HAL
open science

A Model for Matching Semantic Maps between Languages (French/ English, English / French)

Sabine Ploux, Hyungsuk Ji

To cite this version:

Sabine Ploux, Hyungsuk Ji. A Model for Matching Semantic Maps between Languages (French/ English, English / French). Computational Linguistics, 2003, 29 (2), pp.155-178. hal-00933321

HAL Id: hal-00933321

https://hal.science/hal-00933321v1

Submitted on 20 Jan 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Model for Matching Semantic Maps between Languages (French/English, English/French)

Sabine Ploux*
Institut des Sciences Cognitives

Hyungsuk Ji*
Institut des Sciences Cognitives

This article describes a spatial model for matching semantic values between two languages, French and English. Based on semantic similarity links, the model constructs a map that represents a word in the source language. Then the algorithm projects the map values onto a space in the target language. The new space abides by the semantic similarity links specific to the second language. Then the two maps are projected onto the same plane in order to detect overlapping values. For instructional purposes, the different steps are presented here using a few examples. The entire set of results is available at the following address: http://dico.isc.cnrs.fr.

1. Goals

This article presents a spatial model that projects the semantic space of a source language word onto a semantic space in the chosen target language. Although the study presented in this article can be described from various angles, we place it within the framework of artifactual simulations of the translation process, and more specifically, access to the target language's lexicon. The model is described as a construction process designed to reproduce cognitive functions and their extensions. Future research will include the study of the psycholinguistic validity of such a spatial representation. Now let us briefly describe the scientific basis of the study.

- Three major areas are generally distinguished in the study of the translation process (see Vinay and Darbelnet [1996]), the lexicon (or the study of notions), sentence generation (putting words together), and the message (which brings communicative factors into play). The first area involves choosing the right word, which is usually left up to the intuition and expertise of the translator. Our model deals with accessing the lexicon of the target language starting from a notion in the source language. The utility of this research lies in the fact that different languages break down reality in different ways.
- Although the translation process has been mastered by a number of experts, it is usually still dependent upon the utilization of tools like dictionaries. The model proposed here relies on semantic maps and offers an alternative method based on the concepts of lexical access and lexical neighborhood.
- The work by Anderson (1983) and Collins and Loftus (1975) on the organization of the lexicon is based on priming and the automatic

* UMR 5015 CNRS-Université Lyon I, 67 bd Pinel, F-69 675 Bron Cedex. E-mail:{ploux,ji}@isc.cnrs.fr.

© 2003 Association for Computational Linguistics

spreading of activation to the prime's neighboring concepts. As an alternative to these local semantic networks, Masson (1995) proposed a connectionist model that takes into account the subjects' reaction time during priming experiments (the correspondence is based on the assumption that semantic or phonologic proximity and ease of access are correlated). Rouibah, Ploux, and Ji (2001) showed that experimental data on interactions between phonology and semantics could be simulated by distances on lexical maps. One advantage of this proposal is that experimental and artifactual findings converge; another is its ability to describe a real lexicon. Although the relevance of our model to the representation of the mental lexicon will not be discussed in this article (attempts to gain insight into this correlation are currently underway in other studies), this point is not unrelated to the suitability of our approach to modeling translation as a cognitive function.

2. Description of the Model

No two lexicons are related by a one-to-one correspondence (Abplanalp 1998). In other words, the way words are used to refer to extralinguistic reality varies across languages. Some examples of this are cross-language differences in color naming and, borrowing Chuquet and Paillard's (1989) English-French examples, differences like:

- room: *pièce, chambre, bureau*

(or in an abstract domain)

- *esprit*: mind, spirit, wit

Certain authors (Abplanalp 1998) insist how impossible it is to translate at the word level and propose recourse to the conceptual level as a theoretical alternative. Concepts are thought to depend on human cognitive abilities that are general and shared by all. Although the correspondence between words and concepts remains a controversial topic of study (Reboul 2000), the concept/word opposition is nevertheless relevant to any model of translation, even an artifactual one like ours. As we shall see, even when heeding the specific organization and breakdown of each individual language, the matching operation does not take place at the word level but at the substrate level (defined below), where the set of meanings of each word "cuts out" a form.

First, we will present the model we devised to describe the organization of languages. Then we will explain the source-to-target spreading method used.

2.1 A Model Based on Semantic Similarity

The model was initially developed on the basis of a semantic similarity: synonymy. Note, however, that the data and the model are independent, so this same framework can be used to organize other types of similarity (contextual, phonological [Rouibah, Ploux, and Ji 2001], etc.). Other authors also organize the lexicon or other kinds of knowledge on the basis of similarity. For example, in Edelman's (1998) spatial model of internal representations of the world's objects, spatial proximity reflects object similarity. WordNet (Fellbaum 1998) and EuroWordNet (Vossen 1998) organize the lexicon conceptually as a network of terms, each of which is associated with a partition into

Synsets (a **Synset** being a small group of synonyms that label a concept). Our model differs from Edelman's in that it deals with lexical semantics, not perceived objects. It also differs from Miller's (1990) approach, in three respects:

- the grain of the semantic units
- the lexical structure generation mode
- the resulting geometry and organization

Most models[1] use separate units to represent words or concepts (symbols, points in a space, nodes on a graph, etc.). Relationships between units are expressed as proximity links (in spatial models) or as arcs between nodes (in networks). Our model is spatial, but it differs from local models in that each term is represented by a region in the space, part of which it shares with other terms. This region is constructed automatically according to lexical similarity links (such as those given by a synonym dictionary). It is not the result of supervised learning, nor is it a manual, ontological description of how the lexicon is organized. The next section will break the semantic-space construction process into steps in presenting the initial data, the granular approach, and the resulting organization.

2.2 Method

2.2.1 Initial Data. Three databases were used: two synonym databases (one containing French terms and one containing English terms) and a translation database (French-English, English-French) that maps each term to similar words in the other language. The links between an entry and the terms that follow it were not chosen "by hand." The data were taken mainly from published dictionaries and thesauruses.[2] It is updated and supplemented regularly by the addition of new links between words (synonymy or translation links). The method used to generate the French synonym database (described in detail in Ploux (1997) was applied again to generate the English and translation databases. The first step required creating an intermediate database containing the set of all links attested in available work in lexicography. In this preliminary database, a term was deemed similar to another term if at least one lexicographer had established the link. The final database was obtained through symmetrization of the links produced in the first step. While maintaining the shifts in meaning that occur when there is nontransitivity and that, as we shall see, are essential for developing the model, we created new links to symmetrize any initially one-directional ones.[3] Table 1 gives a typical example of the structure of the initial data. Table 2 gives a global evaluation of the number of entries and links in the lexical databases. Note that we are not attempting here to define the term *synonymy*. We rely on lexicographic publications, which as Edmonds and Hirst (2002) remarked, "have always treated synonymy

1 Masson's (1995) model assigns each concept a basin of attraction in a multidimensional space of activation. This framework authorizes a certain form of internal variability for the set of patterns corresponding to a concept. Nevertheless the basins are disjoint and do not overlap as do the nodes in local semantic networks. Furthermore, this model, built essentially for the purposes of validating hypotheses and comparing psycholinguistic results, is applicable only to a highly limited vocabulary and is therefore a poor representative of the natural lexicon.

2 For the French database, we used files compiled by the National Institute for the French Language (INALF: Institut National de la Langue Française) from seven different 19th- and 20th-century synonym dictionaries; for the English and translation databases, we used files obtained from the French company MEMODATA.

3 Note that symmetrization does not make the semantic spaces of the two terms equivalent.

Table 1
Format of data files.

Headword:	$\text{Similar}_1, \text{Similar}_2, \text{Similar}_3, \ldots$
insensible: (extracted from the similar English word database)	apathetic, benumbed, callous, comatose, impassive, imperceptible, impercipient, indiscernible, insensitive, pachydermatous, senseless, thick-skinned, unaffected, unaware, unconscious, undetectable, unfeeling, unsympathetic
insensible: (extracted from the similar French word database)	adamantin, anesthésié, apathique, aride, assoupi, blasé, calleux, calme, cruel, de marbre, desséché, dur, détaché, endormi, endurci, engourdi, flegmatique, . . .
insensible: (extracted from the French-English translation database)	imperceptible, insensitive, numb, unfeeling

Table 2
Number of entries and links in the lexical databases.

