DOI: 10.1111/lnc3.12286

TEACHING AND LEARNING GUIDE

WILEY

Teaching & Learning Guide for: The semantic map model

Thanasis Georgakopoulos | Stéphane Polis

University of Liège & F.R.S.-FNRS, Belgium

This guide accompanies the following article:

Georgakopoulos, T., & Polis, S. (2018). "The semantic map model: State of the art and future avenues for linguistic research". Language and Linguistics Compass *12/2 (2018) pp*. 1-33 *[https://doi.org/10.1111/lnc3.12270]*.

Author's Introduction

The semantic map model is relatively new in linguistic research, but it has been used intensively during the past three decades for studying a variety of cross-linguistic and language-specific questions. This method enables the capturing of regular patterns of semantic structure based on similarities of form-meaning correspondence across languages. It has been fruitfully applied to the study of a variety of topics in linguistic typology, semantics, and historical linguistics and plays a prominent role in modern linguistic theory. This teaching and learning guide aims to provide readers with (a) key relevant works in the field, (b) additional information about online electronic resources and software solutions for graph visualization, and (c) material that can be used in order to introduce the semantic map model and its applications in the context of courses on linguistic typology, historical linguistics, and computational linguistics. Four modules are suggested for each course (two shared introductory modules and two specific modules). Alternatively, the eight modules can be clustered in the framework of a single seminar on semantic maps.

Author Recommends:

van der Auwera J., & Plungian, V. A. (1998). Modality's semantic map. Linguistic Typology, *2*(1), 79–124. https://doi.org/10.1515/lity.1998.2.1.79.

In this study, van der Auwera and Plungian discuss earlier accounts of modality, unify them, and propose a diachronic semantic map of this semantic domain. This map integrates the premodal, modal, and post-modal meanings and is claimed to have universal relevance. In the paper, different conventions are used to visualize the types of relationships between meanings (such as hyperonymy, hyponymy, metonymy, and metaphor). This is the first study using the semantic map model that visually combines three different kinds of linguistic information, namely, synchrony, diachrony and types of semantic relationships. It includes many methodological insights that are crucial for later studies in the field.

Haspelmath, M. (2003). The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In M. Tomasello (Ed.), *The new psychology of language, vol.* 2 (pp. 211–243). New York: Erlbaum.

In this seminal article, Haspelmath gives a detailed description of graph-based semantic maps (also known as "classical" semantic maps, implicational or connectivity maps). He discusses and illustrates the advantages of this approach over alternate solutions, for example, the *list method* and the *general meaning method*. He further describes the necessary steps for building a semantic map and, by illustrating concrete cases (e.g., from the domain of indefiniteness and the instrumental area), he underlines both the predictive power of the model and its limitations. In the final part of the paper, Haspelmath gives clear directions for future research, for example, regarding the extension of the model from grammatical to lexical meanings and considers the potential integration of a diachronic dimension. This article is a useful resource for reading courses, as it not only introduces the model in a clear step-by-step fashion but also situates it within the wider theoretical debate.

François, A. (2008). Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In M. Vanhove (Ed.), *From polysemy to semantic change. Towards a typology of lexical semantic associations* (pp. 163–215). Amsterdam/Philadelphia: John Benjamins.

This paper provides a blueprint for constructing graph-based lexical semantic maps. Drawing on the classical semantic map method outlined in Haspelmath (2003), François discusses the possibility of comparing lexicons across languages. He coins the term "colexification" in order to refer to lexical items that "lexify" multiple senses and exemplifies his approach with a map for the rich polysemy patterns associated with the notion BREATHE. This map is drawn on an empirical basis and structures the meanings associated with BREATHE in the world's languages. His colexification model aims at providing a better understanding of lexical organization in synchrony, by keeping a balance between the search for language universals and the need for respecting each language's uniqueness. It is an excellent resource for anyone interested in lexical typology.

Croft, W., & Poole, K. T. (2008). Inferring universals from grammatical variation: Multidimensional scaling for typological analysis. Theoretical Linguistics, *34*(1), 1–37. https://doi.org/10.1515/thli.2008.001.

This position paper by Croft and Poole is published in a special issue of the journal *Theoretical Linguistics*. Based on the observation that classical semantic maps are not mathematically well defined, the authors present an alternative visual approach, describing how multidimensional scaling (MDS) can be used in order to uncover typological universals. They argue that MDS is more effective with large-scale and complex datasets than the traditional model. Crucially, the result of the analysis is not a graph structure (as with classical semantic maps): The meanings are represented by points, distributed across a two-dimensional space using multivariate statistical techniques, and the distance between the points is indicative of their (dis)similarity. The special issue of the journal hosted four responses to and comments on the paper, which were followed by a final response from Croft and Poole.

