The Rationality of Semantic Change

Omer Korat
omerk@stanford.edu

Stanford University

Introduction

This study argues that semantically related function words affect each other's distribution over syntactic environments. Words that can have the same meaning are observed to have opposite trends of change in frequency across different syntactic structures which correspond to the shared meaning. This phenomenon is demonstrated to have a rational basis: it increases communicative efficiency by prioritizing words differently in the environments on which they compete. As words immigrate to new syntactic environments over time, they tend to push out words that populated these environments prior to immigration.

The idea of semantic contrast being a catalyst for change has been explored previously [1]. The principle is based on the intuition that when speakers choose linguistic expressions, they do so because they mean something they would not mean by some other expression. In Rational Speech Acts Theory [2] these notions have been generalized to account for pragmatic inference in the general case by assuming that the probability of an utterance is proportional to its information gain over its cost. Here, these ideas are applied to provide a rational account for the interaction between the changes of semantically related words.

The Prediction

When two words w_1 and w_2 can have the same meaning in some syntactic environment E, speakers can gain extra information by partitioning E into subsets X and Y, and use w_1 more frequently in X, and w_2 more frequently in Y. This increases the informativity of utterances by making it easier to retrieve the intended sub-environment.

For example, such an interaction is predicted between *hence* and *therefore*. Originally, *hence* was a locative used to indicate place of origin ("from hence"). An increase in the DM use of *hence* would lead to competition with *therefore* on the sentence-initial DM environment ("hence, John is smart"), so this environment would be partitioned into sentence-initial DM and mid-sentence DM ("John is therefore smart"), such that *hence* is preferred in the former, and *therefore* is preferred in the latter. This way, speakers can save cognitive effort when choosing among DMs which compete on the same meaning - namely, justification.

References

[1] Eve V Clark. On the pragmatics of contrast. *Journal of child language*, 17(02):417–431, 1990.

[2] Michael C. Frank and Noah D. Goodman. Predicting pragmatic reasoning in language games. *Science*, 336(3):998–998, 2012.

[3] Beatrice Santorini Kroch, Anthony and Lauren Delfs. The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME). CD-ROM, first edition, release 3, 2004. Department of Linguistics, University of Pennsylvania.

[4] Arja Nurmi Anthony Warner Susan Pintzuk Taylor, Ann and Terttu Nevalainen. The York-Helsinki Parsed Corpus of Early English Correspondence (PCEEC). Oxford Text Archive, first edition, 2006. Department of Linguistics, University of York.

Setup

The co-distributions of groups of words that compete on related uses in the Penn Parsed Corpora of Early Modern English [3] and the Parsed Corpora of Early English Correspondence [4] were collected. All formally distinguishable patterns of each word were identified throughout the corpus. For example, the contrast use of *but* was annotated as a conjunction, while the exception use was annotated as a preposition. For each comparison set of words W which compete on some environment E, the pattern which defines E was chosen to be the weakest possible regular expression over tree structures which captures exactly one of the uses of each $w \in W$.

The words selected for this study are *very*, *thus*, *but*, *except*, *though*, *therefore*, *still*, *yet*, *from*, *hence*, *as* and *when*.

Example: Still and *yet* compete on their positive polarity use (denoted by the variables *_adv_pos, demonstrated in (1). Additionally, they compete on an adverbial use (*_adv) following a raised clause introduced by a complementizer/preposition, as demonstrated in (2).

(1) And consequently, they may still with greater ease begin with it, ...

(2) If I can come again, we are still to have our ball.

Statistical Model

Relative and absolute frequency counts were collected for each use of each word, and the counts for each competing pair were compared to each other over time. The hypothesis states that for each pair of uses competing on an environment, there exists a point in time t such that one of the uses becomes more frequent following t and the other becomes less frequent following t, meaning the trends cross at some point in time. Formally, the trends (i.e. true population regression lines) for the frequencies of each word should have opposite slopes. This means that for two uses U and V of a word, there is a point in time (τ') starting from which, one use's frequency grows over time, while the other use's frequency decreases over time. Formally, we have that:

$$|F\tau'(U - F'_\tau(V)| = O((\tau - \tau')^2)$$

For some point in time τ, Where F'_τ measures the frequency of a use at point τ', following the standard definition for Big O complexity.

