Probabilistically Modeling Semantic Change

Aleksandr Nisnevich
David Hall
Daniel Klein

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2015-36
http://www.eecs.berkeley.edu/Pubs/TechRpts/2015/EECS-2015-36.html

May 1, 2015

Probabilistically Modeling Semantic Change

by Aleksandr Nisnevich

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II**.

Approval for the Report and Comprehensive Examination:

Committee:

Professor Dan Klein
Research Advisor

(Date)

* * * * * * *

Professor Pieter Abbeel
Second Reader

(Date)

Probabilistically Modeling Semantic Change

Alex Nisnevich, David Hall, and Dan Klein
EECS, UC Berkeley
alex.nisnevich@berkeley.edu, {dlwh,klein}@cs.berkeley.edu

Abstract

We present a new model of semantic change that traces the evolution of a word's meaning through a phylogeny. By mapping meaning changes to WordNet paths, we are able to identify common kinds of semantic change. We find optimal weights for our model using expectation-maximization (EM), and use the results to conjecture about the relative frequency of different kinds of semantic change.

1 Introduction

Lexical semantic change has been poorly understood. We know the ways in which lexical change can manifest itself, but we still have only a rough sense of which changes are the most frequent, or even how often changes tend to occur overall. Do words tend to become more general over time, as in the case of English *dog*? Or more specific, like English *deer*? Is there a "rate" of semantic change, in analogy to the (controversial) rate of cognate loss from [10]? In this paper, we take a computational approach to studying semantic change, building a probabilistic model that simulates the evolution of meanings of words over time.

Previously, authors have investigated computational approaches to phylogenetics and dating, offering insights into the spread of large language families like the Indo-European family and Austronesian language family [5]. More recently, efforts have been made at phonetic reconstruction using probabilistic models [2]. However, unlike other diachronic phenomena such as sound change, semantic change has been difficult to model because semantic relationships between meanings are difficult to define systematically.

Historically, semantic change has been considered to be " 'fuzzy', highly irregular, and difficult to predict"[13]. In contrast to this view, Wilkins (1993) introduced the idea that there are natural tendencies of semantic change that can be rigorously studied for the purposes of historical reconstruction [13], particularly in the semantic domain of body parts.

For example, one of the natural tendencies proposed by Wilkins is that "for a term for a visible person-part to shift to refer to the visible whole for which it is a part, [while] the reverse change is not natural."[13] In other words, Wilkins claims that within the semantic field of body parts, *part holonymy* (part-to-whole semantic change, such as *brain* → *head*) should be much more common than *part meronymy* (whole-to-part semantic change, such as *belly* → *navel*).[1]

[1]While there are other types of meronymy/holonymy, such as *substance meronymy/holonymy* and *member meronymy/holonymy*, our model predicts these types of semantic change to be rare enough to be negligible, so for the remainder of the paper I will use *part holonymy* and *holonymy* interchangeably, and likewise for *part meronymy* and *meronymy*.

In line with Wilkins' position, we propose the use of probabilistic models as a way of estimating the parameters governing semantic change. In this paper, we present a new model of lexical semantic change that traces the evolution of a word's meaning through a phylogeny (sections 2 and 3), as well as a procedure for training this model using available semantic networks and machine learning techniques (sections 4 and 5).