	Number of entries	Mean number of synonyms per entry	Mean number of terms proposed by the translation database
French	54,690	7.5	2.3
English	148,247	6.8	1.9

as near-synonymy."[4] However, having more flexible semantic links does not detract from the accuracy of the model. No other operations are carried out on the data sets before application of the model.

2.2.2 Semantic Units. To represent variations in a word's meaning, each word is associated with a spatial form (or space) (Ploux 1997; Ploux amd Victorri 1998). The points in the space are finer units of meaning than the word itself. In our computational simulation, the points are represented by cliques. A **clique** is a set of terms related to each other by synonymy.[5] The conjunction of all terms in the same clique crystallizes and constrains the meaning given to the word. These cliques thus constitute good candidates for generating the substrate upon which the form will take shape. The presentation of the results and the features of the model will be illustrated using examples from the headword *good* for English and from the headword *insensible* for French. The Appendix provides the full results, as well as the definition of the word *insensible* taken from a French dictionary. These examples are illustrative of the main characteristics of the entire data set.

4 Moreover, for the two languages under study here, there are notable differences in how lexicographers understand and use the concept of synonymy. Synonymy relations in French dictionaries, for example, are not always symmetrical and are rarely transitive. What is more, the links have a broader scope. For instance, the words *abri* (shelter) and even *masure* (shed) are given as synonyms of *maison* (house). To make the databases homogeneous during the matching operation, a new version of the English database was supplemented with certain hypernym links often given as synonyms in French dictionaries. The software offers the user the opportunity to see the output obtained using the two versions of the English database, displayed under the headings *standard search* and *enriched search*.

5 By definition, this is a maximal, connected component of the synonym graph. Words are placed at the nodes of the graph, and arcs between two nodes represent a synonymy link.

The synonym list contains a heterogeneous set of scrambled terms:

- For the French headword *insensible*, some of the terms represent a moral value (*dur, sans-coeur, . . .*), others a physical value (*inerte, engourdi, . . .*), and still others a perceptual value (*imperceptible, inapparent . . .*).
- The headword *good* includes many similar terms. As a first approximation, only the most representative are given here. Some of the terms represent a generic value (*right, sound, . . .*), others refer to a capability (*able, . . .*) or have an affect-related value (*benevolent, . . .*), while still others represent a quality of taste (*tasty, . . .*).

The clique list contains the cliques generated from this set of terms. Cliques represent rather precise units of meaning.[6]

- Here are some examples of cliques representing the moral value of the French headword *insensible*:

  20: cruel, dur, impitoyable, implacable, inexorable, inflexible, inhumain, insensible
  21: cruel, dur, impitoyable, implacable, inexorable, inflexible, insensible, sévère
  22: cruel, dur, implacable, inflexible, inhumain, insensible, rigide
  23: cruel, dur, implacable, inflexible, insensible, rigide, sévère

  Some examples of cliques representing the physical value:

  2: anesthésié, insensible
  50: endormi, engourdi, inerte, insensible
  51: engourdi, froid, inerte, insensible
  52: engourdi, immobile, inerte, insensible, paralysé

  And some examples of cliques representing the perceptual value:

  69: imperceptible, inapparent, insensible, invisible
  70: imperceptible, indiscernable, insaisissable, insensible, invisible
  71: imperceptible, indiscernable, insensible, léger

- Here are some examples of cliques representing the more prominent senses of the English headword *good*:

  84: dependable, good, reliable, safe, secure
  87: dependable, good, reliable, solid, sound
  102: fair, good, honest, honourable, just, right, upright

  Some examples of cliques representing a more specific meaning of aptitude or ability:

  6: able, adequate, capable, competent, effective, good
  7: able, adroit, clever, dexterous, expert, good, skilful
  8: able, capable, clever, expert, good, skilful

6 The cliques are numbered here in the order in which the results are presented on the Web site (alphabetical order).

And some examples of cliques with affect-related values:

111: friendly, gentle, good, kind, kindly, nice, sweet
112: friendly, good, gracious, kind, kindly, nice, sweet
113: friendly, good, helpful, kind

Note that a given term may belong to several cliques (this characteristic is due to the nontransitivity of the relation). It appears in each clique with a precise meaning that is constrained by the presence of its neighbors.

- For example, the following cliques have terms in common; the first has a stronger moral value than the second:

  15: calme, flegmatique, froid, impassible, imperturbable, insensible
  18: calme, immobile, inanimé, insensible

- In the same manner, there are shared terms in the next two cliques of *good*, the first related to taste, the second to personal qualities:

  80: delectable, delicious, good, lovely, savoury, scrumptious, tasty
  82: delicious, good, lovely, nice, pleasant

This last point brings us to the study of semantic variations. The following clique path, in which each clique shares at least one term with the next, moves in a relatively continuous way from one value to another.

- Transition from a moral value to a physical value:

  21: cruel, dur, impitoyable, implacable, inexorable, inflexible, insensible, sévère
  34: dur, froid, impitoyable, implacable, insensible, sévère
  35: dur, froid, inaccessible, indifférent, insensible
  39: dur, impassible, indifférent, insensible, stoïque
  15: calme, flegmatique, froid, impassible, imperturbable, insensible
  16: calme, froid, inanimé, insensible
  63: froid, inanimé, inerte, insensible
  83: inanimé, inerte, insensible, mort

  The continuity between the moral and physical values has its counterpart in their usage. For example, one can use the term *engourdi* in French to qualify the disposition of a person who exhibits little moral reactivity, as in:

  > Il allait comme dans un songe, l'esprit engourdi, paralysé, sans chagrin vibrant, saisi par une sorte d'engourdissement moral qui l'empêchait de souffrir, éprouvant même un allégement qu'augmentaient les exhalaisons tièdes épandues dans la nuit.[7] (Maupassant 1881, page 350)

  Moreover, as we shall see later, this type of continuous link between two values, which acts as a metaphor here, is expressed more explicitly in the English example below.

7 Although the term *engourdi* is not specifically translated, to help the reader understand this fine shade of meaning, here is a translation of the above passage (Maupassant 2002): *He walked as if he were in a dream; his thoughts were paralyzed, although he felt no great grief, for he was in a state of mental torpor that prevented him from suffering, and he even felt a sense of relief which was increased by the mildness of the night.*

Table 3
Evaluation of clique granularity.

Entry	Number of cliques containing entry	Number of distinctions found in published dictionaries	Number of Synsets in WordNet
défendre	44	9–13	—
distraction	39	3–10	—
fou	319	10–23	—
jouer	95	15–46	—
maison	123	9–42	—
vert	50	9	—
blue	54	22–34	26
house	82	11–24	11
good	193	24–50	30
look	104	18–73	13
mind	87	41–68	13
play	240	77–84	47

- Transition from a taste-related value to an affective value:

  80: delectable, delicious, good, lovely, savoury, scrumptious, tasty
  78: delectable, delicious, excellent, exquisite, good, lovely, scrumptious
  77: delectable, delicious, enjoyable, good, pleasant
  79: delectable, delicious, good, lovely, pleasant
  82: delicious, good, lovely, nice, pleasant
  114: friendly, good, kind, kindly, nice, pleasant, sweet
  111: friendly, gentle, good, kind, kindly, nice, sweet

By contrast, for the French headword *insensible*, there is greater discontinuity between the perceptual value and the others. At the present stage of our project, clique lists are in alphabetical order, and the underlying semantic topology has not yet been built. The geometric model we are now going to present achieves this step. Table 3 contains an evaluation of the granularity generated by the cliques.

