USING CAUSAL INFERENCE TO DETECT DIRECTIONAL TENDENCIES IN SEMANTIC EVOLUTION

JOHANNES DELLERT
Seminar für Sprachwissenschaft, Universität Tübingen
Tübingen, Germany
jdellert@sfs.uni-tuebingen.de

This paper proposes a novel application of causal inference in the area of semantic language evolution, which attempts to infer unidirectional trends of lexical change exclusively from massively cross-linguistic dictionary data. First, we show how colexification between concepts can be modeled mathematically as mutual information between concept variables. Core notions of causal inference (most prominently, the unshielded collider criterion) are then applied to predict the dominant directionality in pathways of semantic change. The paper concludes by revisiting a few well-known examples of synchronic polysemies, and illustrating how the method succeeds in building hypotheses about their historical development.

1. Introduction

As a language evolves over time, all parts of the language system undergo constant and sometimes drastic change. Compared to phonological and syntactic change, semantic change is much more subject to historical developments and accidents, making it notoriously erratic and hard to predict (Hollmann, 2009). Still, the long search for patterns which allow to predict at least some kinds of semantic change has started to yield some promising results. Most prominently, Traugott and Dasher (2002) manage to establish a general tendency for words to evolve towards more speaker-oriented meanings.

If for a given concept, we want to predict how the words for that concept are likely to evolve, we have the problem that extralinguistic factors can strongly influence each individual case, and can easily invert the expected course of events. For instance, while it is quite common for words denoting large crops to be used colloquially as words for the concept HEAD:N (e.g. French *citron* “lemon”, German *Birne* “pear”, and Russian *repa* “turnip”), the reverse development has occurred in Thai, where the word *hua* “head” has become the count noun for large fruit.

Therefore, we can always only talk about tendencies, and a typical question we might hope to answer has the following form: Is a word denoting the concept X more likely to additionally develop the sense Y, than a word denoting Y is to develop the sense X? For the concepts SUN:N and DAY:N, it seems obvious that semantic change will always occur in one direction. But are such unidirectional patterns exceptional, or do they occur along many pathways in conceptual space? Previous attempts to identify such tendencies have relied on intimate knowledge of historical developments within a small number of languages, e.g. English and Japanese in the case of Traugott and Dasher (2002). Such a small language sample detracts from the reliability of broad generalizations, a problem which also concerns more recent computational approaches such as Sagi, Kaufmann, and Clark (2011). If we want to come to more general conclusions, we will need to consider a wider range of languages, as did classical work in lexical typology like Viberg (1984) or Wilkins (1996) for small semantic domains.

To gather evidence for and against a postulated unidirectional change, we could extensively study all available historical sources and etymological literature for a large sample of languages. Unfortunately, only very few languages and even fewer language families are sufficiently well-documented over relevant timespans. Moreover, the semantic reconstructions presented in etymological dictionaries are often speculative, and therefore cannot be used as empirical evidence.

Evans and Wilkins (2000) propose synchronic polysemies as a more readily available source of evidence for semantic evolution, and argue that cross-linguistic polysemies should be used systematically to constrain the semantic reconstructions postulated in etymologies. Analyses of semantic change processes have established (Zalizniak et al., 2012) that a word undergoing semantic change will typically pass an intermediary polysemous stage where it denotes both the original and the newly developing sense. Synchronic polysemies thus provide us with a snapshot of semantic evolution in action, which we can exploit as observational data in order to retrieve hints about tendencies in their historical development.

With massively cross-linguistic lexical databases slowly becoming available, we can use the power of statistical methods to leverage large amounts of polysemy data for inferring new general tendencies in semantic evolution, or to gather additional evidence in favor of or against existing theories. The purpose of this work is to show that causal inference has the potential to become such a method.

2. Causal Inference

Causal inference (Pearl, 2009) is a relatively recent branch of statistics which has developed partial solutions to the classical problem that correlation between two statistical variables does not imply a causal relationship. By taking additional variables and their covariance patterns into account, the existence and the direction of immediate causation between statistical variables can often be inferred from observational data alone, given some reasonable background assumptions.

