indirect. The colours on the map indicate the distribution of Dutch causative constructions with *doen* "do" (pink) and *laten* "let" (light-green) with regard to these semantic functions.

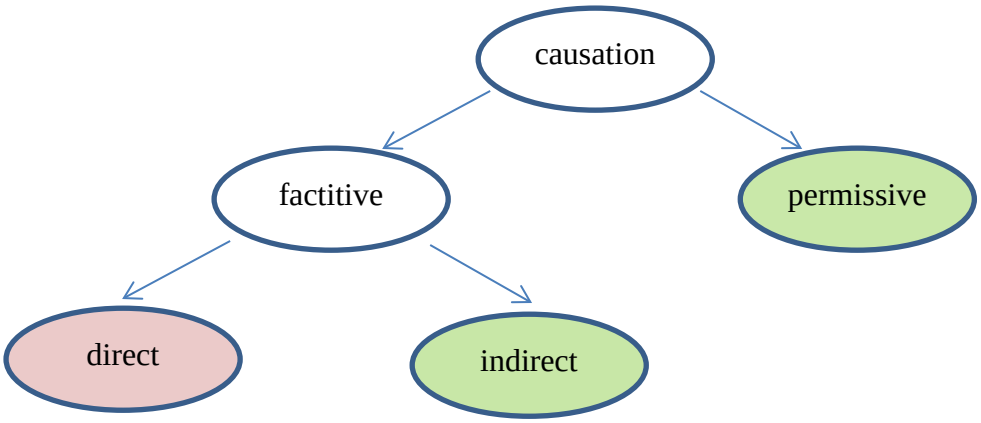


Figure 4. A hierarchical semantic map of causation, with the functions performed by Dutch causative *doen* (pink) and *laten* (light-green).

The second type of maps represents points that are located at different distances from one another. The points are usually exemplars from parallel corpora or non-verbal experimental stimuli. They represent a specific situation, rather than a broad semantic function. The greater the distance, the more frequently the situations that correspond to the exemplars or stimuli are expressed by the same forms in different languages. Such maps are constructed with the help of multivariate statistical methods, such as Multidimensional Scaling (e.g. Wälchli & Cysouw 2012) or Correspondence Analysis (Majid *et al.* 2012). See an example in Figure 5, which displays a Multidimensional Scaling map with exemplars of analytic (periphrastic) causative constructions from a parallel corpus of eighteen European languages from Levshina (Under review). The dots are different semantic situations of causation. The lines show the areas with a high density of exemplars of analytic causatives in five Romance languages: French *faire* + V_{INF}, Italian *fare* + V_{INF}, Spanish *hacer* + V_{INF}, Portuguese *fazer* + V_{INF} / V_{SUBJ} and Romanian *a face* + *să* V_{SUBJ}. The interpretation of the dimensions and clusters is the researcher's responsibility. In this example, the horizontal dimension corresponds to the distinction between permissive (left) and factitive causation (right), whereas the vertical dimension in the region on the right separates forceful causation (top) from non-forceful causation (bottom).