2.2.3 Output Geometry and Organization. To construct the semantic space, a conventional correspondence factorial analysis[8] (Benzécri 1992) was conducted between the cliques and the synonyms. For each entry, the initial matrix M_{ij} contains *nc* rows (where *nc* stands for the number of cliques) and *ns* columns (where *ns* stands for the number of terms). It is defined by the formula $M_{ij} = 1$ if clique i contains term j, and 0 if not. The results showed that the χ_2 distances[9] calculated using this method furnish a coherent representation of semantic variations. Table 4 presents the configurations

8 Correspondence analysis is a factor analysis method that uses categorical variables (that is, noncontinuous or discretized ones).

9
$$d(c_i, c_k) = \sum_{j=1}^{n} \frac{x}{x_{.j}} \left(\frac{x_{ij}}{x_{i.}} - \frac{x_{kj}}{x_{k.}} \right)^2$$
where c_i and c_k are two cliques, n is the number of synonymous terms, $x_{i.}$ the number of terms in c_i (respectively c_k), $x_{.j}$ the frequency of term t_j and x the sum of the frequencies of all terms (or the total number of terms in all cliques).

Table 4
Comparison of Euclidean distance and χ_2 distance on the principal plane for the above cliques.

	Euclidian distance	χ_2 **distance**
$d(c_{23}, c_{12})$	1.7855	1.7357
$d(c_{17}, c_{23})$	0.6306	0.0170
$d(c_{12}, c_{17})$	1.2382	1.17213

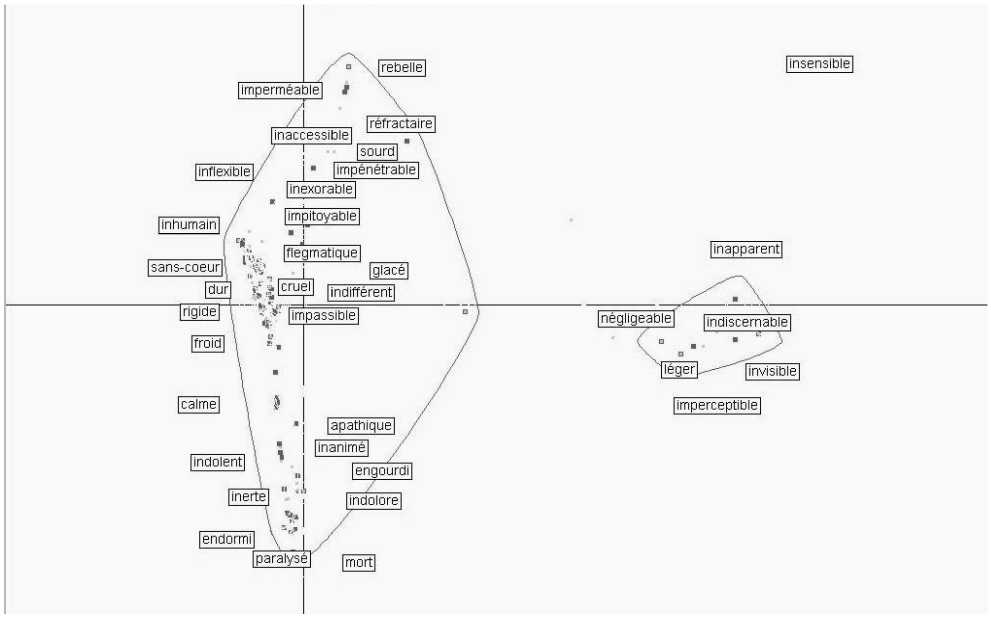

Figure 1
Two-cluster semantic space for the French headword *insensible*.

on the principal plane for the Euclidean distance and the χ_2 distance, reduced to the same proportion. The headword *fast* has many cliques, including

- c12: express, fast, quick, rapid, swift
- c17: fast, fastened, fixed, secure
- c23: fast, firm, lasting, stable, tight

The values obtained using the χ_2 distance are more suited to semantic categorization than those obtained using Euclidean distance; cliques representing the same class are closer together (even if they do not share a larger number of terms) than ones representing different meanings.

The dimension of the geometric space is equal to the smaller of the two numbers, *ns* or *nc*. To show the results visually, the projections onto the principal axes are presented in Figures 1 and 2. (The horizontal axis in the figures is the best representative of the form delineated by the cluster of points such that the distances between the points are maintained to the optimal degree; the vertical axis, perpendicular to the first, is the second best representative, and so on.) Cliques are represented by points, and each

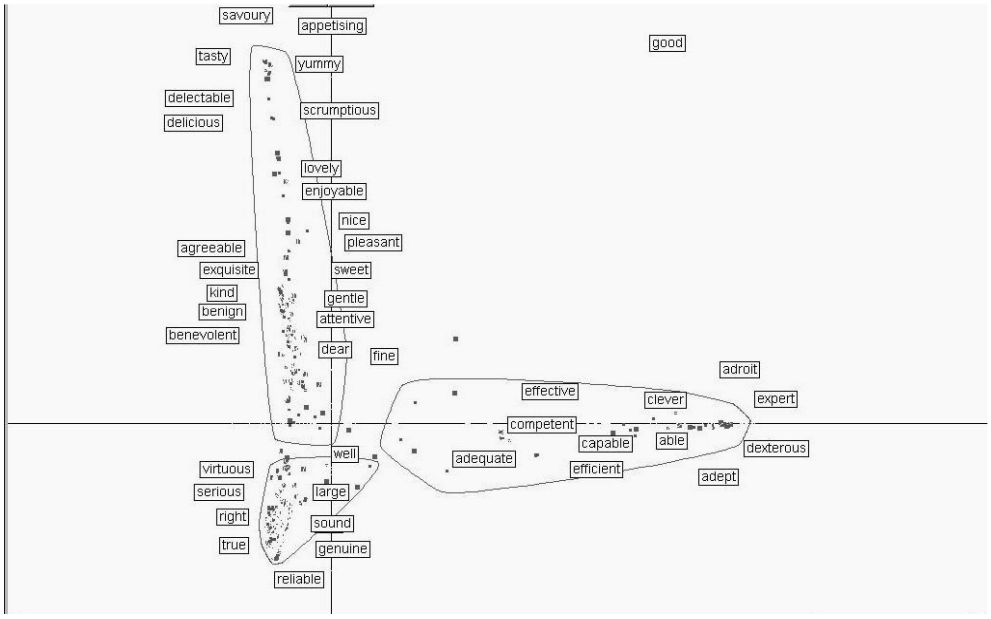

Figure 2
Three-cluster semantic space for the English headword *good*.

term by the region in the space delineated by the set of cliques that contains it.[10] Using the examples again, let us review the main characteristics of the resulting organization. The same type of organization is found in all cases.

2.2.4 Distinguishing Semantic Values. The model plots the different values on the map. Distinct notions are clearly separate, and gradual variations are maintained.

- In the *insensible* example (Figure 1), we can see two clusters as a first approximation, one smaller cluster labeled by the terms *imperceptible, inapparent, indiscernable, négligeable*, etc., and representing the perceptual value of the word, and one larger cluster containing the moral and physical values. In the center of the second cluster, we find the terms *dur, inhumain, sans-coeur, cruel, etc.*, which are prototypes of the word's moral value. Two branches come out of this center, one that qualifies a more specific value (*réfractaire, rebelle, impénétrable, etc.*), and one that leads to the physical value.[11]
- In the *good* example (Figure 2), the cliques and terms are plotted on the map in accordance with the proximities of the values and their links. On the principal plane, the cluster of points extends in two directions: the first axis represents the capability value, and the second the affective value. The affective value gradually turns into a taste-related value (*tasty, . . .*). These two main directions are interconnected by the generic value (*right, true, . . .*) located near the origin.

10 An appropriate algorithm generates the envelope (i.e., the set of cliques that contains the term) for a given term.
11 In all figures in this article, the principal classes are outlined. (A publication about the principles of this automatic classification model is now in preparation; only the results are given here.)

Table 5
Some examples of spatial interconnections between semantic values.

Entry	Value at the origin (labeled by a prototype)	Examples of off-centered values (labeled by prototypes)
défendre	protéger	1. excuser 2. interdire, . . .
maison	domicile	1. commerce 2. lignée, . . .
insensible	sans-coeur	1. imperceptible, 2. engourdi, . . .
home	abode	1. family, 2. interior, . . .
good	right	1. able, 2. delicious, . . .