2 Definitions

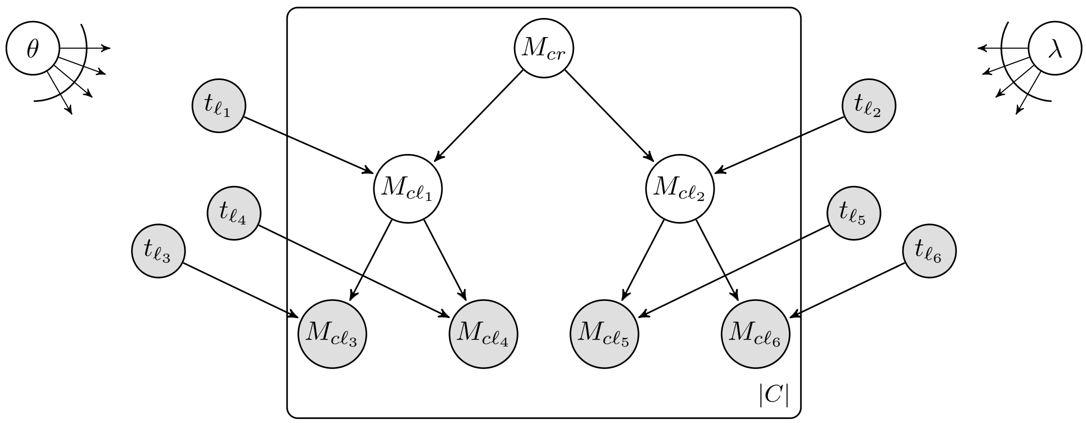

Figure 1: **Graphical representation of our time-aware model for a sample phylogeny**. (Note that in general, only a subset of the full phylogeny is relevant for any given cognate set.)

- $|C|$ is the number of cognate sets
- M_{cr} represents the reconstructed meaning of cognate set c in the root language
- $M_{c\ell}$ represents the meaning of cognate set c in language ℓ, with the shaded nodes representing observed languages
- t_ℓ is the length (in years) of the branch leading to language ℓ
- λ is the model parameter specifying the overall rate of change
- θ is the model parameter specifying the multinomial distribution from which semantic changes are chosen

We assume the existence of a (known) phylogeny T over languages, indexed by ℓ. The phylogeny T also has branch lengths t. We will denote the length of a branch leading to language ℓ as t_ℓ. We also assume the existence of a set C of cognate sets W_c, with reflexes $W_{c\ell}$ in each modern language, though reflexes are possibly unobserved in some or many languages.

Our interest is not in the words themselves, but in their meanings. Specifically, we wish to model how words change their meanings as they evolve through time. We assume that each reflex of each root W_{cr} has some meaning M_{cr}. Each reflex's meaning is generated conditional on the meaning of that word's parent. That is, the probability of change $p(M_{c\ell} = m' | M_{c\mathbf{par}(\ell)} = m, T = t_\ell)$, for meanings m and m' and where $\mathbf{par}(\ell)$ is the parent of the language ℓ. We are interested in learning the parameters of this distribution.

3 The Model

In our model (see Figure 1), the distribution $p(M_{c\ell} = m' | M_{c\mathbf{par}(\ell)} = m, T = t_\ell)$ is parameterized as a continuous time Markov jump process, where the number of jumps (meaning changes) is governed by a Poisson process. That is, the number of changes in a length of time t is given by $\mathrm{Poi}(t\lambda)$, where λ is the overall rate of change.

Meanings change according to a random walk on a *semantic graph* – a labeled multigraph consisting of edges connecting meanings labeled by the kind of change. When a change happens, the kind of edge to take is chosen according to some multinomial distribution θ, and then the new meaning is chosen uniformly from among those meanings that are connected to the current label by that meaning. If no such edge is available, the process repeats until an acceptable walk is found. We are ultimately interested in the parameters λ and θ, telling us the rate of semantic change and the relative likelihood of different kinds of changes, respectively.

The likelihood is given by:

$$\begin{aligned} & p(M_{c\ell} = m' | M_{c\mathbf{par}(\ell)} = m; T = t_\ell, \lambda, \theta) \\ & = \sum_{\pi \text{ is from } m \text{ to } m'} \left(\mathrm{Poi}(|\pi|; t\lambda) \cdot \prod_{e \in \pi} p(\mathrm{type}(e); \theta) \cdot \frac{1}{\#\text{neighbors of } m \text{ with type}(e)} \right) \end{aligned} \quad (1)$$

Through the rest of this paper, we abbreviate $p(M_{c\ell} = m' | M_{c\mathbf{par}(\ell)} = m; T = t_\ell; \theta; \lambda)$ by $\mathcal{P}(m' | m; t_\ell; \theta; \lambda)$, and omit the parameters θ, λ when they are implied.

4 WordNet Search

In order for our model to be trained, we must first determine the nature of the semantic relationship between each pair of meanings within each cognate set. We employ WordNet[4], a semantic graph of 117,000 nodes (synsets), each of which consists of one or more wordforms, a definition, and a set of semantic relationships to other nodes.