Link: https://github.com/jaytimm/MDS_for_Linguists

The link makes available the resources for linguistic typologists interested in applying NOMINATE MDS techniques to linguistic data as presented in Croft and Poole.

Cysouw, M., Haspelmath, M., & Malchukov, A. L. (2010). Special issue "semantic maps: Methods and applications". Linguistic Discovery, *8*(1).

This special issue in the journal *Linguistic Discovery* is the most systematic effort to discuss the different aspects of the semantic map method. This issue contains 14 original research articles and can be seen as a companion volume. All of the manuscripts were opened up for commentary from other researchers. The total number of articles including the commentaries and the authors' replies is 42. The contributions focus on a wide range of topics, including

- advantages and limitations of semantic maps;
- classical semantic maps versus proximity maps using multidimensional scaling techniques;
- semantic maps and diachrony;
- how to build semantic maps;
- challenges for the method (e.g., contiguity violations; large datasets);
- the theoretical status and scope of semantic maps within linguistic theory (e.g., how conceptual are semantic maps?); and
- extending the scope of semantic maps (e.g., by including pragmatic dimensions or by investigating word formation patterns).

Links:
Journal: http://journals.dartmouth.edu/cgi-bin/WebObjects/Journals.woa/xmlpage/1/issue/34
Conference: https://www.eva.mpg.de/lingua/conference/07-SemanticMaps/index.html

Regier, T., Khetarpal, N., & Majid, A. (2013). Inferring semantic maps. Linguistic Typology, *17*(1), 89–105. https://doi.org/10.1515/lity-2013-0003

In this paper, Regier, Khetarpal, and Majid show that the semantic map inference problem is formally identical to another problem: inferring a social network from disease outbreaks in a population. They acknowledge the fact that semantic map inference is computationally intractable (as suggested by Croft & Poole 2008), but show that an efficient approximation algorithm exists (Angluin, Aspnes, & Reyzin, 2010). Having tested the algorithm on the cross-linguistic data of Haspelmath (1997) about the indefinite pronouns and of Levinson et al. (2003) on spatial categories, they conclude that the approximations produced by the algorithm are of high quality, which means that they produce equal or better results than the manually plotted maps.

Link: http://lclab.berkeley.edu/regier/semantic-maps/

The link contains the python code for the network inference algorithm as well as a sample input file, which is useful for scholars who want to test the method.

Additional references:

Angluin, D., Aspnes, J., & Reyzin, L. (2010). Inferring social networks from outbreaks. In M. Hutter, F. Stephan, V. Vovk & T. Zeugmann (Eds.), *Algorithmic learning theory 2010* (pp. 104–118). Berlin: Springer. https://doi.org/10.1007/978-3-642-16108-7_12

Haspelmath, M. (1997). *Indefinite pronouns*. Oxford: Oxford University Press.

Levinson, S., & Meira, S., & the Language and Cognition Group (2003). 'Natural concepts' in the spatial topological domain—Adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. Language *79* (3), 485–516. https://doi.org/10.1353/lan.2003.0174.

van der Auwera, J. (2013). Semantic maps, for synchronic and diachronic typology. In A. G. Ramat, C. Mauri & P. Molinelli (Eds.), *Synchrony and diachrony. A dynamic interface* (pp. 153–176). https://doi.org/10.1075/slcs.133.07auw.

In this paper, van der Auwera defends the classical semantic map method and outlines the main advantages of this approach for studying both synchronic variation and semantic change. He reviews the pros and cons of the graph model as compared to the statistical approach that produces scatterplots using MDS (see Croft & Poole 2008 above). Addressing the main criticisms against classical semantic maps, he claims that maps produced with the statistical approach lack two essential properties: the adjacency requirement (which he sees as a defining feature of semantic maps) as well as the predictive power (in particular the implicational universals that cannot be inferred from scatterplots).

Online Materials

1. Linguistic resources

(a) Database of crosslinguistic colexifications (CLICS): http://clics.lingpy.org/

See List, J.-M., Mayer, T., Terhalle, A., & Urban, M. (2014). CLICS: Database of cross-linguistic colexifications. Marburg: Forschungszentrum Deutscher Sprachatlas (Version 1.0, retrieved from http://CLICS.lingpy.org, accessed on 2017–7-6).