2.2.5 Spatially Interconnecting the Values. Table 5 shows the hierarchy of the spatial organization. The middle column contains the generic values (when they exist) that interconnect the different meanings of the word. Highly specific values are far from the origin. This organization follows directly from the calculation of the profile matrix, which assigns more weight to infrequent terms and to cliques containing few elements.

2.3 Matching

As stated above, the breakdown and overlapping of the lexicon varies from one language to the next. However, several studies (Illes and Francis 1999; Ikeda 1998) have found evidence that the two languages of a bilingual person access a common semantic system. To handle the problem of lexical differences in our translation model, connections link semantic units rather than words. Because they are finer-grained than words, semantic units are assumed to be less sensitive to the way a given language "cuts up" the world, and as such, they are better candidates for achieving a closer fit between the two languages. For a given set of cliques in the source language, the model constructs the set of cliques to be used for the translation. The two spaces (one associated with each set of cliques) are then projected onto a map that maintains the matches. The example of *insensible* is a good representative of the various patterns that can appear. It has two very different, nearly homonymic semantic values, as well as some other values whose meanings overlap considerably. For this reason, we present the results for the matching operation using this example. The four steps in this construction process are described below.

> *Step 1. Constructing the source semantic space*. In order to build a semantic space in the target language associated with a term in the source language, the system starts by generating the set of all cliques containing the requested word. This step is identical to the one described in Section 2.2.2.
>
> *Step 2. Searching for relevant target language units for translation*. For all initial terms similar to the input word, the translation database furnishes the corresponding terms in the target language. Some of these terms are relevant to the initial generic meaning; others are clearly far removed from that meaning. For example, the synonyms *timide* and *léger* of the term *insensible* can be translated respectively as (. . . , *shy*, . . .) for *timide* and (. . . , *airy*, . . .) for *léger*, neither of which is useful in generating this headword's target semantic space. To find the relevant senses, the model compares the source language cliques to the cliques generated from the set of terms proposed by the translation database. Target clique relevance

is calculated as follows: Let S be a clique in the source language composed of the terms $(t_i^s)_{i=1\cdots n_s}$, and let C be clique $(t_j^c)_{i=j\cdots n_C}$ in the target language. The model evaluates the relevance of the translation based on the rank of matrix M_{SC}, composed of zeroes and ones, calculated using the formula $M_{SC}[i][j] = 1$, if t_j^c translates t_i^s, and $M_{SC}[i][j] = 0$, otherwise. The rank defines a spreading parameter (in the model, a rank of zero means that the two cliques are unrelated and the target clique represents an out-of-range meaning in the translation operation; a rank of three or more represents a highly cohesive semantic link).

If this last constraint is imposed on all cliques, the model will output a relatively small number of terms belonging to the target's semantic field.[12]

Step 3. Constructing the source-point/target-point geometry. The factorial analysis algorithm (presented in Section 2.2.3) is followed to determine the correspondences between the source cliques and the target cliques that were retained in step 2, because they are relevant to at least one clique in the source language. The correspondences are determined by taking the product of the following matrices:

$$M_{tr} = M_{cs}^{S} * T_{sc} * M_{cs}'^{C}$$

where M_{cs}^{S} is the source-clique/source-term matrix defined as in monolingual processing (see Section 2), T_{sc} is the matrix that defines the translation between the source terms and the target terms ($T_{sc}[i][j] = 1$ if and only if term j translates term i in the initial database), and $M_{cs}'^{C}$ is the transposed target-clique/target-term matrix.

For a subset of the French cliques of *insensible*, the closest three English cliques are given below for each French clique, along with a table of the corresponding distances calculated on the principal plane (Table 6). The maps reproduced in Figures 3–5 summarize the resulting distances for the headword *insensible*.

- cf_{28}: cruel, dur, féroce, impitoyable, implacable, inexorable, inhumain, insensible

  ce_{67}: cruel, ferocious, fierce, ruthless, savage
  ce_{84}: cruel, inhuman, merciless, pitiless, ruthless, savage
  ce_{28}: bitter, cruel, fierce, ruthless, savage

- cf_{40}: dur, indifférent, inhumain, insensible, sans-coeur

  ce_{36}: callous, hard, hardened
  ce_{33}: callous, cruel, hard, hard-hearted, heartless
  ce_{92}: difficult, hard, tough

- cf_{78}: imperméable, insensible, rebelle, réfractaire, sourd

  ce_{148}: insensitive, unmoved

12 Our software proposes two types of lexical access. The first is more restrictive and sets the rank at three or more; the second supplies a broader vocabulary and sets the rank at two or more.

Table 6
Distances between French and English cliques on the principal plane. (For all cliques, the distances ranged between 0.0035 and 4.0183.)

	cf_{28}	cf_{40}	cf_{50}	cf_{51}	cf_{68}	cf_{71}	cf_{78}
ce_{17}	2.6483	2.1008	0.3486	0.1542	3.2331	3.6239	1.3379
ce_{28}	0.2567	0.3009	2.0498	2.3846	3.1128	3.5984	1.0726
ce_{33}	0.5407	0.0543	1.7672	2.1031	3.0130	3.5002	0.7897
ce_{36}	0.5625	0.0439	1.7467	2.0831	2.9967	3.4839	0.7670
ce_{40}	1.3223	0.7680	1.0502	1.3933	2.6981	3.1721	0.0907
ce_{67}	0.1151	0.4404	2.1912	2.5256	3.1666	3.6503	1.2135
ce_{84}	0.2228	0.3323	2.0854	2.4206	3.1108	3.5959	1.1056
ce_{87}	2.3812	1.8385	0.0813	0.2630	3.2189	3.6326	1.0886
ce_{89}	2.5708	2.0272	0.2656	0.0944	3.2805	3.6799	1.2730
ce_{92}	0.5461	0.0633	1.7609	2.0965	3.0209	3.5081	0.7856
ce_{97}	2.5336	1.9861	0.2365	0.1846	3.1868	3.5866	1.2240
ce_{98}	2.3637	1.8176	0.0676	0.3000	3.1508	3.5641	1.0598
ce_{100}	2.2895	1.7453	0.0164	0.3593	3.1568	3.5761	0.9926
ce_{112}	3.0294	2.7339	2.8600	3.0620	0.3066	0.7473	2.5137
ce_{114}	3.6318	3.3733	3.5037	3.6875	0.4153	0.0788	3.1821
ce_{129}	2.9849	2.6955	2.8534	3.0606	0.3108	0.7718	2.4883
ce_{130}	1.3400	0.7892	0.9756	1.3163	2.8715	3.3439	0.0867
ce_{137}	3.1895	2.9208	3.1005	3.3031	0.0685	0.5290	2.7350
ce_{148}	1.3696	0.8147	0.9837	1.3268	2.7454	3.2172	0.0551
ce_{149}	3.6058	3.3440	3.4678	3.6514	0.3879	0.1143	3.1488
ce_{152}	3.6335	3.3752	3.5061	3.6899	0.4171	0.0765	3.1843

ce_{130}: impassive, indifferent, phlegmatic, stoical
ce_{40}: callous, impassive, insensible, unfeeling

- cf_{50}: engourdi, froid, inerte, insensible

  ce_{100}: dull, inanimate, inert, lifeless
  ce_{98}: dull, expressionless
  ce_{87}: dead, inanimate, inert, lifeless

- cf_{51}: engourdi, immobile, inerte, insensible, paralysé

  ce_{89}: dead, numb, paralytic
  ce_{17}: asleep, numb
  ce_{97}: dull, dulled

- cf_{68}: imperceptible, insensible, invisible

  ce_{137}: imperceptible, indiscernible, invisible
  ce_{112}: frivolous, indifferent, insignificant, trifling, unimportant
  ce_{129}: impalpable, imperceptible, intangible, invisible

- cf_{71}: imperceptible, insensible, insignifiant, léger

  ce_{152}: light, slight, trifling, trivial
  ce_{114}: frivolous, light, trifling, trivial
  ce_{149}: insignificant, slight, trifling, trivial, unimportant

Step 4. Defining the lexical regions. As above, for each language, a term is represented by the clique region that contains it.