To find WordNet paths between arbitrary pairs of meanings, we perform a depth-limited breadth-first search. However, as Table 1 shows, a simple BFS yields poor path coverage even at a relatively high depth[2]. This happens for two reasons. First, WordNet primarily only covers relationships of hyper/hyponymy and holo/meronymy, so any pair of meanings that are related in a more complicated way will most likely not be linked on WordNet. Secondly, many of the meanings that we're interested in (e.g. "small bone") are not in WordNet at all, because WordNet only contains meanings that can be expressed as single lexical entries in English.

To deal with the first issue, we modify our search procedure by using the definition (gloss) of each node to find potential relationships to other nodes (described in section 4.1). To deal with the second issue, we use a set of parsing rules to connect complex meanings to existing meanings in WordNet (described in section 4.2). As Table 1 shows, these two techniques together result in an enormous improvement in WordNet coverage for meaning pairs.

[2]Trying a depth higher than 3 would both be very computationally expensive and lead to linguistically unsound results, since it would be very strange for the meaning of a word to change in such a complicated way in a single transition.

Coverage by depth and search type	Depth 1	Depth 2	Depth 3
BFS	7.2%	9.8%	13.8%
BFS with gloss search (4.1)	13.8%	23.3%	39.5%
BFS with gloss search and complex meanings (4.2)	15.4%	27.2%	**46.6%**

Time taken (in seconds)	Depth 1	Depth 2	Depth 3
BFS	50.3 s	63.7 s	270 s
BFS with gloss search (4.1)	59.1 s	97.2 s	574 s
BFS with gloss search and complex meanings (4.2)	86.2 s	194 s	1876 s

Table 1: Percentage of meaning pairs in the Indo-European body parts semantic field that have WordNet paths between them, using different search methods and depths, as well as the amount of time that each method takes (for a set of 93 protoforms, each having between 2 and 74 meanings).

Condition	**Path type**
1. Dest contains "part of (Det) Node"	MERONYM
2. Node contains "part of (Det) Dest"	HOLONYM
3. Dest contains "kind of (Det) Node"	HYPONYM
4. Node contains "kind of (Det) Dest"	HYPERNYM
5. Dest contains "Node" *or* Node contains "Dest"	"related"

Table 2: Rules used to determine the kind of relationship that a given node (Node) has to the destination node (Dest) by comparing the definitions and wordforms provided by WordNet for each node. Here, *X contains "Y"* means that X's definition contains one of Y's wordforms, and *(Det)* represents the optional presence of a determiner (e.g. *a*, *an*, *the*).

4.1 Searching through Glosses

To deal with the fact that the set of relationships that a given WordNet node has to other nodes is frequently incomplete, we additionally compare each node's definition and wordforms to the definition and wordforms of the destination node, using the list of conditions in Table 2 to determine when two nodes are related, and what kind of relationship they have.[3]

For example,

- The nodes [step, **stair**: support consisting of a place to rest the foot while ascending or descending a stairway] and [tread: structural member consisting of the horizontal **part of a stair** or step] are given a *meronym* link (by condition #1).
- The nodes [**body**: the entire structure of an organism] and [blood: the fluid that is pumped through the **body** by the heart] are given a *related* link (by condition #5).

[3] A quick note on the semantic terminology used here: if X is a kind of Y, then X is a *hyponym* of Y and Y is a *hypernym* of X. If X is a part of Y, then X is a *meronym* of Y and Y is a *holonym* of X. If nodes X and Y are related in some way but we can't easily characterize the relationship between them, we simply call them *related*.

Treat ...	as ...
“Det N"	“N"
“Adj N"	Hyponym(“N")
“kind of N"	Hyponym(“N")
“part of N"	Meronym(“N")
“N_1 of N_2"	Hyponym(“N_1") *or* Meronym(“N_2")

Table 3: Rules used to connect more complex meanings to meanings already in WordNet. (N = noun, Adj = adjective, Det = determiner)

4.2 Connecting Complex Meanings to WordNet

Many meanings that we are interested in do not appear verbatim in WordNet but can be connected to meanings in WordNet through kind-of (hyponymy) or part-of (meronymy) relations. We accomplish this by running the meaning through the Stanford part-of-speech tagger[11] and applying the rules in Table 3, repeating the process recursively until either a meaning within WordNet is found or no more rules can be applied.