A new version of CLICS has recently been published online; it includes information about polysemy patterns in 1220 language varieties of the world (see List, J.-M., Greenhill, S., Anderson, C., Mayer, T., Tresoldi, T., & Forkel, R. (Eds.) (2018). CLICS: Database of Cross-Linguistic Colexifications. Max Planck Institute for the Science of Human History: Jena (retrieved from http://clics.clld.org).

This online database of synchronic lexical association provides information about colexification patterns attested for 1,280 concepts in 221 language varieties of the world. CLICS assembles data from four main sources: the Intercontinental Dictionary Series (IDS), the World Loanword Database (WOLD), the Logos dictionary, and the Språbanken project. The web interface turns the raw data into colexification networks visualized with a force-directed graph layout. Especially worth noting are the weighted edges of the networks—which make it immediately clear which patterns are frequent and which not—and the interactive components that allow for analysis of the genetic and areal distributions of these patterns (cf. Figure 1). Although a colexification network is not, strictly speaking, a semantic map, the data made available on CLICS can be used as a convenient source for building synchronic lexical semantic maps.

(b) Open Multilingual WordNet: http://compling.hss.ntu.edu.sg/omw/

See Bond, Fr. & Paik, K. (2012). A survey of wordnets and their licenses. In *Proceedings of the 6th Global WordNet Conference (GWC 2012)*, Matsue, 64–71 (available online: http://web.mysites.ntu.edu.sg/fcbond/open/pubs/2012-gwc-wn-license.pdf).

Open Multilingual WordNet gives access to wordnets in multiple languages, which are linked to the original Princeton WordNet of English (PWN). The basic units of WordNet are sets of cognitive synonyms (called synsets), each expressing a distinct concept. For instance, "pass, hand, reach, pass on, turn over, give" all belong to a synset which can be defined as "place into the hands or custody of." Synsets are interlinked through lexical relations (any word can occur in several synsets) and

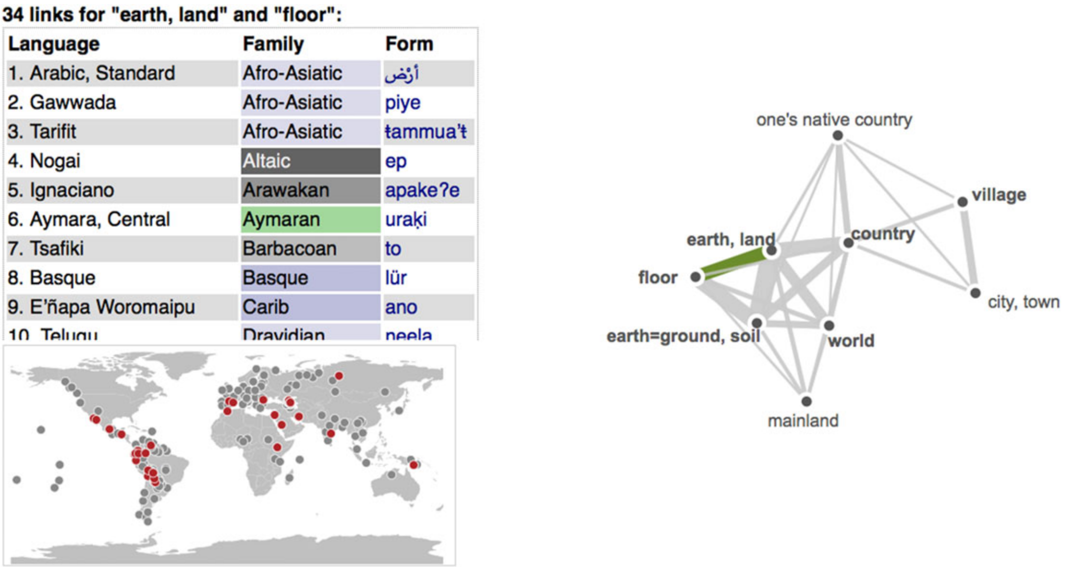

FIGURE 1 The community network for the concept "earth, land" in CLICS

conceptual-semantic relations (hyperonymy, hyponymy, meronymy, etc.). Open Multilingual WordNet can be queried using the python Natural Language Tool-Kit (NLTK) WordNet interface (http://www.nltk.org/howto/wordnet. html). An extended version, the "Extended Open Multilingual Wordnet" (http://compling.hss.ntu.edu.sg/omw/summx.html), is also available, but since the additional data were automatically collected from Wiktionary and the Unicode Common Locale Data Repository, its accuracy rate is lower (estimated at approximately 94%).