The next section will use examples to illustrate the results obtained. The entire set of results is available at http://dico.isc.cnrs.fr.

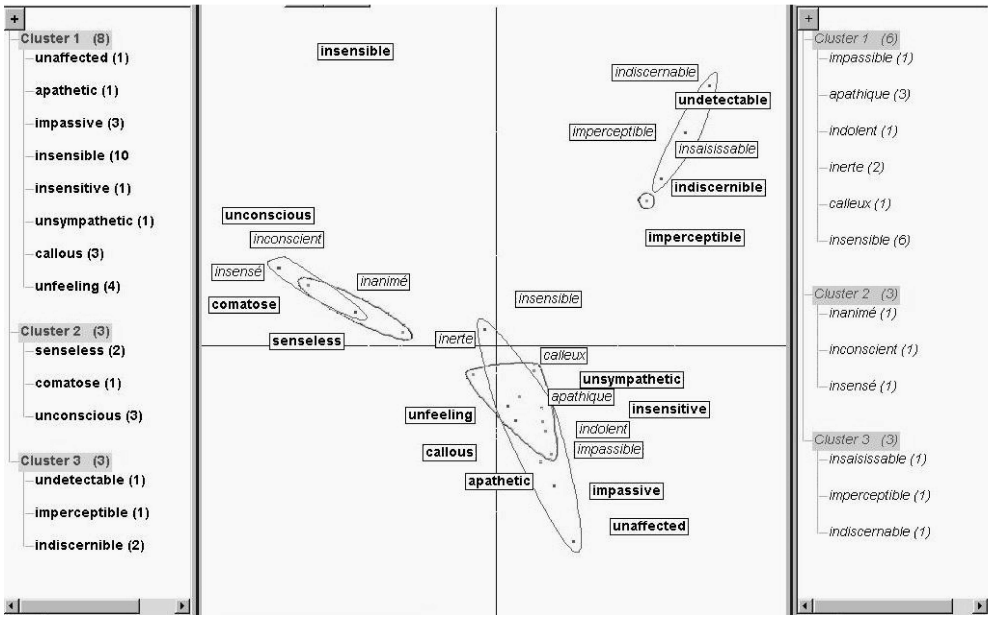

Figure 3
English-French space matching for the English headword *insensible*.

3. Results

The advantages of the model presented are (1) access to an extended lexicon and a broad semantic field and (2) coherence of the matching between the semantic values in each language. The results for *insensible* will be used again in this section to illustrate the second advantage.

3.1 Access to an Extended Semantic Field and Lexicon

The model fulfills two functions: It searches for a suitable lexicon and organizes the terms found. For each entry, the initial data provides a short list of terms representing certain prototypes of the word's translation. Table 1 lists the four English terms proposed as translations for the French word *insensible*. It can happen that certain semantic values in the source language are not represented in the translation database. For example, *insensible* has no corresponding French word in our database of English word translations. However, the model builds the appropriate values in French (Figure 3).

The model builds a much larger vocabulary that includes the initial terms from the translation database and some semantic neighbors. Table 7 presents an overall evaluation of the results.

Table 7
Assessment of lexical access spreading to the target language.

Mean number of terms supplied by the translation database from a sample of 60 terms	Mean number of terms supplied by the semantic maps of the same sample
14.1	92.9

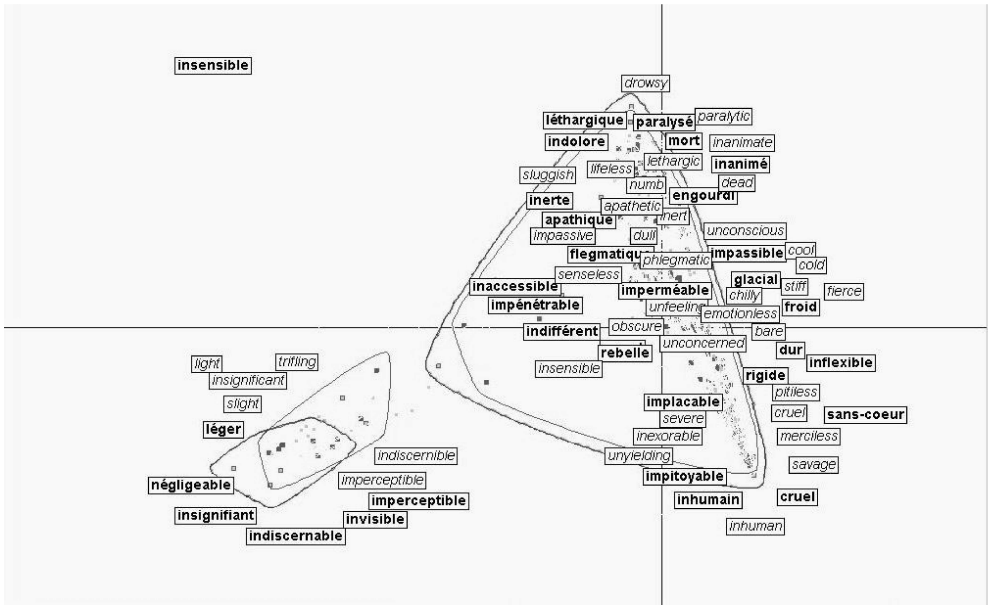

Figure 4
Two-cluster separation of the French and English spaces for the French headword *insensible*.

3.2 Coherence of the Semantic Matching

The final step in the model consists of establishing a correspondence between the semantic values of the cliques and the terms in the two languages. By application of the above algorithm, the cliques and terms of the two languages are plotted on the same map. This map thus provides a summary of the semantic proximities in each language. In order to demonstrate the coherence of the semantic-value matching after projection onto the target language, the clusters obtained from the French and English cliques for the term *insensible* are superimposed on one another. Figures 4 and 5 present the division of the output into two and four clusters. (The French clusters in these figures are marked by a darker line and set in a darker typeface than the English ones.) As in the two-cluster semantic space for the French word *insensible*, Figure 4 separates the perceptual value from the other values.

The three-cluster separation then differentiates the physical-moral value from the moral value. Figure 5 shows the division within the physical-moral value between what is more specifically physical and what pertains to emotional insensitivity (*emotionless, réfractaire, etc.*) or to the inability to discern that sensitivity (*impénétrable, etc.*).

Note that although all values initially present in the monolingual space are represented, a reorganization process still takes place during pairing with the target language. In French, the terms (*réfractaire, inacessible, . . .*) were separated from the terms (*inerte, engourdi, . . .*) by the group made up of the terms (*dur, sans-coeur, . . .*), but now they are located close to the center. This layout probably results from (1) the effect of the greater number of terms like (*inert, numb, sluggish, chilly, . . .*), which, in English, unlike in French, encompass emotional and physical insensitivity and therefore bring these two values closer together on the map, and (2) the prototypical, central nature of this value in English, as expressed by the terms (*impassive, insensible, insensitive, . . .*).