For example,

- “small bone" = Hyponym(“bone")
- “part of the hand" = Meronym(“hand")
- “part of a horse hoof" = Meronym(Hyponym(“hoof"))
- “ham of horse" = Hyponym(“ham") *or* Meronym(“horse").

Note that a given meaning could potentially map to multiple WordNet nodes. In the case that there are multiple start and/or end nodes, the WordNet search will find paths between *any* start node and *any* end node.

4.3 Overall Algorithm

The overall WordNet search algorithm, incorporating the improvements described above, is:

```
FIND-WORDNET-PATH(startMeaning, destMeaning, depth)
startPaths ← FIND-SYNSETS-AND-PATHS(startMeaning)
destPaths ← FIND-SYNSETS-AND-PATHS(destMeaning)
pathsFound ← ∅
for d = 1 to depth do
   for all startPath ∈ startPaths do
      for all destPath ∈ destPaths do
         startSynset ← SYNSET(startPath)
         destSynset ← SYNSET(destPath)
         if startSynset = destSynset then
            add (startPath + REVERSED(destPath)) to pathsFound
         else if ARE-RELATED(startSynset, destSynset) then
```

```
            link ← RELATED-LINK(startSynset, destSynset)
            add (ADD-LINK(startPath, destSynset, link) + REVERSED(destPath)) to pathsFound
        end if
      end for
    end for
    newPaths ← ∅
    for all path ∈ startPaths do
      for all (dest, link) ∈ GET-LINKED-SYNSETS(path) do
        newPath ← ADD-LINK(startPath, dest, link)
      end for
    end for
    startPaths ← newPaths
  end for
  return pathsFound
```

where:

- FIND-SYNSETS-AND-PATHS(*meaning*) returns a set of *path*s corresponding to the given *meaning*, if any exist. Each *path* consists of a *synset* (i.e. WordNet node) and a list of *link*s. Table 3 is used to determine the links (if any) that connect meanings to WordNet nodes. For example, we can represent FIND-SYNSETS-AND-PATHS("small bone") as $\{(\text{Bone}, [\text{Hyponym}])\}$, where BONE is the synset that has "bone" as a wordform.[4]

- SYNSET(*path*) returns the synset of the given *path*.

- REVERSED(*path*) returns a reversed copy of *path* – for example, for an arbitrary synset s,

  $$\text{Reversed}\big((s, [\text{Hyponym}, \text{Related}, \text{Meronym}])\big) = (s, [\text{Holonym}, \text{Related}, \text{Meronym}])$$

- ARE-RELATED(*start*, *dest*) uses Table 2 to determine if there is any relationship between the *start* and *dest* synsets.

- RELATED-LINK(*start*, *dest*) uses Table 2 to return the link type between the *start* and *dest* synsets, assuming that they are related.

- ADD-LINK(*path*, *synset*, *link*) returns a copy of *path* with *link* added to its list of links and its synset replaced with *synset*.

- GET-LINKED-SYNSETS(*startSynset*) returns a set of (*destSynset*, *link*) pairs representing destination synsets related to *startSynset* and the *link* between each pair of synsets.

- + is the concatenation operator for paths:

  $$(s_1, [link_{11}, \cdots, link_{1m}]) + (s_2, [link_{21}, \cdots, link_{2n}]) = (s_2, [link_{11}, \cdots, link_{1m}, link_{21}, \cdots, link_{2n}])$$

4.4 Making WordNet Relationships More Linguistically Plausible

One type of semantic change that is held to be natural, especially within the semantic field of body parts, is the change between parts in the same category (e.g. *snout* ↔ *nose* or *beak* ↔ *face*)[13].

[4] In practice, WordNet usually has multiple synsets corresponding to any given word.

However, WordNet generally represents transitions of this nature as two changes (*hyponymy* $\rightarrow$ *hypernymy*) rather than one. In order for our model to treat this kind of transition as a single change, we compress [HYPONYM, HYPERNYM] sequences within paths into a single change that we call SISTER.