2. Visualization tools

(a) Gephi: https://gephi.org

See Bastian M., Heymann S., Jacomy M. (2009). *Gephi: An open source software for exploring and manipulating networks*. International AAAI Conference on Weblogs and Social Media.

Gephi is a free and open source program, which supports a great variety of graph import/export formats and offers many types of layouts. It comes with a *Quick Start Tutorial*, a *Visualization Tutorial*, and a *Layouts Tutorial*, which guide users through the basic and advanced settings of the program. Gephi facilitates the exploration of polysemy networks and semantic maps thanks to the wealth of statistical and filtering methods available. The "Force Atlas" algorithm often proves to be the most appropriate for linguistic analysis, and statistics about the modularity of the graph, which automatically identify coherent sub-networks, assist in making sense of complex data (see Figure 2).

(b) Cytoscape: http://www.cytoscape.org/

See Shannon P., Markiel A., Ozier O., Baliga N. S., Wang J. T., Ramage D., Amin N., Schwikowski B., Ideker T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research, *13*(11), 2498–504.

Cytoscape is another powerful open source solution for network visualization and analysis. Originally designed for biological research, it uses *network* to refer to *graph* and can use *interactions* to refer to *edges*. Less user-friendly than Gephi, Cytoscape has several features not found with other tools: the capabilities to handle multiple links between a pair of nodes, a wider range

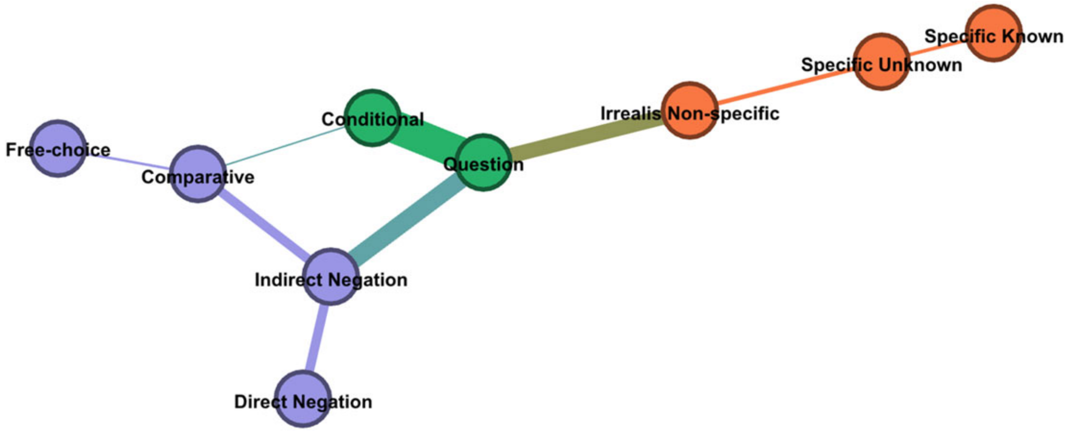

FIGURE 2 Visualization of Haspelmath's (1997) data on indefinite pronouns (with the Force Atlas layout and modularity analysis)

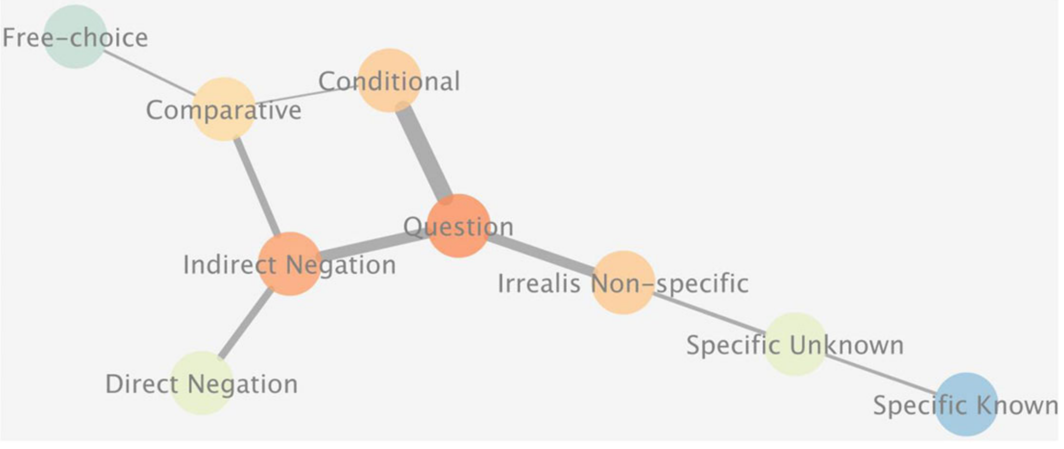

FIGURE 3 Visualization of Haspelmath's (1997) data on indefinite pronouns (with weighted edges and centrality analysis for the nodes)