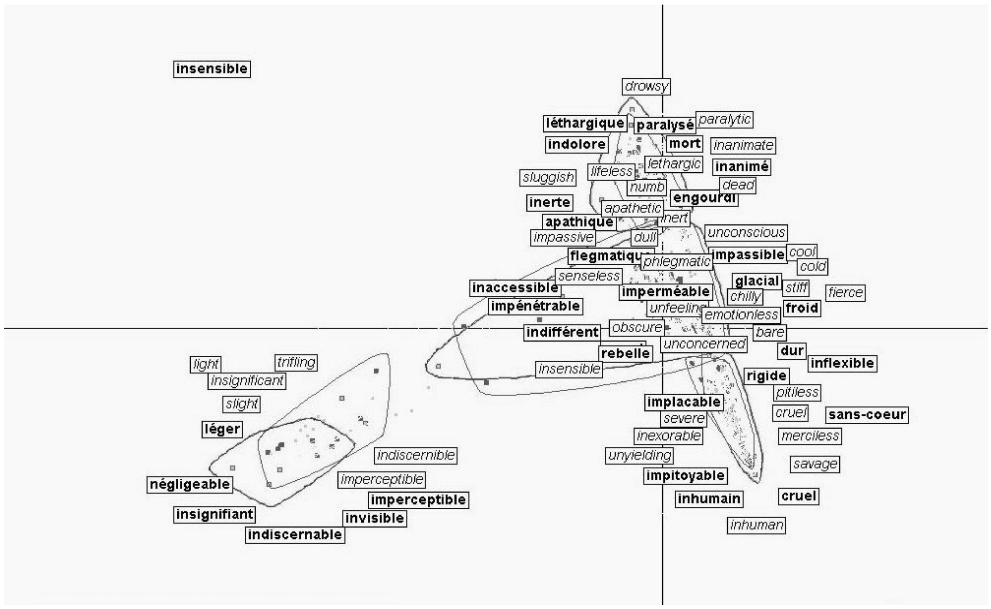

Figure 5
Four-cluster separation of the French and English spaces for the French headword *insensible*.

4. Discussion

We have presented a model for matching a semantic space in a source language and a semantic space in a target language. This model, currently built from lexical similarity relations (synonymy or near-synonymy and translations), uses several representation levels: cliques, which represent very precise units of meaning; terms, which are represented geometrically by a region in the space containing a set of cliques; and clusters, which are generated from the results of a spatialization process that singles out a term's main semantic values. (Again, this last representation level is merely mentioned in the present article; the method used to generate it and the rationale for its use in semantic classification will be described in detail in a forthcoming publication.) The matching between the French and English spaces is achieved by mapping the cliques of the two languages to each other. The model software allows a user to choose a candidate word in the target language according to its synonym neighborhood. A map showing each language's neighborhoods and separate clusters for each semantic value helps the user make the choice. This system and its interactive interface is a useful tool appreciated by researchers, translators, writers, and other users. Although this alone is enough to justify the model, it would be worthwhile to incorporate it into a more complete automatic language processing system. We are now working on enhancing the system by including context relations, and by bringing to bear a word's argument structure, qualia structure, and lexical inheritance.

Within the past 10 years, original contributions have been made in the areas of compositional semantics and lexical context assignment (see Ide and Veronis [1998] for the state of the art on word sense disambiguation). Most studies have dealt with the sentence, but some have looked at the discourse and text levels. Based on a generative framework, Pustejovski (1995) proposed a computational model that adds a representation of a word's structures (event structure, argument structure, qualia structure, and

lexical inheritance structure), along with transformation rules for combining units. In their study, Asher and Lascarides (1995) showed that lexical semantics and discourse structure may interfere with discourse structure and devised heuristics to disentangle the effects of these two interacting levels. Other authors (Foltz, Kintsch, and Landauer 1998; Kintsch 2001; Schütze 1998) have developed an approach based solely on automatic corpus analysis in which co-occurrences and their frequencies are used to generate the semantic space associated with a given word. Edmonds and Hirst (2002) proposed a model with two tiers: a fine-grained synonym tier and a coarse conceptual tier. Unlike Edmonds and Hirst's approach, which rests on an ontological model and conceptual representations, our model is capable of detecting semantic distinctions solely on the basis of similarity links. This feature is one of the model's assets, but it is also a limitation, which provides the incentive for the enhancements we are currently developing. Here is a brief preview of our ongoing projects:

- Certain words are poorly represented in terms of synonymy. This is the case for words that are essentially nonpolysemous, like *computer* or *daisy*, and thus have very few synonyms. Such entities are better delineated by an ontological, hierarchical representation and by their qualia structure than by synonymy links. Grammatical words also have few synonyms, so they too need to be represented in a formalism more suited to their own features than the one proposed in this article.
- Usage contexts or domains of application are not currently given for the different semantic values detected by the model. For example, the perceptual value of the word *insensible* is employed to modify external phenomena, whereas the moral and physical values apply to animate beings. It would thus be useful, as in a standard dictionary, to specify the different types of terms the values obtained can modify.
- Our research should help improve map drawing. At the present time, map neighborhoods rely solely on semantic criteria, which sometimes leads to the map's including terms with similar meanings but different syntactic category memberships than the initial word.

These projects should contribute to furthering research on language and automatic language processing. As stated in the article's introduction, we are also working on the cognitive relevance of our model. We have already conducted an initial study aimed at determining whether a spatial model is an appropriate way of representing the structure of the mental lexicon. Our work on this problem draws from a preliminary study (Rouibah, Ploux, and Ji 2001) which proposes a homomorphism between lexical distance (the organizing principal of our model) and reaction time (the parameter used in lexical access experiments). This idea is based on the finding that lexical distance is subject to the same effects as reaction time.

Appendix

Example of a classification, for the French term *insensible* (taken from *Le Petit Robert* version 1.2). Rough English translations are given in parentheses.
insensible:

- I Qui ne sent pas, ne ressent rien. (Not sensing, feeling nothing.)

1. Qui n'a pas de sensibilité physique. inanimé, mort. (Having no physical sensitivity. inanimate, dead.)
2. Qui n'éprouve pas les sensations habituelles, normales. (Not experiencing the usual, normal sensations) (insensible à la douleur, au froid, à la chaleur. (insensitive to pain, to cold, to heat.)
3. Qui n'a pas de sensibilité morale; qui n'a pas ou a peu d'émotions. (Having no moral sensitivity; having few if any emotions.) apathique, calme, détaché, froid, impassible, imperturbable, indifférent. cruel, dur, égoïste, endurci, impitoyable, implacable, inexorable. imperméable, indifférent. sourd. étranger, fermé, inaccessible; réfractaire. (apathetic, calm, detached, cold, impassible, imperturbable, indifferent. cruel, hard, egotistical, hardened, pitiless, implacable, inexorable. impervious, indifferent. deaf. foreign, closed, inaccessible; resistant.)

- II

1. Qu'on ne sent pas, qu'on ne perçoit pas ou qui est à peine sensible, perceptible. imperceptible, léger. (Not being sensed, not being perceived or being just barely sensible, perceptible. imperceptible, slight.)
2. Graduel, progressif. (Gradual, progressive.)

System output for a request to generate the semantic space associated with the French headword *insensible*.
Your query was: insensible. There are 71 synonyms and 93 cliques.

Table 8
Synonym list for the headword *insensible* (French lexical database).

insensible:	adamantin, anesthésié, apathique, aride, assoupi, blasé, calleux, calme, cruel, de marbre, desséché, dur, détaché, endormi, endurci, engourdi, flegmatique, frigide, froid, féroce, glacial, glacé, immobile, impassible, imperceptible, imperméable, imperturbable, impitoyable, implacable, impénétrable, inabordable, inaccessible, inanimé, inapparent, indifférent, indiscernable, indolent, indolore, inerte, inexorable, inflexible, inhumain, ininflammable, insaisissable, insignifiant, invisible, invulnérable, léger, léthargique, mort, neutre, négligeable, obtus, paralysé, progressif, rebelle, rigide, réfractaire, sans coeur, sans entrailles, sans coeur, sec, sourd, stoïcien, stoïque, suprasensible, sévère, timide, égoïste, étranger, étroit.

Table 9
Clique list for the headword *insensible* (French lexical database).