Additionally, when two meanings are part of the same synset in WordNet, we treat the transition between them as a "free" change – in other words, links between meanings that belong to the same synset are not counted in the length of a path.[5]

5 Training the Model with EM

To find the weights for our model, we use EM. In the Expectation step, we compute the set of expected changes $\pi_{c\ell}$ for each cognate set c and language ℓ, and the likelihood of each such change $L_{c\ell}$, via Viterbi. In the Maximization step, we maximize the log-likelihood $L(\theta, \lambda; \mu)$ to find the optimal parameters θ, λ.

5.1 The Expectation Step

In the E-step, we first determine the most likely meaning for each language within each cognate set (that is, $M_{c\ell}\ \forall c \in C, \ell \in T$) by the Viterbi algorithm[12].

We compute the $\uparrow$-messages from the leaves to the root:

$$\mu_{\ell \to \mathbf{par}(\ell)}(m) = \max_{m'} \mathcal{P}(m'|m; t_\ell)\mu_{\ell\uparrow}(m') \tag{2}$$

$$\alpha_{\ell \to \mathbf{par}(\ell)}(m) = \arg\max_{m'} \mathcal{P}(m'|m; t_\ell)\mu_{\ell\uparrow}(m') \tag{3}$$

$$\mu_{\ell\uparrow}(m) = \begin{cases} -\infty & \text{if } \ell \text{ is an observed language and } m \text{ is not an observed meaning for } \ell \text{ in } c \\ \sum_{\ell' \in \mathbf{children}(\ell)} \mu_{\ell' \to \ell}(m) & \text{otherwise} \end{cases} \tag{4}$$

where $\mathcal{P}(m'|m; t_\ell)$ is defined as in (1).

We then proceed from the root r downwards to find the most likely meaning for each language.

$$M_{cr} = \arg\max_{m} \mu_{\ell\uparrow}(m) \tag{5}$$

$$M_{c\ell} = \alpha_{\ell \to \mathbf{par}(\ell)}(M_{c\mathbf{par}(\ell)}) \quad \forall \ell \neq r \tag{6}$$

Next, for each language within each cognate set, we determine the most likely change from $m = M_{c\mathbf{par}(\ell)}$ to $m' = M_{c\ell}$ by finding the path π from m to m' that maximizes the path likelihood:

$$\prod_{e \in \pi} \left(p(\text{type}(e); \theta) \cdot \frac{1}{\#\text{neighbors of } m \text{ with type}(e)} \right) \tag{7}$$

The E-step returns the set of expected $\pi_{c\ell}$s and the likelihood of each $\pi_{c\ell}$, which we will call $L_{c\ell}$.

[5] However, treating same-synset transitions as free chances can cause unexpected problems for us due to the fact that we store meaning strings rather than synsets in M. This is discussed further in section 7.1.

5.2 The Maximization Step

In the M-step, we maximize the sum of the log likelihoods of each transition taken.

Let $\pi_{c\ell}$ be the most likely change from $M_{c\mathbf{par}(\ell)}$ to $M_{c\ell}$ and $L_{c\ell}$ be the likelihood of $\pi_{c\ell}$.

We can now compute the new λ and θ parameters and the objective as follows:

$$\lambda = \frac{\#\text{ individual changes}}{\text{total time span}} = \frac{\sum_{c\in C}\sum_{\ell}|\pi_{c\ell}|}{|C|\sum_{\ell} t_\ell} \tag{8}$$

$$\forall\text{ type}\quad \theta_{\text{type}} = \frac{\#\text{ changes of given type}}{\#\text{ individual changes}} = \frac{\sum_{c\in C}\sum_{\ell}\sum_{e\in\pi_{c\ell}}\mathbb{I}[e=\text{type}]}{\sum_{c\in C}\sum_{\ell}|\pi_{c\ell}|} \tag{9}$$

$$obj = \sum_{c\in C}\sum_{\ell}\log L_{c\ell} \tag{10}$$

5.3 The EM Algorithm

The overall algorithm is:

EM(C)
$threshold \leftarrow 0.01\cdot|C|$
$\theta \leftarrow [0,\ldots,0]$
$\lambda \leftarrow 10^{-4}$
$i \leftarrow 0$
$obj_0 \leftarrow \infty$
repeat
  $i \leftarrow i+1$
  for all $c\in C$ **do**
    $\pi_c, L_c \leftarrow$ Expectation(c,θ,λ)
  end for
  $obj_i,\theta,\lambda \leftarrow$ Maximization(C,π,L,θ,λ)
until $|obj_i - obj_{i-1}| < threshold$
return θ,λ,obj_i