The central idea of causal inference is to exploit patterns of conditional independence. A correlation between two variables may vanish when conditioning on other variables, i.e. considering every combination of their possible values separately. The pattern in which some correlations disappear upon conditioning on

sets of other variables provides hints which help us to exclude some causal hypotheses, sometimes leaving only one possible direction of causation.
Causal inference depends on a stability condition which is equivalent to postulating that a true causal connection between two variables can safely be assumed to exist if the dependence between them cannot be explained away by observing any subset of the remaining variables. Applying this criterion in a principled way to a larger number of variables gives us what is called a **causal skeleton**, an undirected graph linking all pairs of variables whose covariance cannot be explained away by conditioning on other possibly intervening variables.
The key to turning such a skeleton into a partially directed causal graph is to consider **unshielded triples**, i.e. triples of variables of the form $A - B - C$. Consider the different conditional independence patterns one would expect for all possible causal patterns. If the true pattern is $A \leftarrow B \rightarrow C$, we would expect some correlation between A and C which disappears when conditioning on the common cause B. An analogous argument applies to the cases $A \rightarrow B \rightarrow C$ and $A \leftarrow B \leftarrow C$. By contraposition, we can thus infer the **unshielded collider** $A \rightarrow B \leftarrow C$ if conditioning on B was not necessary to explain away any possible correlation between A and C. This pattern of reasoning lies at the heart of causal inference algorithms such as the PC algorithm (Spirtes, Glymour, & Scheines, 2000), and is exactly what we will use to infer the directionality of lexical change.

3. Measuring Conditional Independence between Concepts

In order to apply causal inference to the domain of semantic evolution, we model concepts as statistical variables. Our variables will be language-independent concepts represented by German (or English) glosses, and the observations are realizations of these concepts across a large number of languages. If we do this for two related concepts, there will be some languages where the two concepts are **colexified**, i.e. they can be denoted by the same polysemous lexeme. For instance, the concepts SUN:N and DAY:N are colexified because the Hungarian word *nap* denotes both senses (as do equivalents in many other languages).
Assuming a conceptual space which is given as a set of language-independent senses, the subset which can be expressed by a word (or lexeme) w in some language is called the **isolectic set** of that word (François, 2008). The isolectic set of Hungarian *nap* could be represented as {SUN:N, DAY:N, SOLAR:A}.
For ease of exposition, we assume a one-to-one-mapping from German (or English) glosses to language-independent concepts. In my data, each “concept” is defined by a single gloss in German. In what follows, “colexification” therefore means that two German glosses occur together on one side of a dictionary equation, and should perhaps more accurately be called co-translation into German. Because rare polysemies represent intermediary stages of semantic evolution and are not stable over time, it is possible to count every colexification equally, without correcting for genealogical relatedness.

We now turn to measuring the connectedness of concepts based on overlaps of isolectic sets across many languages. Mathematically, we will model stochastic dependence between two concepts in terms of non-vanishing mutual information. Observations will be in the form of isolectic sets, and the mutual information of variables for colexified concepts will be non-zero. The joint information measure R over sets of variables will be designed to be a **submodular information measure**, which means that it meets the following axioms:

1. $R(\emptyset) = 0$
2. $S \subseteq T \Rightarrow R(S) \leq R(T)$ for all sets of variables S and T
3. $R(S) + R(T) \geq R(S \cup T) + R(S \cap T)$ for all sets of variables S and T

Every submodular information measure gives rise to a measure of conditional mutual information which still has all the essential properties needed for causal inference (Steudel, Janzing, & Schölkopf, 2010).
To measure joint information in a set of concepts $\{c_1, \ldots, c_n\}$, we use a very simple and trivially submodular measure R based on the sets $iso(c_i)$ of isolectic sets containing each concept c_i:

$$R(c_1, \ldots, c_n) := \left| \bigcup_{i=1}^{n} iso(c_i) \right|$$

This is mathematically equivalent to the measure in Steudel et al. (2010, Section 5.4), where the authors propose to use an analogous measure on sets of content words to measure causal influences between texts.
Informally, the joint information content R of a set of concepts is thus the number of isolectic sets in which these concepts are involved. If the isolectic sets of two different lemmas have exactly the same elements, they are still counted separately. Whenever an isolectic set covers multiple concepts from the set $\{c_1, \ldots, c_n\}$, $R(c_1, \ldots, c_n)$ will be different from the sum $\sum_{i=1}^{n} iso(c_i)$. It is this difference that the resulting measure of mutual information will quantify. In the case of two concepts, the derived mutual information $i(c_i, c_j)$ simply counts the number of isolectic sets in which the two concepts c_i and c_j co-occur, i.e. the number of colexifications in the data:

$$\begin{aligned} i(c_i, c_j) &:= R(c_i) + R(c_j) - R(c_i, c_j) \\ &= |iso(c_i)| + |iso(c_j)| - |iso(c_i) \cup iso(c_j)| = |iso(c_i) \cap iso(c_j)| \end{aligned}$$

For instance, if our entire dataset consisted of three lemmas with the isolectic sets $\{$SUN:N,DAY:N$\}$, $\{$SUN:N,SOLAR:A$\}$, and $\{$SUN:N,DAY:N,SOLAR:A$\}$, we would have $i(\{$SUN:N,DAY:N$\}) = 3 + 2 - 3 = 2$.

