

Semantic Changes in Apparent Time

JEAN-PHILIPPE MAGUÉ

Université de Lyon

0. Introduction

Semantic changes have been scientifically studied for more than 150 years (Nerlich 1992). All along this history, successive generations of scholars have adopted at least three different theoretical frameworks (Magué 2005). Chronologically, the first trend focused on the identification of the different kinds of semantic changes a lexeme can undergo. This *taxonomist* trend culminates with Ullman (1962). The second trend adopts a *typologist* point of view and is characterized both by the advocacy of cross-linguistic studies and the focus on semantic field rather than isolated lexemes. A typical work in the trend is Viberg (1983). Finally, a *cognitivist* trend has more recently emerged, which aims at explaining the cognitive mechanisms that underlie semantic change (e.g., Sweetser 1990).

Despite the great variety of theoretical approaches the study of semantic changes has gone through, the methodologies used have surprisingly remained the same: only completed semantic changes are studied, either by the analysis of synchronic manifestations, i.e., polysemy or sets of cognates, or by the analysis of the development of a new meaning from corpus evidences. What makes this fact even more surprising is that, on the other hand, the study of phonological changes has undergone a methodological revolution (which has entailed theoretical breakthroughs) during the last 40 years with the emergence of the Labovian variationist sociolinguistics (Labov 1963, 2001).

Sociolinguistics studies the correlation between linguistic variation and socio-economic factors. Among those factors, age of the speaker is of particular interest. Assuming the *Apparent Time Hypothesis* (Bailey et al. 1991), which holds that speakers acquire their idiolect mainly during a critical period in their childhood, correlation between age and linguistic variation is the synchronic manifestation of a change in progress. Most of ongoing linguistic changes observed that way are phonological changes (Labov 1963), few are morphosyntactic ones (Parrott 2002), but, to our knowledge, semantic changes have remained left aside from variationist sociolinguistics. A possible explanation for this state of affairs lies in methodological difficulty to measure precisely and objectively enough the semantic variation. While phonetic variation is directly observable from speakers'

productions, since sounds are precisely the public part of linguistic communication, meanings are mental entities and are not directly made public during communication. The experimenter who wishes to study the semantic variation faces thus the double challenge of obtaining an objective representation of the private mental meaning and of measuring semantic variation from this representation. The goal of this paper is to present a method to achieve this double challenge.

The method presented here is based on the work in the field of quantitative anthropology of Romney et al. (2000) which addresses the issue of inter-cultural differences in the representation of various cultural domains. The main idea it relies on, is to apply statistical treatment to semantic similarity judgments performed by speakers between words belonging to a same semantic field.

1. Materials and Method

The semantic field we analyzed was built with the French word *maison* ‘house’ and 20 of its synonyms¹ chosen for their high frequency variation between the first and the second halves of the 20th century (Table 1). Subjects were given a questionnaire, presenting pairs of words followed by a 10cm-long axis. Each of the 210 possible pairs was presented once to each subject, and the order of the words within the pairs was counterbalanced between subjects. The order of the pairs was randomized for each subject. Subjects had to judge the semantic similarity of each pair of words by placing a mark on the corresponding axis: on the left extremity for unrelated words, on the right one for perfect synonyms and on intermediate positions for intermediate semantic similarities. The position of the marks were measured and scaled to lie between 0 and 1. The answers of each subject i were then represented by a 21×21 symmetric matrix A_i . Few subjects skipped some of the pairs. To deal with the missing values, all the subsequent analyses were performed on the matrices $M_i = (\text{corr}(A_i) + I) / 2$, where $\text{corr}(A_i)$ is the correlation matrix of A_i . This operation had also the effect of removing noise from the data, since in the matrix M_i , the similarity between two words is given by the similarity judgment patterns of those two words against all others.