of visual attributes, and many additional plug-ins. However, note that the layouts in Cytoscape are not as interactive as in Gephi.

Sample Syllabus:

The aim of the following portion of a syllabus is to introduce students to the semantic map model, the different methodologies, and their respective role in modern linguistic debates. The syllabus covers 8 weeks of teaching. It can be envisioned either as a single course on semantic maps or as 4-week units that can fit into general classes on linguistic typology, historical linguistics, and computational linguistics, respectively. In the latter case, the first 2 weeks are common to all classes and introduce the model, while the second two are targeted to the specific fields. Upon completion of the entire module, students (a) will understand in detail the basic assumptions and principles of the semantic map model; (b) will become familiar with the advantages and shortcomings of the method as well as with the different ways of building a map; (c) will have a deep understanding of the literature in this area of research and a grasp of the wide range of linguistic domains in which semantic maps are used; (d) will be able to interpret maps that incorporate information of different kinds, for example, diachronic and weighted semantic maps; and (e) will be able to construct their own semantic maps using language samples of different sizes. The class is geared towards advanced undergraduate or masters students.

Week 1. A general introduction to semantic maps

The syllabus first introduces the main theoretical premises and basic assumptions of the semantic map approach. This is followed by an overview of the different domains in which semantic maps can be used, for studying both grammar and the lexicon. Students will also be familiarized with the two main types of maps, namely the classical and proximity ones. Haspelmath (2003) and van der Auwera and Temürcü (2006) are good resources for background information about the method, while Koptjevskaja-Tamm et al. (2015), with its focus on lexical typology, is particularly useful for highlighting the scope of the model.

Suggested readings:

van der Auwera, J., & Temürcü, C. (2006). Semantic Maps. In K. Brown (Ed.), *Encyclopedia of language and linguistics (2nd ed.), vol. 11* (pp. 131–134). Oxford: Elsevier.

Croft, W., & Poole, K. T. (2008). Inferring universals from grammatical variation: Multidimensional scaling for typological analysis. Theoretical Linguistics, *34*(1), 1–37. https://doi.org/10.1515/thli.2008.001.

Haspelmath, M. (2003). The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In M. Tomasello (Ed.), *The new psychology of language*, *vol. 2* (pp. 211–243). New York: Erlbaum.

Koptjevskaja-Tamm, M., Rakhilina, E., & Vanhove, M. (2015). The semantics of lexical typology. In N. Riemer (Ed.), *The routledge handbook of semantics* (pp. 434–454). London & New York: Routledge.

Week 2. Classical semantic maps including visualization

During the second week, students learn how to build classical semantic maps. They become familiar with the process of meaning identification as well as with the principal constraints in linking the various meanings. The notational conventions used in most studies employing the model are also presented. In terms of reading assignments, the students start with van der Auwera (2013), which presents the different representational possibilities of classical semantic maps. They continue with Zwarts (2010), which discusses the inductive (bottom-up) and the deductive (top-down) approaches to constructing semantic maps. The reading list further includes Cysouw (2007), which proposes to resort to weighted semantic maps for providing visual information about the frequency of polysemy patterns across languages.

Suggested readings:

van der Auwera, J. (2013). Semantic maps, for synchronic and diachronic typology. In A. G. Ramat, C. Mauri & P. Molinelli (Eds.), *Synchrony and diachrony. A dynamic interface* (pp. 153–176). https://doi.org/10.1075/slcs.133.07auw.

Cysouw, M. (2007). Building semantic maps: The case of person marking. In B. Wälchli & M. Miestamo (Eds.), *New challenges in typology* (pp. 225–248). Berlin/New York: Mouton de Gruyter.