6 Experiment

To get a measure of our model's accuracy and to see what predictions it makes, we tested it on the 156 cognate sets in the Body Parts category within the Indo-European Lexicon assembled by the LRC [8]. In order to improve WordNet coverage, we ignored all meanings that were not nouns and threw out the cognate sets whose reconstructed protoform wasn't a noun, leaving 93 cognate sets. Semantic paths were found between 46.6% of possible meaning pairs in this dataset. For our phylogeny, we used a distance-labelled Indo-European phylogeny adapted from [3].

6.1 Accuracy

We assessed our model through a protoform reconstruction task. For each cognate set, we compared the root meaning (protoform) predicted by our model to the annotated protoform given in the dataset, and we scored our model based on the percentage of protoforms that it predicted correctly.

We compared the performance of our model in this task to two other models: a *parsimony* model that takes into account time-spans but not WordNet paths (essentially treating every path as a *none* path) and a *time-unaware* log-linear model that takes into account WordNet paths but not time-spans. We also compared each model against a *most common meaning (MCM)* baseline score, obtained by simply taking the meaning that occurs most frequently among observed languages for each cognate set.

	Body Parts only (93 sets)	All semantic fields (430 sets)
Our model	45.2%	30.7%
Parsimony	40.7%	31.6%
Time-unaware	38.7%	27.4%
MCM	46.6%	45.1%

As shown above, our model outperformed the other two when trained on the Body Parts semantic field. None of the models performed as well as the MCM baseline on this task, but this is not surprising: the LRC dataset is largely based on the work of Pokorny[6], a German linguist, and thus appears to both have on average a higher coverage of Germanic reflexes than of those in other families and a slight bias toward picking Germanic meanings as the annotated protoform. Thus, MCM has a appears to have a structural advantage in this task that the models do not have.

When we tried expanding the experiment to cover all semantic fields (as opposed to just Body Parts), all of models performed significantly worse, and our model ended up performing slightly worse than the parsimony model. The most likely explanation for this is that the meanings in the Body Parts semantic field tend to have well-defined relationships between them, while meanings in other semantic fields tend to have more poorly defined relationships between them.

6.2 Results

After training on the Body Parts semantic field, the final parameters are $\lambda = 4.05 \cdot 10^{-5}$ and θ as shown in Table 4 below.

Semantic change type	**Expected count**
related	33.81%
hypernym	16.79%
none	15.11%
hyponym	12.47%
sister	11.51%
part_holonym	5.04%
part_meronym	4.08%
all others	1.20%

Table 4: Counts for different semantic changes predicted by our model.

7 Conclusion

The results in 6.1 are promising in that they suggest that, at least within the Body Parts semantic field, WordNet paths and timespans are both important factors in our model. However, the task that we are using to measure model accuracy is problematic one, since it relies on data (reconstructed protoforms) whose accuracy is generally difficult to demonstrate. Perhaps a better task (albeit one that cannot be automated) would be to assess the opinions of historical linguists regarding how plausible the trees reconstructed by our model look compared to those reconstructed other models. Another interesting task would be to use our model to try to reconstruct the phylogeny itself, using techniques similar to [3], and to see how similar the result is to established Indo-European phylogenies.

As far as the questions that motivated this research go, the final λ predicts that semantic changes occur on average once every 23,700 years in each branch of the phylogeny for any given word[6], and the final θ predicts that most common categorizable semantic changes are hypernymy, hyponymy, and sisterhood (hyponym $\rightarrow$ hypernym), in that order, with hypernymy being about 1.3 times as common than hyponymy. In line with Wilkins's theory, our model predicts holonymy to be more common than meronymy for body parts, though the difference is less extreme than what Wilkins predicted.

Our results do have a high margin of error, because almost half of the changes predicted are either not categorizable (*related*) or entirely semantically implausible (*none*). Both the hypernymy $>$ hyponymy and the holonymy $>$ meronymy result are likely to not be statistically significant. In order to improve on this, we would most likely have to improve our WordNet coverage.