Conditional mutual information between two concepts c_i and c_j given a set of concepts $S := \{s_1, \ldots, s_n\}$ is then defined in the following way:

$$i(c_i, c_j; S) := R(c_i, s_1, \ldots, s_n) + R(c_j, s_1, \ldots, s_n) \\ - R(c_i, c_j, s_1, \ldots, s_n) - R(s_1, \ldots, s_n)$$

The submodularity of the information measure R ensures that $i(c_i, c_j; S)$ is always nonnegative (Steudel et al., 2010, Lemma 1).

Intuitively, $i(c_i, c_j; S)$ counts the colexifications between c_i and c_j which cannot be explained away by colexification with any of the concepts in S. In our example, we have $i(\{$SUN:N,DAY:N;SOLAR:A$\}) = 2 + 2 - 2 - 1 = 1$, but $i(\{$DAY:N,SOLAR:A;SUN:N$\}) = 2 + 2 - 2 - 2 = 0$, which means that DAY:N and SOLAR:A are independent given SUN:N, and we get the unshielded triple DAY:N$-$SUN:N$-$SOLAR:A in the causal skeleton.

Turning to the question how to detect the directionality of the established causal links (or possible paths of semantic evolution), we reconsider the inituition behind the unshielded collider criterion. To infer a causal arrow $c_1 \to c_2$, we need a third concept c_3 which forms a unshielded collider $c_1 \to c_2 \leftarrow c_3$. This means that if c_1 and c_3 are colexified, none of the isolectic sets in question must extend only to c_2, but there need to be one or several other concepts c_s on which we can condition to remove the link between c_1 and c_3. Whenever we see such a configuration, it becomes more likely that words for c_2 were extended to cover the sense c_1, because otherwise we would expect some of these words to also have been extended to c_3. For instance, if words for UNDERSTAND:N were likely to evolve into words for HEAR::V, the colexification between SEE::V and UNDERSTAND:N would cause some isolectic areas to cover all three concepts. The absence of such isolectic areas provides us with evidence that the true pattern is much more likely to be SEE::V $\to$ UNDERSTAND::V $\leftarrow$ HEAR::V.

If we check for many different concepts c_3 whether they form unshielded colliders together with c_1 and c_2, the small sample size will often lead to conflicting evidence, i.e. some unshielded colliders which imply $c_1 \to c_2$ and others which imply $c_2 \to c_1$. In such a case, a scoring scheme can be used to decide whether one of the directions is more probable. The current version of my implementation simply weights each arrow $c_1 \to c_2$ for each concept c_3 where $R(c_1, c_3) > 1$ or $R(c_2, c_3) > 1$ by the factor $w(c_1 \to c_2; c_3) := \frac{R(c_1,c_2) \cdot R(c_3,c_2)}{R(c_2)}$, i.e. the number of colexifications between c_1 and c_3 which we would have expected if the causal pattern were $c_1 \leftarrow c_2 \to c_3$ or $c_1 \leftarrow c_2 \leftarrow c_3$. If the summed arrow score $sc(c_1 \to c_2) := \sum_{c_3} w(c_1 \to c_2; c_3)$ is more than 20% higher than $sc(c_2 \to c_1)$, the current implementation returns the arrow $c_1 \to c_2$.

4. Examples

The ideal way to evaluate the method's potential would be to collect a large number of clear-cut example cases where the etymological literature across language families only reconstructs semantic evolution in one direction. Unfortunately, even the largest available database of semantic shifts (Zalizniak et al., 2012) only contains very few instances of multiply attested unidirectional semantic shifts, and many of these involve very specialized concepts which one cannot expect to find across many dictionaries (e.g. *amber*, *catnip*, or *woodlouse*).
Studies on cross-linguistic patterns of semantic change such as Koch (2008) do yield some more useful examples, but they typically specialize on very small semantic domains. It is thus extremely difficult to find enough examples of cross-linguistically attested unidirectional semantic shifts for a numerical evaluation in terms of precision and recall. Considering only some very prominent examples instead, it is at least possible to illustrate how the inference method works, and to highlight both the power and the inadequacies of causal inference in this application. To ensure reproducibility of these initial results, isolectic sets for all three examples, and log files allowing to trace the computations in detail, are available as supplementary materials on the author's webpage.