1 : adamantin, dur, insensible
2 : anesthésié, insensible
3 : apathique, endormi, indolent, insensible
4 : apathique, endormi, inerte, insensible
5 : apathique, flegmatique, impassible, imperturbable, indifférent, insensible
6 : apathique, indifférent, indolent, insensible
7 : apathique, inerte, insensible, mort
8 : apathique, insensible, léthargique
9 : aride, desséché, froid, insensible, sec
10 : aride, froid, indifférent, insensible, sec
11 : aride, froid, insensible, sec, sévère
12 : assoupi, endormi, engourdi, insensible
13 : blasé, flegmatique, froid, indifférent, insensible
14 : calleux, dur, endurci, insensible
15 : calme, flegmatique, froid, impassible, imperturbable, insensible
16 : calme, froid, inanimé, insensible
17 : calme, immobile, impassible, insensible
18 : calme, immobile, inanimé, insensible
19 : cruel, dur, féroce, impitoyable, implacable, inexorable, inhumain, insensible
20 : cruel, dur, impitoyable, implacable, inexorable, inflexible, inhumain, insensible
21 : cruel, dur, impitoyable, implacable, inexorable, inflexible, insensible, sévère
22 : cruel, dur, implacable, inflexible, inhumain, insensible, rigide
23 : cruel, dur, implacable, inflexible, insensible, rigide, sévère
24 : cruel, dur, indifférent, inhumain, insensible
25 : de marbre, glacial, impassible, insensible
26 : desséché, dur, froid, insensible, sec
27 : dur, endurci, impitoyable, implacable, inflexible, insensible
28 : dur, endurci, impitoyable, insensible, sans coeur
29 : dur, endurci, indifférent, insensible, sans coeur, sec
30 : dur, froid, glacial, impassible, insensible
31 : dur, froid, glacial, insensible, sec
32 : dur, froid, impassible, implacable, insensible
33 : dur, froid, impassible, indifférent, insensible
34 : dur, froid, impitoyable, implacable, insensible, sévère
35 : dur, froid, inaccessible, indifférent, insensible
36 : dur, froid, indifférent, insensible, sec
37 : dur, froid, insensible, sec, sévère
38 : dur, impassible, implacable, inflexible, insensible
39 : dur, impassible, indifférent, insensible, stoïque
40 : dur, impitoyable, inhumain, insensible, sans coeur
41 : dur, indifférent, inhumain, insensible, sans coeur
42 : dur, inhumain, insensible, sans coeur
43 : dur, inhumain, insensible, sans entrailles
44 : dur, insensible, invulnérable
45 : dur, insensible, rigide, sec, sévère
46 : dur, insensible, rigide, stoïque, sévère
47 : détaché, flegmatique, imperturbable, indifférent, insensible
48 : détaché, indifférent, insensible, étranger
49 : endormi, engourdi, indolent, insensible
50 : endormi, engourdi, inerte, insensible
51 : engourdi, froid, inerte, insensible
52 : engourdi, immobile, inerte, insensible, paralysé
53 : engourdi, insensible, léthargique
54 : engourdi, insensible, rigide
55 : flegmatique, froid, impassible, imperturbable, indifférent, insensible
56 : frigide, froid, glacé, insensible
57 : froid, glacial, glacé, impassible, insensible
⋮

Table 10
Examples of cliques generated in the target language for the headword *insensible*.

⋮

3 apathetic, cold, dull, indifferent, languid
4 apathetic, cold, unfeeling
5 apathetic, cool, impassive, indifferent
6 apathetic, cool, indifferent, unconcerned
7 apathetic, dull, languid, sluggish
8 apathetic, impassive, indifferent, languid
9 apathetic, impassive, indifferent, phlegmatic

⋮

14 apathetic, phlegmatic, sluggish
15 arid, dried, parched
16 arid, dry, parched
17 asleep, numb
18 austere, bare
19 austere, bitter, harsh, severe
20 austere, cold
21 austere, grave, hard, harsh, severe
22 austere, hard, hard-hearted, harsh, stern
23 austere, hard, hard-hearted, heartless, stern
24 austere, hard, harsh, rigid, severe, stern, strict

⋮

28 bitter, cruel, fierce, ruthless, savage
29 bitter, cruel, harsh, ruthless
30 bitter, cruel, harsh, severe
31 callous, cold, dead, indifferent
32 callous, cold, senseless, unfeeling
33 callous, cruel, hard, hard-hearted, heartless
34 callous, cruel, hard-hearted, heartless, unfeeling
35 callous, cruel, heartless, inhuman
36 callous, hard, hardened
37 callous, hard-hearted, insensitive, unfeeling
38 callous, hardened, insensitive, unfeeling
39 callous, impassive, indifferent
40 callous, impassive, insensible, unfeeling
41 callous, insensible, insensitive, unfeeling
42 callous, insensible, senseless, unfeeling
43 calm, calmness, composure, cool, quiet
44 calm, composed, cool, impassive, imperturbable
45 calm, composed, cool, quiet

⋮

54 cold, dead, frigid, indifferent
55 cold, dry, dull, frigid, languid
56 cold, dull, frigid, indifferent, languid
57 cold, freezing, frigid, frosty, icy
58 cold, frigid, frosty, frozen, icy
59 cold, frigid, icy, indifferent
60 cold, senseless, unconscious
61 cool, detached, indifferent, unconcerned
62 cool, emotionless, impassive, imperturbable
63 cool, impassive, indifferent, stoical
64 cramped, dry, stiff

65 cramped, stiff, tight
66 crisp, frosty
67 cruel, ferocious, fierce, ruthless, savage
68 cruel, grave, hard, harsh, severe
69 cruel, hard, hard-hearted, harsh, stern

⋮

77 cruel, heartless, inexorable, pitiless, relentless
78 cruel, heartless, inexorable, relentless, stern
79 cruel, heartless, inhuman, merciless, pitiless, ruthless
80 cruel, heartless, merciless, pitiless, relentless, ruthless, unfeeling
81 cruel, implacable, inexorable, pitiless, relentless
82 cruel, implacable, merciless, pitiless, relentless
83 cruel, inexorable, relentless, severe, stern
84 cruel, inhuman, merciless, pitiless, ruthless, savage
85 dead, extinct, inanimate, lifeless
86 dead, idle, inert
87 dead, inanimate, inert, lifeless
88 dead, indifferent, inert
89 dead, numb, paralytic
90 deaf, indifferent
91 difficult, hard, stiff
92 difficult, hard, tough
93 difficult, obscure
94 dozing, drowsy
95 drowsy, lethargic, sleepy
96 dry, severe, stiff
97 dull, dulled

⋮

98 dull, expressionless
99 dull, faint, languid
100 dull, inanimate, inert, lifeless
101 dull, indifferent, inert, languid
102 dull, indifferent, inert, neutral
103 dull, inert, languid, lethargic, sluggish

⋮

111 frivolous, idle, light, trivial
112 frivolous, indifferent, insignificant, trifling, unimportant
113 frivolous, insignificant, trifling, trivial, unimportant
114 frivolous, light, trifling, trivial
115 hard, hardened, tough
116 hard, heartless, relentless, unyielding
117 hard, inflexible, relentless, stern
118 hard, inflexible, relentless, unyielding
119 hard, inflexible, rigid, stern
120 hard, inflexible, rigid, stiff, stubborn, unyielding
121 hard, inflexible, rigid, tough, unyielding
122 hard, rigid, severe, tough

⋮

128 immobile, inert, motionless
129 impalpable, imperceptible, intangible, invisible
130 impassive, indifferent, phlegmatic, stoical
131 impassive, indifferent, unmoved
132 impenetrable, inaccessible, unapproachable
133 impenetrable, incomprehensible, inscrutable, unfathomable
134 impenetrable, incomprehensible, obscure
135 impenetrable, unapproachable, unfathomable

136 imperceptible, indiscernible, insensible
137 imperceptible, indiscernible, invisible
138 implacable, inexorable, inflexible, relentless

⋮

146 inflexible, intractable, stubborn, unyielding
147 insensible, senseless, unconscious
148 insensitive, unmoved
149 insignificant, slight, trifling, trivial, unimportant
150 lethargic, phlegmatic, sluggish
151 lethargic, sleepy, sluggish
152 light, slight, trifling, trivial

⋮

Table 11
Clique list for the headword *good* (English standard lexical database).