The experiment was performed on two groups of native French speakers, differing only on their mean age. The younger group was composed of 47 subjects (36 females, i.e., 76.6%) with a mean age of 21 years (± 1.4 , min = 17, max = 26). The older group was composed of 16 subjects (11 females, i.e., 68.8%) with a mean age of 56 years (± 3.1 , min = 49, max = 63). While the two groups differed in terms of age ($t(61) = 33.98$, $p < 10^{-15}$), they were matched in respect to gender ($\chi^2(1, N=63) = 0.388$, $p > .5$). In order to match the groups in their level of competence in French, subjects were asked the number of pages written in French they read per day. In the younger group, subjects read on average 39.1 pages while in the older group subjects read 25.2 pages ($t(44) = 0.91$, $p > .35$).

- (1) The 21 words used in the experiment, their frequencies during the first and second halves of the 20th century and the variation of their frequency

¹ Obtained from the online synonyms dictionary <http://dico.isc.cnrs.fr>

Semantic Changes in Apparent Time

between the two periods. Frequencies were obtained from the online dictionary *Le Trésor de la Langue Française Informatisé*.²

		Frequency 1900 -1950 (per million)	Frequency 1950 - 1999 (per million)	Frequency variation (%)
<i>case</i>	'hut'	982	1,352	+38
<i>chalet</i>	'chalet'	353	1,106	+213
<i>château</i>	'castle'	10,743	5,802	-46
<i>chaumière</i>	'thatched cottage'	1,236	314	-75
<i>clinique</i>	'private clinic'	417	1,038	+149
<i>construction</i>	'construction'	3,442	4,811	+40
<i>entreprise</i>	'enterprise'	3,032	6,104	+101
<i>établissement</i>	'establishment'	1,930	2,999	+55
<i>firme</i>	'firm'	56	3,441	+6,045
<i>habitation</i>	'dwelling'	1,081	659	-39
<i>hôpital</i>	'hospital'	2,707	3,922	+45
<i>immeuble</i>	'building'	784	1,135	+45
<i>intérieur</i>	'interior'	15,514	24,662	+59
<i>logement</i>	'accommodation'	1,456	995	-32
<i>logis</i>	'home'	3,279	1,582	-52
<i>maison</i>	'house'	63,774	50,652	+21
<i>manoir</i>	'manor'	459	285	-38
<i>masure</i>	'hovel'	749	399	-47
<i>propriété</i>	'property'	8,326	4,764	-43
<i>réduit</i>	'cubbyhole'	381	531	+39
<i>résidence</i>	'residence'	424	574	+35

2. Results

2.1. Direct Analyses

When all the judgments are considered together, the older group judged the words significantly more similar, albeit very slightly, than the younger one. Their average scores were respectively 0.58 and 0.56 ($t(27781) = 4.64, p < 10^{-5}$). Considering each pair of words individually, four of them presented this difference: *chalet / chaumière*, *château / clinique*, *château / entreprise*, *château / établissement*, and *château / hôpital* (all p 's < 0.01). Nevertheless, *château / intérieur* and *immeuble / réduit* were judged more similar by the younger group than by the older one (p 's < 0.01).

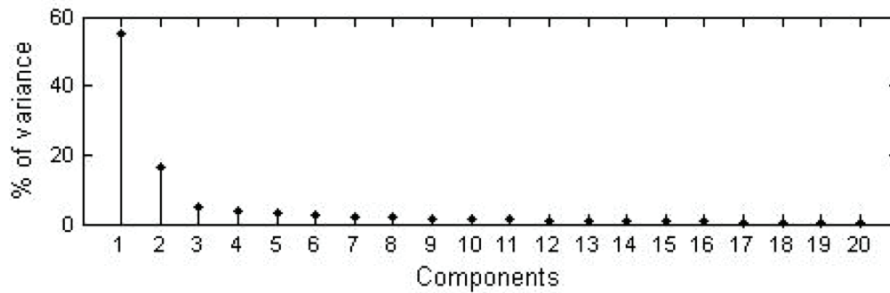
2.2. Principal Components Analysis

Following Romney et al. (2000), we have performed a *Principal component analysis* (PCA) in order to obtain for each subject a semantic space in which words

² <http://atilf.atilf.fr>

are represented by points, and in order to obtain the semantic similarity between two words by the geometric distance between the points representing them. The percentage of variance explained by each component is given in figure 2. Given those percentages (the first and the second component explaining respectively 55% and 17 % of the variance), only the two first components will be considered in the rest of the paper. Before focusing on inter-generational differences in the semantic spaces, it is worth considering the common semantic space.