Zwarts, J. (2010). Semantic map geometry: Two approaches. Linguistic Discovery, *8*(1), 377–395. https://doi.org/10.1349/ps1.1537-0852.a.357

Typology

Week 3. The scope of the model: The linguistic domains covered

During Week 3, the students gain in-depth knowledge of the different types of meanings that can be included in a semantic map, namely, grammatical, lexical, and constructional. They also become acquainted with methodological issues, such as the process of language sampling in the context of semantic maps. Narrog & Ito (2007) is the point of departure for this week: while its focus is on grammatical semantic maps, it makes a general argument in favor of a systematic use of large datasets to better the quality of semantic maps. François (2008) and Wälchli and Cysouw (2012) follow: They are influential typological studies resorting respectively to connectivity and proximity maps for lexical semantics. Traugott (2016) provides hints on how to use semantic maps with entire constructions (rather than with individual grammatical and lexical items).

Suggested readings:

François, A. (2008). Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In M. Vanhove (Ed.), *From polysemy to semantic change. Towards a typology of lexical semantic associations* (pp. 163–215). Amsterdam/Philadelphia: John Benjamins.

Narrog, H., & Ito, S. (2007). Re-constructing semantic maps: The comitative-instrumental area. STUF - Language Typology and Universals, *60*(4), 273–292. https://doi.org/10.1524/stuf.2007.60.4.273.

Traugott, E. C. (2016). Do semantic modal maps have a role in a constructionalization approach to modals? Constructions and Frames, *8*(1), 98–125. https://doi.org/10.1075/cf.8.1.07tra.

Wälchli, B., & Cysouw, M. (2012). Lexical typology through similarity semantics: Towards a semantic map of motion verbs. In M. Koptjevskaja-Tamm & M. Vanhove (Eds.), *New Directions in Lexical Typology*, [special issue] Linguistics, *50*(3), 671–710. https://doi.org/10.1515/ling-2012-0021

Week 4. Semantic maps in language contact

This week is devoted to demonstrating the applicability of the semantic map model to language contact situations and contact-induced changes. The students first read Gast & van der Auwera (2012) and Tenser (2016) in order to become acquainted with the concept of semantic map harmony (also known as “semantic map assimilation”). Grossman & Polis (2017) comes next so as to illustrate how semantic maps can help when studying the extent to which polysemy networks associated with model language items are copied in the target language. Finally, Gil (2017) is a nice example of the semantic map approach as applied to the analysis of an areal pattern.

Suggested readings:

Gast, V. & van der Auwera, J. (2012). What is 'contact-induced grammaticalization'? Examples from Mayan and Mixe-Zoquean languages. In B. Wiemer, B. Wälchli & B. Hansen (Eds.), *Grammatical replication and borrowability in language contact* (pp. 381–426), Berlin: de Gruyter Mouton.

Grossman, E., & Polis, S. (2017). Polysemy networks in language contact: The borrowing of the Greek-origin Preposition κατά/ⲕⲁⲧⲁ in Coptic. In E. Grossman, P. Dils, T. S. Richter & W. Schenkel (Eds.), *Greek influence on Egyptian-Coptic: Contact-induced change in an Ancient African language* (pp. 335–367). Hamburg: Widmaier Verlag.

Gil, D. (2017). Roon ve, DO/GIVE coexpression and language contact in Northwest New Guinea. *NUSA*, *62*, 41–100.

Tenser, A. (2016). Semantic map borrowing—Case representation in Northeastern Romani Dialects. Journal of Language Contact, *9*(2), 211–245. https://doi.org/10.1163/19552629-00902001

Historical linguistics

Week 5. Diachronic semantic maps: The basics

This week, students learn more about diachronic semantic maps and their use in grammaticalization studies. They will read the scholarly articles listed below, starting with Anderson (1986) and van der Auwera and Plungian (1998), which will give them the theoretical tools to engage in a class discussion about (a) the differences and similarities between diachronic semantic maps and traditional representations of grammaticalization pathways and (b) the types of evidence required in order to "dynamicize" a classical semantic map and how they can be visually represented.

Suggested readings:

Anderson, L. B. (1986). Evidentials, paths of change, and mental maps: Typologically regular asymmetries. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The linguistic encoding of epistemology* (pp. 273–312). Norwood: Ablex.

van der Auwera J., & Plungian, V. A. (1998). Modality's semantic map. Linguistic Typology, *2*(1), 79–124. https://doi.org/10.1515/lity.1998.2.1.79.

Luraghi, S. (2014). Plotting diachronic semantic maps: The role of metaphor. In S. Luraghi & H. Narrog (Eds.), Perspectives on semantic roles (pp. 99–150). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/tsl.106.04lur.