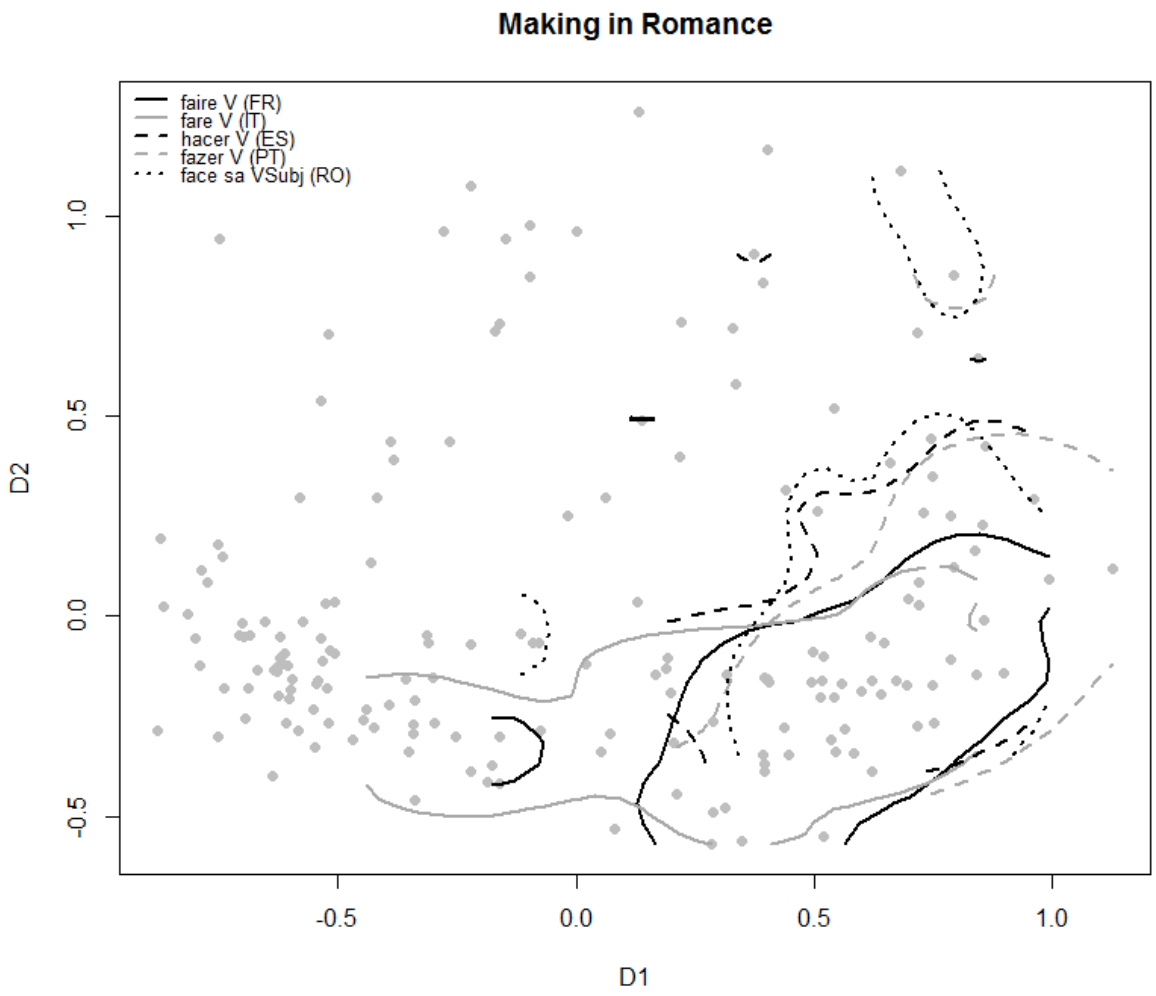


Figure 5. A probabilistic semantic map of analytic causatives based on 18 European languages. The lines show the areas of high probability of exemplars of five Romance causative constructions. This map was created with the help of the R package *smacof*.

In addition to these two main types, there also exist hybrid approaches. For example, the CLICS database (LIST *et al.* 2014), which is available online at http://clics.lingpy.org, represents colexifications of semantic concepts in a large number of languages all over the world. One speaks of colexification when two or more meanings are covered by one lexical item in a language (FRANÇOIS 2008). The screenshot in Figure 6 shows the information one can obtain for the concept HAND. The graph displayed on the right looks like a classical semantic map, with links and nodes that represent semantic functions. At the same time, the line thickness corresponds to the co-occurrence frequency of a pair of semantic functions (i.e. colexification) in the data base. One can see that the concept HAND is the most commonly colexified with the concept ARM. Importantly, the graph, which is created automatically,

connects all pairs of nodes with at least one colexification (thus, the parsimony principle is not observed).