4.1. *Example 1: The Eye of a Needle*

Consider the three concepts EYE:N, EAR:N, and EYEOFNEEDLE:N. Some languages use the same word for EYE:N and EYEOFNEEDLE:N, as English does. Other languages like Polish and Korean use the word for EAR:N to denote the same concept. In either case, it is clear that the words for body parts are used by metaphorical extension to describe an oblong hole in a needle, and any method for analysing semantic shifts should infer that semantic evolution will occur exclusively in the directions EAR → EYEOFNEEDLE and EYE → EYEOFNEEDLE. In this simple case, it turns out that both of the the desired causal arrows can be derived from a single unshielded collider EYE → EYEOFNEEDLE ← EAR. Table 1 displays some of the relevant isolectic sets. The collider is trivial to infer be-

Table 1. Isolectic sets spanning EYEOFNEEDLE:N and EYE:N or EAR:N.

Language	Lemma	Isolectic Set
Basque	*begi*	{EYE:N, KNAG:N, EYEOFNEEDLE:N, STITCH:N, DROPOFGREASE:N, CHEESEHOLE:N}
Dutch	*oog*	{EYE:N, LOOP:N, EYEOFNEEDLE:N}
Korean	*gwi*	{EAR:N, SPOUT:N, CORNER:N, EYEOFNEEDLE:N}
Livonian	*sīlma*	{EYE:N, LOOP:N, SHACKLE:N, EYEOFNEEDLE:N}
Nenets	*xa*	{EAR:N, HANDLE:N, EYEOFNEEDLE:N}
Polish	*ucho*	{EAR:N, HANDLE:N, EYEOFNEEDLE:N}

cause there is no lemma in any language which covers both EAR and EYE, while

both concepts are clearly colexified with EYEOFNEEDLE. We therefore have an unshielded triple where EYEOFNEEDLE is not needed to separate EAR and EYE, because these two concepts are already independent unconditionally. To illustrate the causal argument, for any other causal pattern the expected number of colexifications $w(c_1 \rightarrow c_2; c_3)$ between $c_1 = \text{EYE}$ and $c_3 = \text{EAR}$ is 2.377, as opposed to zero colexifications in the data.
Overall, building on 130 isolectic sets from 77 languages across 19 language families, the algorithm returns EYE $\rightarrow$ EYEOFNEEDLE with a score ratio of 1.269. For EAR $\rightarrow$ EYEOFNEEDLE, the score ratio based on 112 isolectic sets from 76 languages across 20 families is as high as 2.765. In both cases, we get a clear result in favor of the expected directionality, showcasing that the method is able to detect a causal signal representing widespread metaphorical extension.

4.2. *Example 2: Counting and Calculating*

Many languages use the same word for the concepts of counting and calculating. Instances of isolectic sets subsuming both concepts are given in Table 2. From a historical perspective, the counting sense arguably is the earlier one, calculation being a later cultural achievement. Is this fact mirrored by the causal signal we can detect from synchronic polysemies? Based on 134 isolectic sets from 68

Table 2. Isolectic sets spanning COUNT:V and CALCULATE:V.

Language	**Lemma**	**Isolectic Set**
Coptic	*op*	{COUNT:V, CALCULATE:V, ESTIMATE:V}
Czech	*počítat*	{COUNT:V, CALCULATE:V}
Indonesian	*membilang*	{COUNT:V, CALCULATE:V, NARRATE:V}
Udmurt	*lydjany*	{COUNT:V, CALCULATE:V}
Spanish	*contar*	{COUNT:V, CALCULATE:V, NARRATE::V}

languages covering 21 language families, we get a score ratio of 1.162 in favor of COUNT:V $\rightarrow$ CALCULATE:V. The evidence is thus not strong enough to make a decision, although it does point into the expected direction. Investigation of the isolectic sets in question shows that the signal would be quite a bit stronger if we could remove the effects of polysemy in the German gloss *zählen* "to count", which is also used in the senses "to be valid" and "to have (inhabitants)".

4.3. *Example 3: Hoping and Expecting*

Finally, we take a look at a pair of verbal concepts whose semantic relationship is not clear on external grounds. Between the concepts of HOPE:V and EXPECT:V (which are frequently colexified), is there a unidirectional pattern we can retrieve from the data? Existing theories appear not to make any prediction here, since both concepts refer to mental states, meaning that they belong to roughly the same level of abstraction. Table 3 gives a number of polysemous verbs which synchronically denote both HOPE:V and EXPECT:V.

