Automatically Populating Acception Lexical Database through Bilingual Dictionaries and Conceptual Vectors

Mathieu Lafourcade

LIRMM
Laboratoire d'informatique, de Robotique
et de Microélectronique de Montpellier
MONTPELLIER - FRANCE.
lafourca@lirmm.fr
http://www.lirmm.fr/ ~lafourca

Abstract. The NLP team of LIRMM currently works on thematic and lexical disambiguation text analysis [*Lafourcade*, 2001]. We built a system, with automated learning capabilities, based on conceptual vectors for meaning representation. Vectors are supposed to encode *ideas* associated to words or expressions. In the framework of Acception Based Lexical Database (instantiated through the Papillon project), we devise some conceptual vector based strategies to automatically populate and check acceptions. In this context, an acception represents a meaning of an entry of a monolingual dictionary, and could be refined and associated to a conceptual vector. Vectors are used for decision making and link quality assessment.

1 Introduction

In the framework of the Papillon project, the NLP team of LIRMM currently works on strategies for automatically populating acception lexical database. Such strategies are based on the simultaneous exploitation of the conceptual vector model, monolingual and bilingual dictionaries. The conceptual vector model aims at representing thematic activations for chunks of text, lexical entries, locutions up to whole documents. Roughly speaking, vectors are supposed to encode *ideas* associated to words or expressions. The main applications of the model are thematic text analysis and lexical disambiguation [*Lafourcade*, 2001] and can found interesting ways for vector refinement through the lexical architecture proposed in Papillon. Practically, we have built a system, with automated learning capabilities, based on conceptual vectors and exploiting monolingual dictionaries. So far, from French, the system learned 87000 lexical entries corresponding to roughly 350000 vectors (the average meaning number being 5). We are conducting the same experiment for English.

With these lexical and vector resources, we can in conjunction with bilingual dictionaries, automatically construct a large set of acceptions (interlingual links).

This *lexical soup* ([**?**]) is built through an iterated process that involves several strategies. The ideas applied to French and English in our experiment are generic and could be extended to any language. The bootstrapping consists in producing a set of acceptions that are corresponding directly to meaning as defined in our French dictionary (we could as well have started from English). A conceptual vector base is thus associated to the acception base, and as such an acception owns a conceptual vector. Then, we add English meanings (as defined in the English monolingual dictionary) to the soup. Here, we have a two-step process. First, associate the meaning to a bilingual correspondence in the English-French dictionary and, second undertake an *acception identification*. This identification process gives one of the following answers about the current meaning: either 1) it should be linked to a given recognized acception, or there is no recognized acception and 2) a new one should be created, 3) another acception should be split (French *abats* to *ofals* and *giblets*) or 4) two or more acceptions should be merged for creating the new one (French *rivière* and *fleuve* to *river*). The decision process is based on thematic distance between vectors of meaning sets previously filtered by bilingual dictionaries.

In this paper, we first expose the conceptual vectors model and the notion of semantic distance and contextualization. Then, we expose the populating strategies that associate meanings to acceptions through sets of bilingual correspondences and conceptual distances.

2 Conceptual Vectors

We represent thematic aspects of textual segments (documents, paragraphs, syntagms, etc.) by conceptual vectors. Vectors have been used in information retrieval for long [*Salton et MacGill*, 1983] and for meaning representation by the LSI model [*Deerwester et al*, 90] from latent semantic analysis (LSA) studies in psycholinguistics. In computational linguistics, [*Chauché*, 90] proposes a formalism for the projection of the linguistic notion of semantic field in a vectorial space, from which our model is inspired. From a set of elementary notions, concepts, it is possible to build vectors (conceptual vectors) and to associate them to lexical items [1]. The hypothesis that considers a set of concepts as a generator to language has been long described in [*Rodget*, 1852] (*thesaurus hypothesis*). Polysemous words combine the different vectors corresponding to the different meanings. This vector approach is based on well known mathematical properties, it is thus possible to undertake well founded formal manipulations attached to reasonable linguistic interpretations. Concepts are defined from a thesaurus (in our prototype applied to French, we have chosen [*Larousse*, 1992] where 873 concepts are identified to compare with the thousand defined in [*Rodget*, 1852]). To be consistent with the thesaurus hypothesis, we consider that this set constitutes a generator space for the words and their meanings. This space is probably not

[1] Lexical items are words or expressions which constitute lexical entries. For instance, ‹*car*› or ‹*white ant*› are lexical items. In the following we will (some what) use sometimes *word* or *term* to speak about a *lexical item*.

free (no proper vectorial base) and as such, any word would project its meaning on this space.