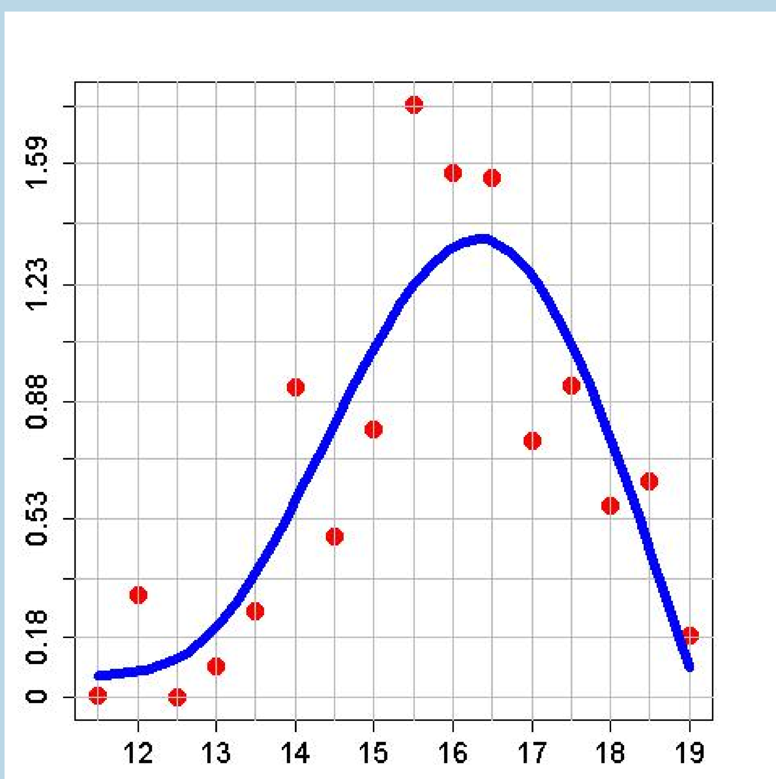
That is, there is a point in time τ starting from which the difference between the frequencies grows quadratically, meaning the trends have to cross.

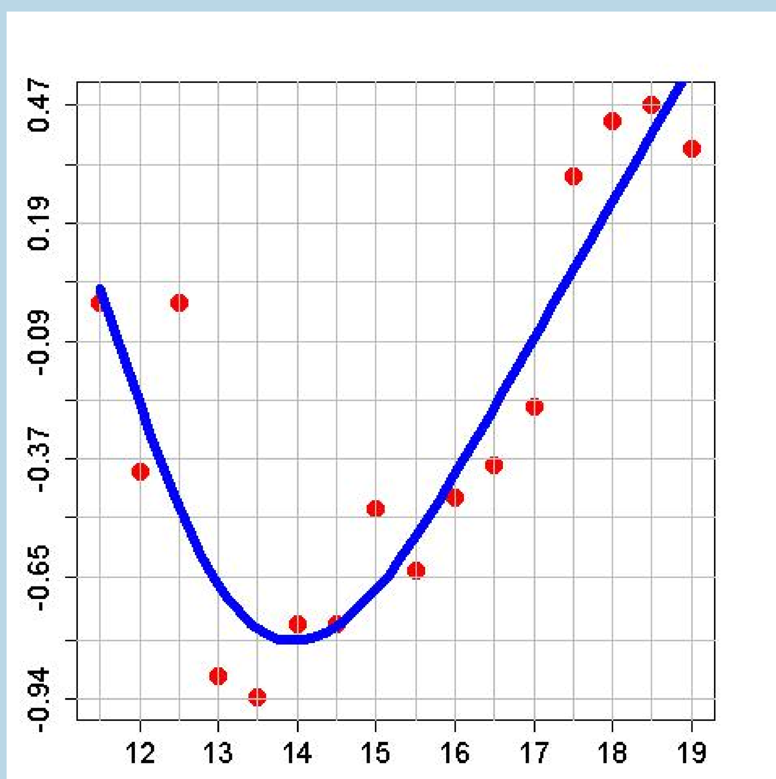
Results

Coefficients and significance levels for all comparison sets are displayed in Table . Each model has 3 coefficients, since the 2 knots partition the intervals into 3 parts. Frequency differences by century are plotted along with model curves. All models were significant, with the exception of the model for *thus* and *very*. All words, with the exception of *still*, were found to change as a linear function of time starting from some year, as described above.

Env.	Use 1	Use 2	Coef1	Coef2	Coef3	Signif. level
Temporal Adv.	`still_adv`	`yet_adv_pos`	0.83	1.35	0.45	**
Degree/Manner	`very`	`thus_adv_deg`	-0.16	-0.06	-0.0072	.
Justification	`hence_dp`	`therefore_dp_top`	-0.1	-0.73	1.22	***
Locative	`hence_adv`	`from`	1.04	1.56	-0.43	***
Exception	`but_p`	`except`	2.06	0.65	-1e-04	***
Contrast	`though`	`but_contrast`	-3.53	-1.62	-1.42	**
Temporal Comp.	`as_p`	`when_p`	0.1	-1.03	0.12	*

Trends by use. Slopes are measured starting from the century the trend started. Signif. codes: *** 0.001 ** 0.01 * 0.05 . .1





Plots by comparison sets for the exception words (right) and the justification words (left). Plots show the difference between the relative frequencies over time. The parabolic shape of the fit corresponds to the X shape formed by the frequency trends.