- (2) Percentage of variance explained by the components. The two first explain together 72% of the total variance.



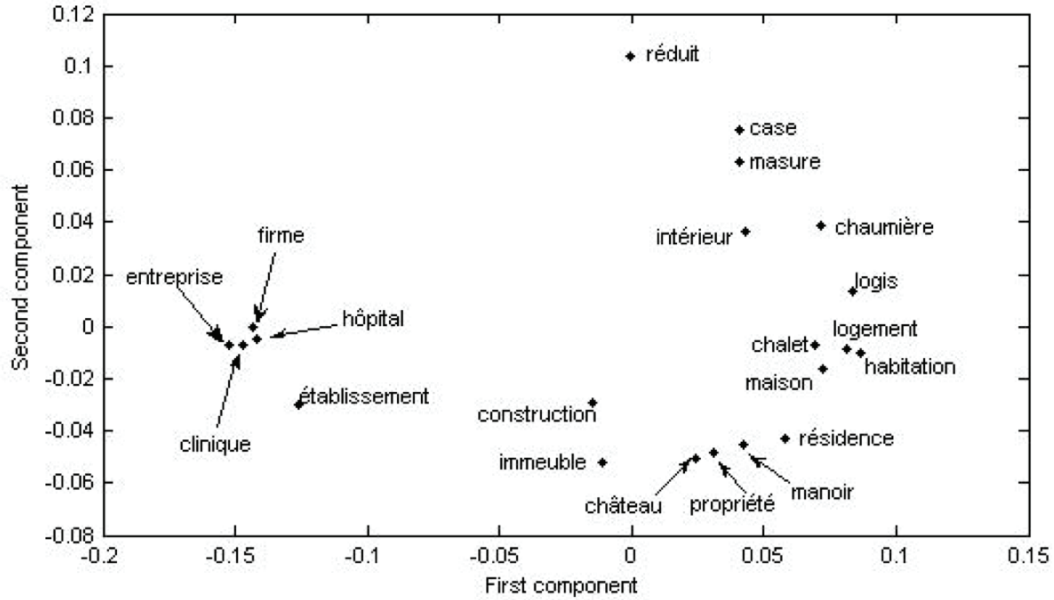
2.2.1. Common Semantic Space

In the common semantic space the position of the words is the average position across all the subjects, regardless of their group (figure 3). Along the first component, words spread from *entreprise, firme, hôpital, clinique* to *logis, logement, habitation*. The first component thus discriminates words along the *habitability* of their referents. Along the second component, words spread from *château, propriété, manoir* to *maison, case, réduit*, making this component axis of *quality* of the lodging.

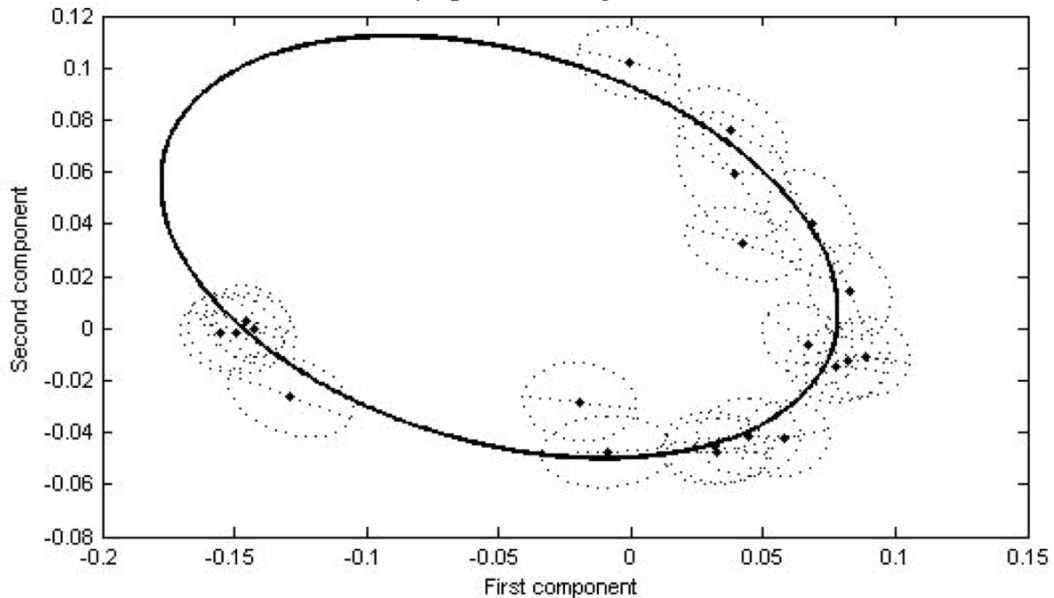
In addition to these axes of habitability and quality, this semantic field is organized by an ellipse (Figure 4, solid line) along which most of the inter-individual variability spreads (Figure 4, dotted lines). As we shall see in the next section, semantic changes occur along this ellipse too.