2.1 Thematic Projection Principle

Let be $\mathcal{C}$ a finite set of n concepts, a conceptual vector V is a linear combination of elements c_i of $\mathcal{C}$. For a meaning A, a vector $V(A)$ is the description (in extension) of activations of all concepts of $\mathcal{C}$. For example, the different meanings of ‹*quotation*› could be projected on the following concepts (the *CONCEPT*[intensity] are ordered by decreasing values):

V(‹*quotation*›) =
STOCK EXCHANGE[0.7], *LANGUAGE*[0.6], *CLASSIFICATION*[0.52], *SYSTEM*[0.33], *GROUPING*[0.32], *ORGANIZATION*[0.30], *RANK*[0.330], *ABSTRACT*[0.25], ...

In practice, the largest $\mathcal{C}$ is, the finer the meaning descriptions are. In return, the computer manipulation is less easy. It is clear, that for dense vectors[2] the enumeration of the activated concepts is long and difficult to evaluate. We would generally prefer to select the thematically closest terms, i.e., the *neighborhood*. For instance, the closest terms ordered by increasing distance of ‹*quotation*› are:

$\mathcal{V}$(‹*quotation*›) = ‹*management*›, ‹*stock*›, ‹*cash*›, ‹*coupon*›, ‹*investment*›, ‹*admission*›, ‹*index*›, ‹*abstract*›, ‹*stock-option*›, ‹*dilution*›, ...

2.2 Angular Distance

Let us define $Sim(A, B)$ as one of the *similarity* measures between two vectors A et B, often used in information retrieval [*Morin*, 1999]. We can express this function as:

$$Sim(A, B) = \cos(\widehat{A, B}) = \frac{A \cdot B}{\|A\| \times \|B\|}$$

with "$\cdot$" as the scalar product. We suppose here that vector components are positive or null. Then, we define an *angular distance* D_A between two vectors A and B as:

$$D_A(A, B) = \arccos(Sim(A, B))$$

Intuitively, this function constitutes an evaluation of the *thematic proximity* and is the measure of the angle between the two vectors. We would generally consider that, for a distance $D_A(A, B) \leq \frac{\pi}{4}$, (i.e. less than 45 degrees) A and B are thematically close and share many concepts. For $D_A(A, B) \geq \frac{\pi}{4}$, the thematic proximity between A and B would be considered as loose. Around $\frac{\pi}{2}$, they have no relation. D_A is a real distance function. It verifies the properties of reflexivity, symmetry and triangular inequality. We can have, for example, the following angles[3]:

[2] Dense vectors are those which have very few null coordinates. In practice, by construction, all vectors are dense.

[3] Examples are extracted from http://www.lirmm.fr/~lafourca (values are in degrees)

D_A(‹*profit*›, ‹*profit*›)=0°	D_A(‹*profit*›, ‹*product*›)=32°
D_A(‹*profit*›, ‹*benefit*›)=10°	D_A(‹*profit*›, ‹*goods*›)=31°
D_A(‹*profit*›, ‹*finance*›)=19°	D_A(‹*profit*›, ‹*sadness*›)=65°
D_A(‹*profit*›, ‹*market*›)=28°	D_A(‹*profit*›, ‹*joy*›)=39°

The first value has a straightforward interpretation, as ‹*profit*› cannot be closer to anything else than itself. The second and the third are not very surprising since a ‹*benefit*› is quite synonymous of ‹*profit*›, in the ‹*finance*› field. The words ‹*market*›, ‹*product*› and ‹*goods*› are less related which explains a larger angle between them. The idea behind ‹*sadness*› is not much related to ‹*profit*›, contrary to its antonym ‹*joy*› which is thematically closer (either because of metaphorical meanings of ‹*profit*› or other semantic relations induced by the definitions). The thematic proximity is by no way an ontological distance but a measure of how strongly meanings may relate to each others.