7.1 Further Steps

There are two areas that could potentially be improved upon: the model itself and the WordNet search.

One major drawback of the current model is that it only considers meanings that have already been observed as possible candidates for the unobserved $M_{c\ell}$s. This is a problem when the meaning that we want at a given node is unobserved – for example, *eye* and *ear* could both derive from *head* by meronomy, but our model would not be able to predict *head* for a parent node if it is not observed anywhere else. Making our model consider all of WordNet as a possible destination for each transition would be computationally infeasible for our exact E-step approach, so we would have to perform an approximate E-step instead, using a Markov chain Monte Carlo method such as Metropolis-Hastings.

The fact that our WordNet search method can still only find paths for 46.6% of possible meaning pairs is likely to be a big obstacle in the performance of our model. Improvement could come in the form of more parsing rules of the sort in Table 3, or by augmenting our usage of WordNet with additional sources of semantic relationships, such as the augmented WordNets generated in [9]. It would also be helpful to try a similar approach to parts of speech other than nouns, using alternative semantic resources such as FrameNet[1] or VerbNet[7].

It may also be helpful to make the model explicitly model transitions between WordNet nodes (that is, synsets) rather than meaning strings. Right now, when a path is found between m and m',

[6] Perhaps a better way of thinking about this is: "Every year, any given word has on average a $4.05 \cdot 10^{-5}$ chance of changing its meaning in any given language."

that path can be between any two WordNet nodes that have m and m' as wordforms, respectively, but because the identities of the nodes themselves are not stored in M, this means that m' could refer to completely different senses of a given word in $m \to m'$ and $m' \to m''$ transitions. This is particularly prevalent when m' is a particularly common word with many disparate definitions.

References

[1] C. Baker, C. Fillmore, and J. Lowe. The Berkeley FrameNet project. In *Proceedings of the COLING-ACL*, pages 86–90, Montreal, Canada, 1998. University of Montreal.

[2] Alexandre Bouchard-Côté, David Hall, Thomas L. Griffiths, and Dan Klein. Automated reconstruction of ancient languages using probabilistic models of sound change. 2013. doi: 10.1073/pnas.1204678110.

[3] Will Chang, David Hall, Chundra Cathcart, and Andrew Garrett. Dating Proto-Indo-European: A revised computational analysis supports the steppe hypothesis. Proceedings of the International Conference on Historical Linguistics, 2013.

[4] Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA, 1998.

[5] Russell D Gray and Quentin D Atkinson. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, 426(6965):435–439, 2003.

[6] J. Pokorny. *Indogermanisches etymologisches Wörterbuch*. Number v. 2 in Indogermanisches etymologisches Wörterbuch. Francke Verlag, 1959. URL `http://books.google.com/books?id=WUJiAAAAMAAJ`.

[7] Karin Kipper Schuler. *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. PhD thesis, University of Pennsylvania, 2006. URL `http://verbs.colorado.edu/ kipper/Papers/dissertation.pdf`.

[8] Jonathan Slocum. Building an Indo-European lexicon using print and online dictionaries. URL `http://www.utexas.edu/cola/centers/lrc/papers/IElex.html`.

[9] Rion Snow, Daniel Jurafsky, and Andrew Y. Ng. Semantic taxonomy induction from heterogenous evidence. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 801–808, Sydney, Australia, July 2006. Association for Computational Linguistics. doi: 10.3115/1220175.1220276. URL `http://www.aclweb.org/anthology/P06-1101`.

[10] Morris Swadesh. Towards greater accuracy in lexicostatistic dating. *International journal of American linguistics*, pages 121–137, 1955.

[11] Kristina Toutanova, Dan Klein, Christopher Manning, and Yoram Singer. Feature-rich part-of-speech tagging with a cyclic dependency network. Proceedings of HLT-NAACL, pages 252–259, 2003.

[12] A.J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *Information Theory, IEEE Transactions on*, 13(2):260–269, April 1967. ISSN 0018-9448. doi: 10.1109/TIT.1967.1054010.

[13] David P Wilkins. *From part to person: Natural tendencies of semantic change and the search for cognates*. Cognitive Anthropology Research Group at the Max Planck Institute for Psycholinguistics, 1993.