Semantic Changes in Apparent Time

- (3) Common Semantic Space. The first component describes the *habitability* of the referent of the words, and the second one the *quality* of lodging.



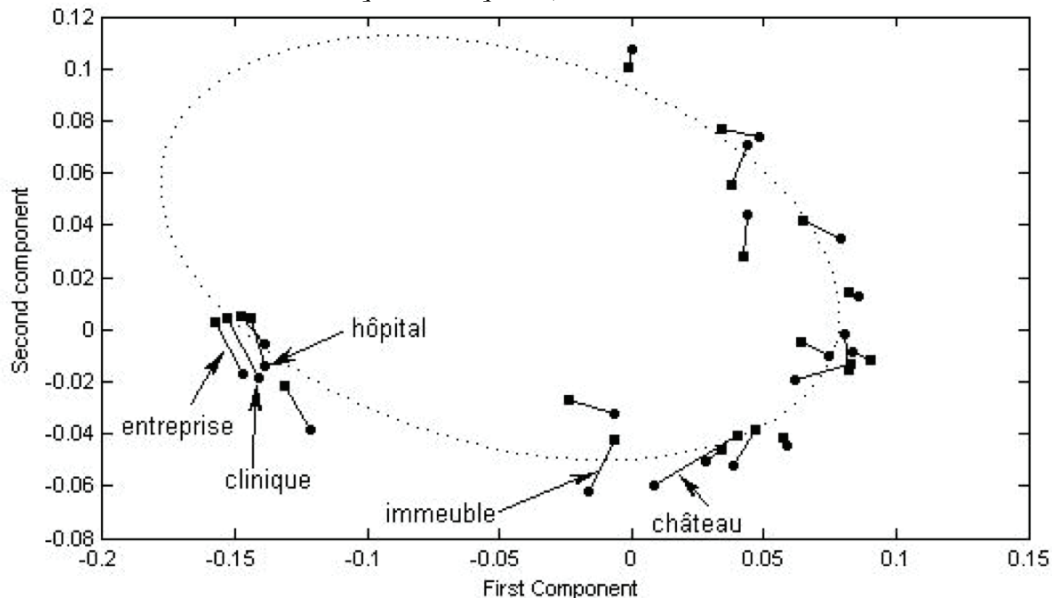
- (4) Elliptical organization of the semantic field (solid line) and confidence ellipses (dotted lines) at $\sigma/2$ with their main axis. Those axes tend to be tangent to the large ellipse structuring the semantic field, indicating that the inter-individual variability spreads along it.



2.2.2. Inter-Generational Differences

A preliminary way to identify inter-generational differences is to compare the mean position of the words between the two groups (figure 5). For most of the words, the change occurs along the ellipse. This change is statistically significant (Hotelling T^2 test) for four of the words: *château* ($T2(T^2)(2) = 18.88, p < .005$), *clinique* ($T2(2) = 7.45, p < .05$), *entreprise* ($T2(2) = 7.00, p < .05$), and *immeuble* ($T2(2) = 6.69, p < .05$). The change for *hospital* also tends to be significant ($T2(2) = 5.88, p = .063$).

- (5) Local inter-generational differences. The mean position of the words is represented for each of the two groups. This position is statistically different for *château*, *clinique*, *entreprise*, and *immeuble*.