Table 3. Isolectic sets spanning HOPE:V and EXPECT:V.

Language	**Lemma**	**Isolectic Set**
Chinese	*xīwàng*	{HOPE:V, EXPECT:V, WISH:V}
Hebrew	*jixel*	{EXPECT:V, HOPE:V}
Japanese	*nozomu*	{EXPECT:V, HOPE:V, WISH:V}
Portuguese	*esperar*	{HOPE:V, EXPECT:V, WAIT:V}
Turkish	*ummak*	{HOPE:V, EXPECT:V, WAIT:V}

Perhaps surprisingly, the result is very strong. Based on 203 isolectic sets from 70 languages covering 22 families, the score ratio is 2.813 in favor of HOPE:V $\rightarrow$ EXPECT:V. The prediction this implies is that words for HOPE:V are more likely to develop the additional sense EXPECT:V than the reverse pattern. This result seems plausible in the light of examples known to the author, e.g. the attested development of Latin *spērāre* into Spanish *esperar*, or the Finnish verb *toivoa* "to hope", which nowadays is also used e.g. for expectations from a person. The algorithm has provided us with a hypothesis based on massively cross-linguistic data, and we could now look into individual language histories to verify this claim, or to find counter-examples.

5. Conclusion

In this paper, we have established that causal inference can be applied to an information geometry defined by cross-linguistic polysemies in order to measure causal influences between concepts. The resulting causal structures can be taken to indicate probable vectors of semantic expansion. For three examples, we have seen that the method does manage to recover some hidden diachronic information from purely synchronic polysemies, and leads to plausible results.

Since much larger sample sizes are typically needed to guarantee correct results in causal inference, the approach must not be interpreted as producing objective proofs of historical events. However, the method does provide an unbiased summary of large amounts of easily available data which are too varied and extensive to be processed by a human expert. It allows us to quickly derive interesting hypotheses about possible directional patterns of semantic evolution, which can then be verified and further elucidated based on the documented history of various languages. Moreover, the new computational tool for quickly developing initial hypotheses about the directionality of semantic shifts will be helpful for researchers seeking to shed more light on this central aspect of language evolution.

Acknowledgments

This research has been supported by the ERC Advanced Grant 324246 EVOLAEMP, which is gratefully acknowledged. The author wishes to thank all members of the EVOLAEMP group for their valuable comments, and three anonymous reviewers for their helpful suggestions.

References

Evans, N., & Wilkins, D. (2000). In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language*, *76*(3), 546–592.

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: Benjamins.

Hollmann, W. B. (2009). Semantic change. In J. Culpeper, F. Katamba, P. Kerswill, & T. McEnery (Eds.), *English language: description, variation and context* (p. 301-312). Basingstoke: Palgrave.

Koch, P. (2008). Cognitive onomasiology and lexical change: Around the eye. In M. Vanhove (Ed.), *From Polysemy to Semantic change: Towards a Typology of Lexical Semantic Associations* (pp. 107–137). Amsterdam: Benjamins.

Pearl, J. (2009). *Causality.* Cambridge University Press.

Sagi, E., Kaufmann, S., & Clark, B. (2011). Tracing semantic change with Latent Semantic Analysis. In K. Allan & J. A. Robinson (Eds.), *Current Methods in Historical Semantics* (pp. 161–183). Berlin, Boston: De Gruyter.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, Prediction, and Search* (2nd ed.). MIT Press.

Steudel, B., Janzing, D., & Schölkopf, B. (2010). Causal markov condition for submodular information measures. In A. Kalai & M. Mohri (Eds.), *Proceedings of the 23rd annual conference on learning theory* (pp. 464–476). Madison, WI, USA: OmniPress.

Traugott, E., & Dasher, R. (2002). *Regularity in semantic change.* Cambridge University Press.

Viberg, A. (1984). The verbs of perception: a typological study. *Linguistics*, *21*(1), 123–162.

Wilkins, D. P. (1996). Natural tendencies of semantic change and the search for cognates. In M. Durie & M. Ross (Eds.), *The Comparative Method reviewed: Regularity and irregularity in language change* (pp. 264–304). New York: Oxford University Press.

Zalizniak, A. A., Bulakh, M., Ganenkov, D., Gruntov, I., Maisak, T., & Russo, M. (2012). The catalogue of semantic shifts as a database for lexical semantic typology. *Linguistics*, *50*(3), 633–669.