Narrog, H. (2010). A diachronic dimension in maps of case functions. Linguistic Discovery, *8*(1), 233–254. https://doi.org/10.1349/ps1.1537-0852.a.352

Narrog, H., & van der Auwera, J. (2011). Grammaticalization and Semantic maps. In H. Narrog, & B. Heine (Eds.), *The Oxford handbook of grammaticalization* (pp. 318–327), Oxford: OUP. https://doi.org/10.1093/oxfordhb/9780199586783.013.0025

Week 6. Challenges for the diachronic semantic maps and alternative methodologies

This week is devoted to methodological challenges for diachronic semantic maps, mainly linked to the fact that traditional diachronic maps are fundamentally qualitative (with attested vs. non-attested pathways of change). For the visualization of semantic changes based on quantitative (corpus) data, alternative methodologies resorting to multidimensional scaling (Hilpert, 2011), possibly combined with a dynamic Behavioral Profile approach (Jansegers & Gries, 2018), can be envisioned. Yet another option is the "wave" method suggested by Andrason (2016). The pros and cons of the different solutions are listed and discussed with the students.

Suggested readings:

Andrason, A. (2016). From vectors to waves and streams: An alternative approach to semantic maps. Stellenbosch Papers in Linguistics, *45*, 1–29. https://doi.org/10.5774/45-0-211

Hilpert, M. (2011). Dynamic visualizations of language change. Motion charts on the basis of bivariate and multivariate data from diachronic corpora. International Journal of Corpus Linguistics, *16*(4), 435–461.

Jansegers, M. & Gries, S. Th. (in press, 2018). Towards a dynamic behavioral profile: A diachronic study of polysemous 'sentir' in Spanish'. Corpus Linguistics and Linguistic Theory.

Malchukov, A. L. (2010). Analyzing semantic maps: A multifactorial approach. Linguistic Discovery, *8*(1), 176–198. https://doi.org/10.1349/ps1.1537-0852.a.350

Computational linguistics

Background in programming languages such as Python and R is not mandatory (since the exercises below are quite basic), but the classes in Weeks 7 and 8 will be demanding for students who never coded.

Week 7. Graph inference and visualization tools

During this week, the students learn how to create a polysemy matrix for a specific concept or set of concepts (e.g., EARTH/LAND in Figure 1 above) based on the data of CLICS (List et al., 2014). Following the method described in Regier et al. (2013), their task is to infer a semantic map using the Python script made available on the authors' website (http://lclab.berkeley.edu/regier/semantic-maps/) and to visualize the graph with such tools as Gephi or Cytoscape (presented in the *Online material*). In a second step, this network is compared with a Formal Concept Analysis (Ryzhova & Obiedkov, 2017) of the same lexical matrix; for this specific exercise the use of the software solutions *Concept Explorer* (http://conexp. sourceforge.net) or *GaLicia* (http://www.iro.umontreal.ca/~galicia/) is recommended.

Suggested readings:

Angluin, D., Aspnes, J., & Reyzin, L. (2010). Inferring social networks from outbreaks. In M. Hutter, F. Stephan, V. Vovk & T. Zeugmann (Eds.), *Algorithmic learning theory 2010* (pp. 104–118). Berlin: Springer. https://doi.org/10.1007/978-3-642-16108-7_12

List, J.-M., Mayer, T., Terhalle, A., & Urban, M. (2014). CLICS: Database of cross-linguistic colexifications. Marburg: Forschungszentrum Deutscher Sprachatlas (Version 1.0, retrieved from http://CLICS.lingpy.org, accessed on 2017–7-6).

Regier, T., Khetarpal, N., & Majid, A. (2013). Inferring semantic maps. Linguistic Typology, *17*(1), 89–105. https://doi.org/10.1515/lity-2013-0003

Ryzhova, D. & Obiedkov, S. (2017). Formal concept lattices as semantic maps. In Ekaterina L. Chernyak (Ed.), Computational linguistics and language science (pp. 78–87). Aachen CEUR Workshop Proceedings.

Week 8. Semantic maps and dimensionality reduction

The aim of this final week is to apply different statistical methods of dimensionality reduction (principal component analysis [PCA], multidimensional scaling [MDS], T-distributed Stochastic Neighbor Embedding [t-SNE]) to the dataset compiled during Week 7. The students will (a) compare the results produced by different techniques (e.g., Levshina, 2015) and (b) assess the complementarity between proximity and graph-based maps. The user guide and R code used in Croft and Poole (2008) for MDS using optimal classification (https://github.com/jaytimm/MDS_for_Linguists) as well as Levshina's tutorial on how to build semantic maps based on contextual features (http://www.natalialevshina.com/statistics. html) can be good starting points for this task.