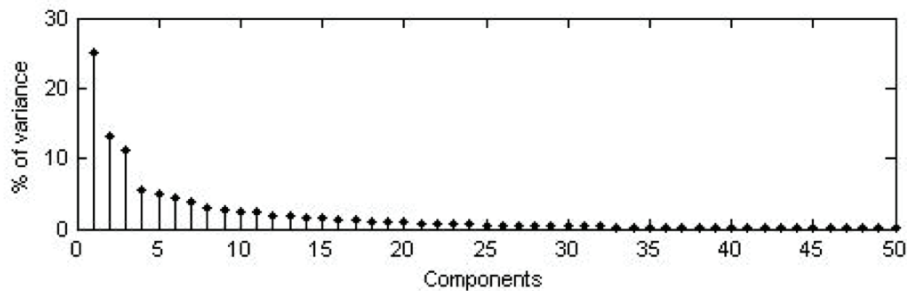


Inter-group differences serve as an alternative to the search for local differences in the semantic spaces of the subjects. To explain the inter-groups differences is to look at differences in the global organization of the semantic field. To address this issue, Romney et al. (2000) proposed that the semantic space of each subject should be represented by the shape of the configuration of the points, which is fully encoded in the set of the 210 distances between all pairs of points. This set of distances is given for each subject s by a 21×21 matrix $D^s = (D^s_{ij})$, where D^s_{ij} is the distance between the points corresponding to the words i and j in her semantic space. This matrix being symmetric, each subject s can be represented by the vector d^s built with the value above the diagonal of D^s , i.e. a point in a 210 dimensions subject space. In order to extract some information from the repartition of the subjects in this space, its dimensionality can be reduced by performing a second PCA (figure 6). From the results of this PCA, it is then possible to quantify the impact of the socio-economic factors on the semantic variation, by

Semantic Changes in Apparent Time

identifying the component associated with each of the factors. Figure 6 gives the percentage of variance explained by the first 50 components after the PCA is performed on the subject space. The second component discriminates the two groups of subjects ($t(61) = 2.02, p < 0.05$). The 13% of the variance this component accounts for can thus be attributed to the age of the speakers. We observed that gender and number of French pages read per day has little influence on the semantic variation. Subjects are discriminated according to their gender by the 17th component ($t(61) = 1.95, p = 0.06$), which explains 1.23% of the variance, and the number of pages read per day is correlated with the 42nd component ($r(61) = 0.45, p < 0.005$), which explains 0.14% of the variance.

- (6) Percentage of variance explained by the first 50 components after a PCA on the subject space.



3. Discussion

The method we have exposed in the previous sections allows us to measure semantic variation among speakers. We show a correlation between this variation and the age of the speakers which is, under the apparent time hypothesis, the synchronic manifestation of a change in progress. While most of the literature on semantic change deals with changes of metaphorical or metonymic nature, the change we observe is rather a change in the internal relationship between the lexemes of the semantic field of lodging in French. This change is more closely related to the change described by Trier (1931), who studied (from corpora) the reorganization of the semantic field of knowledge in Middle High German during the 13th century. In 1200, this semantic field was organized around three lexemes: *Kunst* ‘courtly, chivalric attainments’, *List* ‘non-courtly attainments’ and *Wisheit* ‘human wisdom in all its respects, theological and mundane’ (English glosses are from Traugott & Dasher (2002)). This organization reflected the feudal structure of the German society at this time. One century later, the society was no longer feudal and the semantic field of knowledge had been reorganized in consequence: *List* had moved out of the field and acquired its modern meaning ‘cunning, trick’ while *Wizzen* had moved in the field. Nevertheless, it was not a mere substitution. *Wisheit* had come specialized in religious knowledge, *Kunst* in artistic knowledge, while *Wizzen* covered technical knowledge. Figure 7 is schematic representation of the change.

- (7) Reorganization of the semantic field of knowledge in Middle High German between 1200 and 1300 (from Lehrer, 1985).

Wisheit	
Kunst	List

1200

Wisheit	Kunst	Wizzen
---------	-------	--------

1300

While Trier (1931) analyzed a completed semantic change from corpora, the method we proposed allows observing such an intra-semantic field change during its realization.