⋮

6 : able, adequate, capable, competent, effective, good
7 : able, adroit, clever, dexterous, expert, good, skilful
8 : able, capable, clever, expert, good, skilful
9 : able, capable, competent, effective, efficient, good
10 : absolutely delicious, delectable, delicious, good, gorgeous, lovely, scrumptious, yummy
11 : adept, expert, good, practiced, proficient, skilful, skilled, skillful
12 : adequate, competent, good, satisfactory, sufficient
13 : adequate, full, good
14 : admirable, commendable, deserving, good, meritorious, worthy
15 : admirable, deserving, estimable, good, meritorious, worthy

⋮

27 : advantageous, beneficial, good, helpful, salutary
28 : advantageous, beneficial, good, propitious
29 : agreeable, enjoyable, good, pleasant
30 : agreeable, good, good-natured
31 : agreeable, good, lovely, nice, pleasant, sweet
32 : appetising, appetizing, delicious, good, lovely, nice, savory, savoury, tasty
33 : attentive, good, obliging
34 : attentive, good, sweet, well-behaved
35 : auspicious, benign, good, propitious
36 : auspicious, good, promising, propitious
37 : beneficent, benevolent, benign, good, gracious, kind
38 : beneficent, benevolent, generous, good, kind
39 : beneficent, good, helpful, kind
40 : beneficial, benign, good, propitious
41 : beneficial, friendly, good, helpful
42 : beneficial, friendly, good, propitious

⋮

48 : benevolent, benign, good, gracious, kind, kindly
49 : benevolent, benign, good, gracious, propitious
50 : benevolent, friendly, good, gracious, kind, kindly

⋮

56 : commendable, creditable, deserving, good, meritorious, worthy
57 : commendable, creditable, good, honorable, honourable, worthy
58 : commendable, creditable, good, honourable, meritorious, worthy
59 : commendable, deserving, exemplary, good
60 : competent, expert, good, skilful, skilled, skillful, versed
61 : considerable, fair, good, respectable
62 : considerable, fair, good, serious, substantial
63 : considerable, fair, good, sound, substantial

⋮

71 : creditable, estimable, good, honorable, honourable, worthy
72 : creditable, estimable, good, honourable, meritorious, worthy
73 : dear, good, near
74 : dear, good, precious, sweet
75 : dear, good, precious, valuable
76 : decorous, good, respectable
77 : delectable, delicious, enjoyable, good, pleasant
78 : delectable, delicious, excellent, exquisite, good, lovely, scrumptious
79 : delectable, delicious, good, lovely, pleasant
80 : delectable, delicious, good, lovely, savoury, scrumptious, tasty
81 : delectable, delicious, good, lovely, scrumptious, tasty, yummy
82 : delicious, good, lovely, nice, pleasant
83 : dependable, good, honest, reliable, true, trustworthy
84 : dependable, good, reliable, safe, secure
85 : dependable, good, reliable, safe, trustworthy
86 : dependable, good, reliable, secure, solid

⋮

92 : effective, efficient, good, serviceable
93 : effective, good, in effect, in force
94 : estimable, good, honorable, honourable, respectable, worthy
95 : excellence, good, goodness, merit, virtue, worth
96 : excellent, exemplary, good
97 : excellent, exquisite, fine, good, lovely
98 : excellent, good, noble, worthy
99 : exemplary, good, virtuous
100 : expert, good, practiced, skilful, skilled, skillful, versed, well-versed
101 : exquisite, fine, good, precious
102 : fair, good, honest, honourable, just, right, upright
103 : fair, good, honest, honourable, respectable
104 : fair, good, honest, honourable, righteous, upright
105 : fair, good, honest, serious

⋮

110 : fine, good, well
111 : friendly, gentle, good, kind, kindly, nice, sweet
112 : friendly, good, gracious, kind, kindly, nice, sweet
113 : friendly, good, helpful, kind
114 : friendly, good, kind, kindly, nice, pleasant, sweet
115 : friendly, good, propitious, well-disposed
116 : full, good, large

⋮

121 : gentle, good, noble
122 : genuine, good, honest, right, true
123 : genuine, good, real, solid
124 : genuine, good, real, true
125 : genuine, good, right, sound, true
126 : genuine, good, right, sound, valid

⋮

134 : good, helpful, kind, obliging
135 : good, holy, righteous, virtuous
136 : good, honest, honorable, honourable, moral, righteous, upright, virtuous
137 : good, honest, honorable, honourable, respectable
138 : good, honest, honourable, just, right, true, upright
139 : good, honest, honourable, just, upright, virtuous
140 : good, honest, honourable, moral, right, upright
141 : good, honorable, honourable, virtuous, worthy
142 : good, honourable, meritorious, virtuous, worthy
⋮

Acknowledgments We gratefully acknowledge support of the Agence Universitaire de la Francophonie and the FRANCIL network.

References

Abplanalp, Laure. 1998. *La pertinence et la traduction*. Travaux du centre de traduction littéraire, Lausanne, Switzerland.

Anderson, John R. 1983. *The architecture of cognition*. Harvard University Press, Cambridge.

Asher, Nicolas and Alex Lascarides. 1995. Lexical disambiguation in a discourse context. *Journal of Semantics*, 12(1):9–108.

Benzécri, Jean-Paul. 1992. *Correspondence Analysis Handbook*. Dekker, P, New York.

Chuquet, Hélène and Michel Paillard. 1989. *Approches linguistiques des problèmes de traduction*. Ophrys, Paris.

Collins, Allan M. and Elizabeth F. Loftus. 1975. A spreading-activation theory of semantic processing. *Psychological Review* 82:407–428.

Edelman, Shimon. 1998. Representation is representation of similarities. *Behavioral and Brain Sciences*, 21(4):449–498.

Edmonds, Philip and Graeme Hirst. 2002. Near-synonymy and lexical choice. *Computational Linguistics*, 28(2):104–144.

Fellbaum, Christiane. 1998. *WordNet: an Electronic Lexical Database.* MIT Press, Cambridge.

Foltz, Peter W., Walter Kinsch, and Thomas K. Landauer. 1998. The measurement of textual coherence with latent semantic analysis. *Discourse Processes*, 25:285–307.

Ide, Nancy and Jean Veronis. 1998. Word sense disambiguation: The state of the art. *Computational Linguistics*, 24(1):1–40.

Ikeda, Satoko. 1998. Manual response set in a stroop-like task involving categorization of English and Japanese words indicates a common semantic representation. *Perceptual and Motor Skills*, 87(2):467–474.

Illes, Judy and Wendy S. Francis. 1999. Convergent cortical representation of semantic processing in bilinguals. *Brain and Language*, 70(3):347–363.

Kinsch, Walter. 2001. Predication. *Cognitive Science*, 25:173–202.

Masson, Michael. 1995. A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(1):3–23.

Maupassant, Guy (de). 1881. *Contes et nouvelles*. Gallimard, Bibliothèque de la Pléiade, Paris.

Maupassant, Guy (de). 2002. *Original Short Stories, vol. 2*, trans. Albert M. C. McMaster and L. Quesada Project Gutenberg Release. http://www2.cs.cmu.edu/spok/metabook/maupassant.html.

Miller, George A. 1990. WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–312.

Ploux, Sabine. 1997. Modélisation et traitement informatique de la synonymie. *Linguisticae Investigationes*, 21(1):1–28.

Ploux, Sabine and Bernard Victorri. 1998. Construction d'espaces sémantiques à l'aide de dictionnaires informatisés des synonymes. *Traitement automatique des langues*, 39(1):161–182.

Pustejovsky, James. 1995. *Generative Lexicon.* MIT Press, Cambridge.

Reboul, Anne. 2000. Words, concepts, mental representations and other biological categories. In B. Peeters, editor, *The Lexicon-Encyclopedia Interface*. Elsevier, Amsterdam.

Rouibah, Aïcha, Sabine Ploux, and Hyungsuk Ji. 2001. Un modèle spatial des représentations lexicales impliquées dans la reconnaissance des mots écrits. In Hélène Paugam-Moisy, Vincent Nyckees, and Josiane Caron-Pargue (editors), *La cognition entre individu et société*. Hermès-Science, Paris.

Schütze, Hinrich. 1998. Automatic sense discrimination. *Computational Linguistics*, 24(1):97–124.

Vinay, Jean-Paul and Jean Darbelnet. 1996. *Stylistique comparée du français et de l'anglais*. Didier, Paris.

Vossen, Piek, editor. 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic, Dordrecht, the Netherlands.