The graphical representations of the vectors of ‹*exchange*› and ‹*profit*› shows that these terms are indeed quite polysemous. Two other terms (‹*cession*› and ‹*benefit*›) seems to be more focused on specific concepts. These vectors are the average of all possible meanings of their respective word in the general Thesaurus. It is possible to measure the level of *fuzziness* of a given vector as a clue of the number of semantic fields the word meaning is related to. Because of

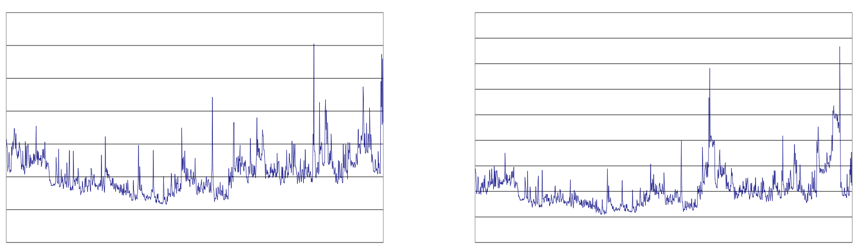

Fig. 1. Graphical representation of the vectors of 2 (rather polysemous) terms *exchange* and *profit*

the vagueness related either to polysemy or to lacks of precision (only 873 general concepts), we have to *plunge* our vectors into a specialized semantic space. However, we cannot cut loose from the general ones for two reasons. First, even non-specialized words may turn out to be pivotal in word sense disambiguation of specialized ones. Second, we cannot know beforehand if a given occurrence of a word should be understood in its specialized acception or more a general one.

2.3 Vector Operators

Vector Sum. Let X and Y be two vectors, we define their *normed sum* V as:

$$V = X \oplus Y \quad | \quad v_i = (x_i + y_i)/\|V\| \tag{1}$$

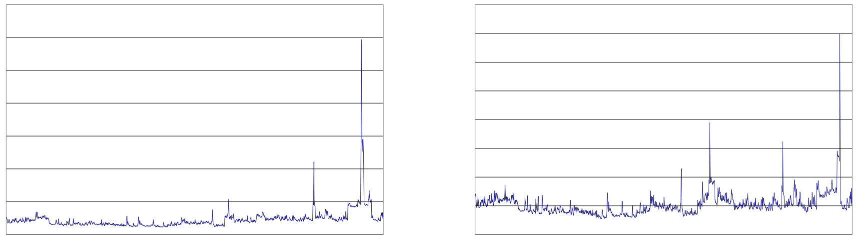

Fig. 2. Graphical representation of the vectors of terms *cession* and *benefit*

This operator is idempotent and we have $X \oplus X = X$. The null vector $\mathbf{0}$ is by definition the neutral element of the vector sum. Thus we write down that $\mathbf{0} \oplus \mathbf{0} = \mathbf{0}$. We then derive by deduction (without demonstration) the *closeness properties* associated to this operator (both local and general closeness).

$$\begin{aligned} D_A(X \oplus X, Y \oplus X) &= D_A(X, Y \oplus X) \\ &\leq D_A(X, Y) \end{aligned} \tag{2}$$

$$D_A(X \oplus Z, Y \oplus Z) \leq D_A(X, Y) \tag{3}$$

Normed Term to Term Product. Let X and Y be two vectors, we define V as *their normed term to term product*:

$$V = X \otimes Y \quad | \quad v_i = \sqrt{x_i y_i} \tag{4}$$

This operator is idempotent and $\mathbf{0}$ is absorbent.

$$\begin{aligned} V = X \otimes X = X \\ V = X \otimes \mathbf{0} = \mathbf{0} \end{aligned} \tag{5}$$

Contextualisation. When two terms are in presence of each other, some of the meanings of each of them are thus selected by the presence of the other, acting as a context. This phenomenon is called *contextualisation*. It consists in emphasizing common features of every meaning. Let X and Y be two vectors, we define $\gamma(X, Y)$ as the contextualisation of X by Y as:

$$\gamma(X, Y) = X \oplus (X \otimes Y) \tag{6}$$

These functions are not symmetrical. The operator γ is idempotent ($\gamma(X, X) = X$) and the null vector is the neutral element. ($\gamma(X, \mathbf{0}) = X \oplus \mathbf{0} = X$). We will notice, without demonstration, that we have thus the following properties of *closeness* and of *farness*):

$$\begin{aligned} & D_A(\gamma(X, Y), \gamma(Y, X)) \\ \leq & \{D_A(X, \gamma(Y, X)), D_A(\gamma(X, Y), Y)\} \\ & \leq D_A(X, Y) \end{aligned} \tag{7}$$

The function $\gamma(X, Y)$ brings the vector X closer to Y proportionally to their intersection. The contextualization is a low-cost meaning of amplifying properties that are salient in a given context. For a polysemous word vector, if the context vector is relevant, one of the possible meanings is *activated* through contextualization. For example, *bank* by itself is ambiguous and it vector is pointing somewhere between those of *river bank* and *money institution*. If the vector of *bank* is contextualized by *river*, then concepts related to finance would considerably dimmed.