This method allows us to quantify the influence of age on semantic variation at 13%. This indicates that factors other than speakers' age determine this variation. The other data we gathered about subjects (i.e., gender and the number of pages read per day), seem to account respectively only for 1.23% and 0.14% of the variation. But given the high number of components (210), random noise can be expected to produce similar results. Thus, we cannot conclude the influence of both factors on the semantic variation. On the other hand, many components, and thus a large part of the variation, remain uninterpreted, in particular the first and the third components which account respectively for 25% and 11% of the variation.

The correlation between semantic variation and age may have another origin other than that of ongoing semantic change. The apparent time hypothesis has never been verified for semantics, and thus we cannot exclude that speakers modify their semantic structure of the semantic field as they get older, such that the younger group would have in 30 years the structure observed today for the older group. Nevertheless, it is not clear why such an age-grading phenomenon would occur. Moreover, the high frequency variations of the lexemes used in our study (Table 1) are clues of a change. It seems then more likely that the semantic variation observed in our study reflects a change in progress rather than an age-grading phenomenon. Ultimately, this would be confirmed by real time studies.

4. Conclusion

Variationist sociolinguistics investigates the relationships between linguistic variation and the many factors that structure a population and relates this linguistic variation with ongoing linguistic change. Yet, sociolinguistics has so far focused almost exclusively on sound changes. In particular, semantic changes and sociolinguistics have remained two disconnected domains. Postulating that one possible reason is the difficulty to observe and measure the semantic variation in the population, this paper has introduced a way to fill this methodological gap. We have applied this method to the semantic field of lodging in French and showed that the variation in the semantic representations of this semantic field mirrors an ongoing internal reorganization, i.e., a semantic change. This study opens new perspectives into the study of semantic changes, which can now be studied through the prism of variationist sociolinguistics and thus benefit from its whole theoretical framework.

Semantic Changes in Apparent Time

References

- Bailey, Guy, Tom Wikle, Jan Tillery and Lori Sand. 1991. The Apparent Time Construct. *Language Variation and Change* 3: 241-164.
- Labov, William. 1963. The Social Motivation of a Sound Change. *Word*, 19, 273–309.
- Labov, William. 2001. *Principles of Linguistic change. Volume II: Social Factors*. Oxford: Blackwell.
- Lehrer, Adrienne. 1985. The Influence of Semantic Fields on Semantic Change. In: Fisiak, J. (ed), *Historical semantics. Historical word-formation*. Berlin: Mouton. 283 - 296.
- Magué, Jean-Philippe. 2005. Changements Sémantiques et Cognition: Différentes Méthodes pour Différentes Échelles Temporelles. Ph.D. diss., Université Lyon 2.
- Nerlich, Brigitte. 1992. *Semantics Theories in Europe 1830–1930. From Etymology to Contextuality*. Amsterdam: John Benjamins Publishing Compagny.
- Parrott, Jeffrey. 2002. Dialect Death and Morpho-Syntactic Change: Smith Island Weak Expletive *it*. In *Papers from NWAV 30*, eds. D.E. Johnson & T. Sanchez. Philadelphia: U. Penn Working Papers in Linguistics.
- Romney, A. K., Moore, C. C., Batchelder, W. H., & Hsia, T. L. 2000. Statistical Methods for Characterizing Similarities and Differences Between Semantic Structures. *Proceedings of the National Academy of Sciences*, 97(1), 518–523.
- Sweetser, Eve. 1990. *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press.
- Traugott, Elizabeth, & Dasher, Richard. 2002. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.
- Trier, Jost. 1931. *Der Deutsche Wortschatz im Sinnbezirk des Verstandes*. Heidelberg: Winter.
- Ullmann, Stephen. 1962. *Semantics. An Introduction to the Science of Meaning*. Oxford: Blackwell.
- Viberg, Ake. 1983. The Verbs of Perception: A Typological Study. *Linguistics*, 21, 123–162.

Jean-Philippe Magué
ENS de Lyon
15 parvis René Descartes
BP 7000
69342 Lyon cedex 07
France

jean-philippe.mague@ens-lyon.fr