Suggested readings:

Croft, W., & Poole, K. T. (2008). Inferring universals from grammatical variation: Multidimensional scaling for typological analysis. Theoretical Linguistics, *34*(1), 1–37. https://doi.org/10.1515/thli.2008.001.

Levshina, N. (2015), *How to do linguistics with R. Data exploration and statistical analysis*, Amsterdam/Philadelphia: John Benjamins.

Wälchli, B., & Cysouw, M. (2012). Lexical typology through similarity semantics: Toward a semantic map of motion verbs. Linguistics, *50*(3), 671–710. https://doi.org/10.1515/ling-2012-0021

Focus Questions

- Explain the two main principles for building a classical semantic map bottom-up and describe the step-by-step procedure to be followed. What are the shortcomings of a purely inductive (bottom-up) approach and how can deductive reasoning (top-down) help improve a semantic map inferred from polysemy data?
- List all the types of information that can be captured by classical semantic maps and describe how these can be represented visually (e.g., frequency of attestation of a meaning might be reflected in the size of the nodes in the graph). How many different kinds of information can be combined in a single semantic map?
- While scholars like Croft have argued that proximity maps (produced with MDS methods) work best with large datasets, others—such as van der Auwera—maintain that connectivity maps possess features that make them more interesting for linguistic typology. Now that classical semantic maps can be inferred from large datasets (Regier et al., 2013), can you reassess the argument of both sides and evaluate the complementarity between the two approaches?
- The semantic map model has been used by researchers in a variety of fields, such as typology, semantics, morphology, historical linguistics, and translation studies. Can you think of an example of its application in each of these domains?
- Diachronic semantic maps are essentially qualitative (attested vs. non-attested pathways of change), while alternative approaches to semantic changes are mostly quantitative, but do not preserve the network-based representation. Can you imagine a way to combine both types of information in a single graph?

TABLE 1 Polysemy data for lexemes expressing the meanings "thread, string" (based on the resources provided by William S. Annis; https://lingweenie.org/conlang/maps/thread.py)

Languages	Forms	Meanings
Malayalam	tantu	filament, tendril, fiber, thread, sinew
Thai	yaawngM	fiber, filament, string, thread, cord, vein
Bambara	gaari	string, thread
Somba-Siawari	kösö	vine, creeper, rope, string, fiber, thread
Turkish	iplik	thread, yarn, fiber, filament
Chinese	xiàn	thread, string, wire, line, clue
Maori	tui	thread, lashing
Russian	верёвка	cord, rope, string
Zulu	intambo	cord, rope, string, thread, wire
Italian	filo	thread, yarn, line, blade, edge, drizzle, string
Buang	aggis	rope, vine, string
Kagwahiva	-enimboa	thread, cord
Nootkah	diqi:ba	net, twine, thread
Rotokas	viripato	thread, string, fishing line, cord
Tlingit	tás	thread, sinew
Inupiaq	ivalu	thread, sinew

Seminar Activity

Divide the class into groups of four. Each group works independently on the small dataset provided in Table 1, with the task of (a) converting the polysemy data provided into a lexical matrix (with language-form as rows and meanings as columns); (b) enriching this matrix with lexical items from languages that are not included in the dataset (one language per student) using dictionaries and/or other lexical resources; (c) manually building a classical semantic map—based on this lexical matrix—that respects both the connectivity hypothesis and the economy principle; and *optionally* (d) using the algorithm provided by Regier et al. (2013; see Week 7) to automatically infer a graph and then comparing the manually and automatically plotted maps.

Afterwards, each group presents its results to the class and assesses the effect of the varying language samples on the maps. All the datasets are then merged to produce a more stable map. This final map is ultimately contrasted with the colexification network of the concept "thread" in CLICS (http://clics.lingpy.org/browse.php?gloss=rope,%20cord), and the tutor highlights the advantages of the semantic map approach.

ORCID

Thanasis Georgakopoulos http://orcid.org/0000-0002-5632-4202
Stéphane Polis http://orcid.org/0000-0003-0350-9753

How to cite this article: Georgakopoulos T, Polis S. Teaching & Learning Guide for: The semantic map model. *Lang Linguist Compass*. 2018;12:e12286. https://doi.org/10.1111/lnc3.12286