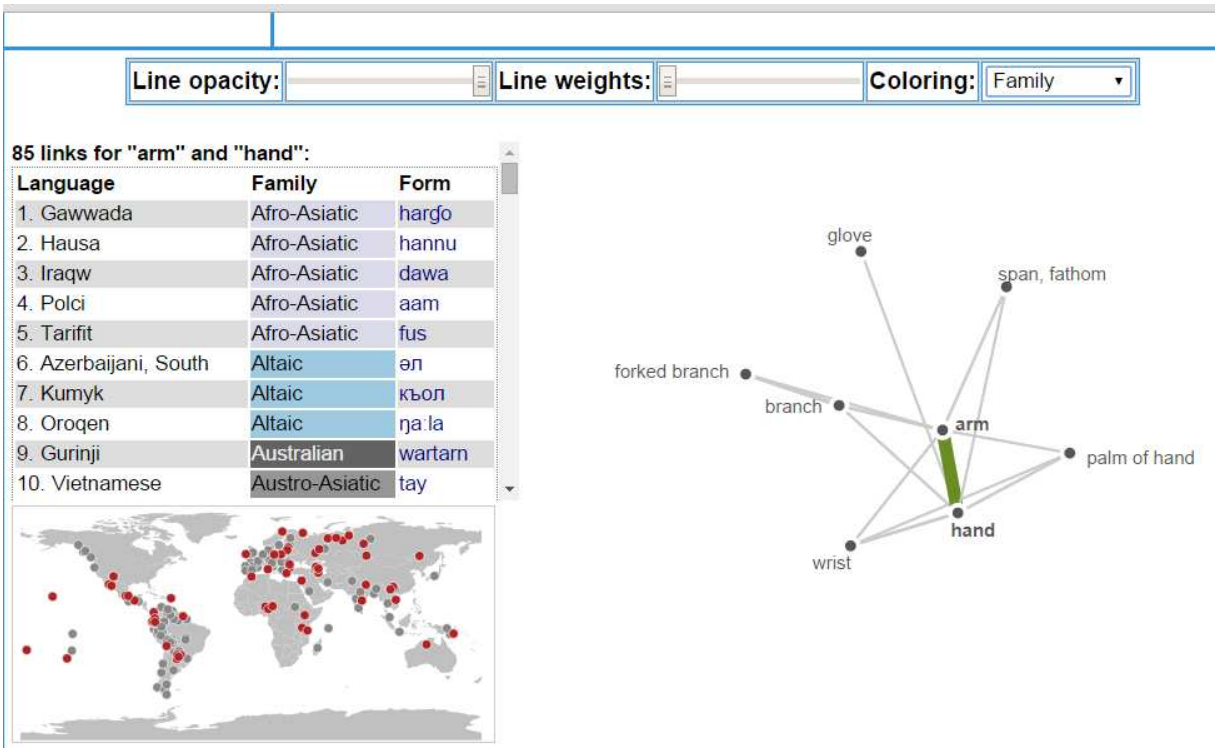


Figure 6. A screenshot from the CLICS database at http://clics.lingpy.org.

Although semantic maps provide a convenient tool for visualization and exploration of cross-linguistic variation in form-meaning mapping, their theoretical interpretation is debatable. While some linguists, e.g. Croft (2010), claim that semantic maps represent the universal conceptual space that is part of the speakers' mental representation, Cristofaro (2010) argues that polyfunctionality of a linguistic construction does not always arise from conceptual similarity between its functions and that semantic maps represent diachronic development, rather than individual knowledge at the synchronic level.

References

CRISTOFARO, S. Semantic maps and mental representation. *Linguistic Discovery* 8(1), 2010, 35-52; CROFT, W. *Radical Construction Grammar: Syntactic theory in typological perspective*, 2001; CROFT, W. What do semantic maps tell us? *Linguistic Discovery* 8(1),

2010, 53-60; FRANÇOIS, A. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In Vanhove, M. (ed.), *From Polysemy to Semantic Change*, 2008, 163-215; HASPELMATH, M. *Indefinite pronouns*, 1997; HASPELMATH, M. The geometry of grammatical meaning: semantic maps and cross-linguistic comparison. In Tomasello, M. (ed.), *The new psychology of language. Cognitive and functional approaches to language structure. Vol. 2*, 2003, 211-242; KOPTJEVSKAJA-TAMM, M., E. RAKHILINA & M. VANHOVE. The semantics of lexical typology. In Rimer, N. (ed.), *The Routledge Handbook of Semantics*, forthcoming; LIST, J.-M., TH. MAYER, A. TERHALLE & M. URBAN. CLICS: Database of Cross-Linguistic Colexifications, 2014. LEVSHINA, N. European analytic causatives as a comparative concept: Evidence from a parallel corpus of film subtitles, under review. MAJID, A., J. S. BOSTER & M. BOWERMAN. The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition* 109, 2008, 235-250; Narrog, H. A diachronic dimension in maps of case functions. *Linguistic Discovery* 8(1), 2010, 233-254; NEDJALKOV, V. *Kausativkonstruktionen*, 1976; REGIER, T. N. KHETARPAL & A. MAJID. Inferring semantic maps. *Linguistic Typology* 17(1), 2013, 89-105; SANSÒ, A. How conceptual are semantic maps? *Linguistic Discovery* 8(1), 2010, 288-309; VAN DER AUWERA, J. Semantic maps, for synchronic and diachronic typology. In Ramat, A. G., C. Mauri & P. Molinelli (eds.), *Synchrony and Diachrony: A Dynamic Interface*, 2013, 153-176; VAN DER AUWERA, J. & V. PLUNGIAN. Modality's semantic map. *Linguistic Typology* 2, 1998, 79-124; WÄLCHLI, B. & M. CYSOUW. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50(3), 2012, 671-710.

Natalia Levshina