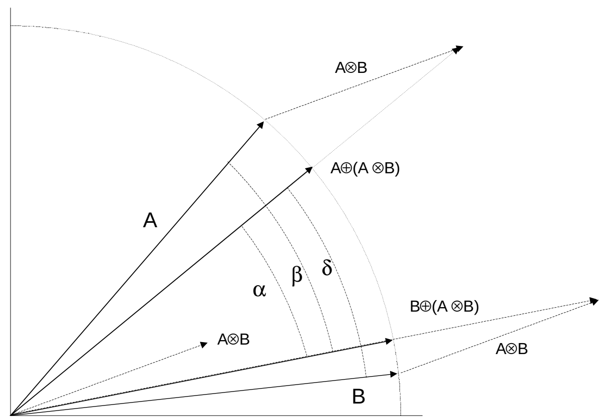

Fig. 3. Geometric represention (in 2D) of the contextualization function. The α angle represents the distance between A and B contextualized by each other.

3 Populating Acception Base

For undertaking the process of populating the acception base, several steps should be considered. First, a bootstrapping is needed to implement an initial set of acceptions. Then, from two bilingual dictionaries (Target Language to Source Language and vice-verse), we associate each equivalent set to corresponding meaning set in the vectorized monolingual dictionaries and then, to the acceptions.

3.1 Constraints

The global architecture of Acception Based Lexical Database has been presented in [*Sérasset and Mangeot*, 2001], and with many practical and theoretical aspects in [*Mangeot*, 2001]. We impose ourselves some constraints that should be respected by links and acceptions.

1. There is at most one link possible from one monolingual meaning to an acception. An entry with n meanings will be *indirectly* linked to at most n acceptions through its meanings (some meaning may not have found their corresponding acception yet)
2. Two meanings (either of the same or different monolingual entries) mays not be linked to the same acception. If they are synonymous, some kind of equivalence relation would be set between the two acceptions [*Lafourcade et Prince*, 2001].

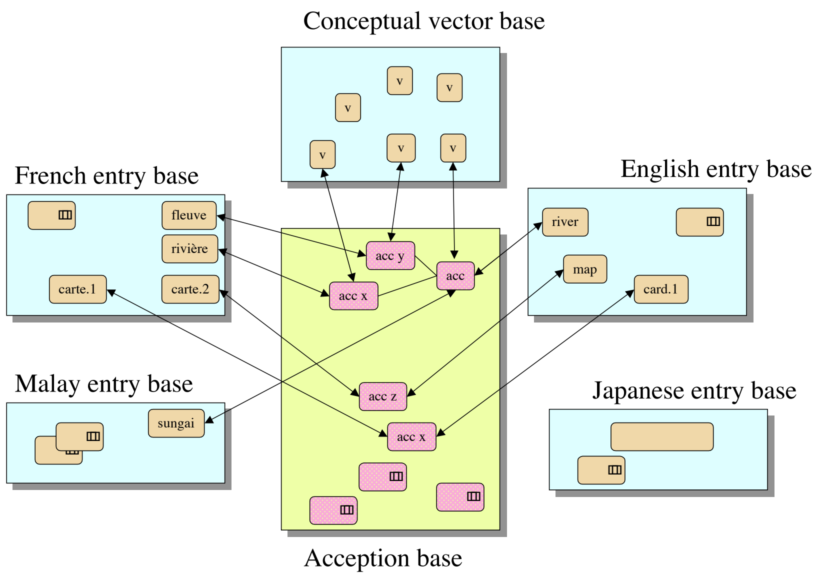

Fig. 4. Acception based architecture. Each monolingual (and artificial, in case of the conceptual vector base) dictionary are linked to the acception base. One meaning is associated to only one acception. Normally, every monolingual meaning and every acception should be linked somehow. In practice, the building process being by essence *any time*, there will always be some (as few as possible) orphaned or dandling objects.

Concerning acceptions, *meaning refinement* links are possible. An example is given in figure 6 with the acceptions linked to meanings *fleuve*, *rivière* and *river*. The link has the effect of meaning fusion: the acception *river* means at the same time *fleuve* and *rivière*.

3.2 Bootstrapping with a Vectorized Monolingual Dictionary

The bootstrapping consists in creating an acception for each meaning defined in the monolingual dictionary. In the context of our approach, we use our vectorized monolingual dictionary and take this opportunity to produce a Conceptual Vector Base where entries are vectors. We call, in this context, the language use for this bootstrapping the source language (Sl). In our experiments, we used French as source language.

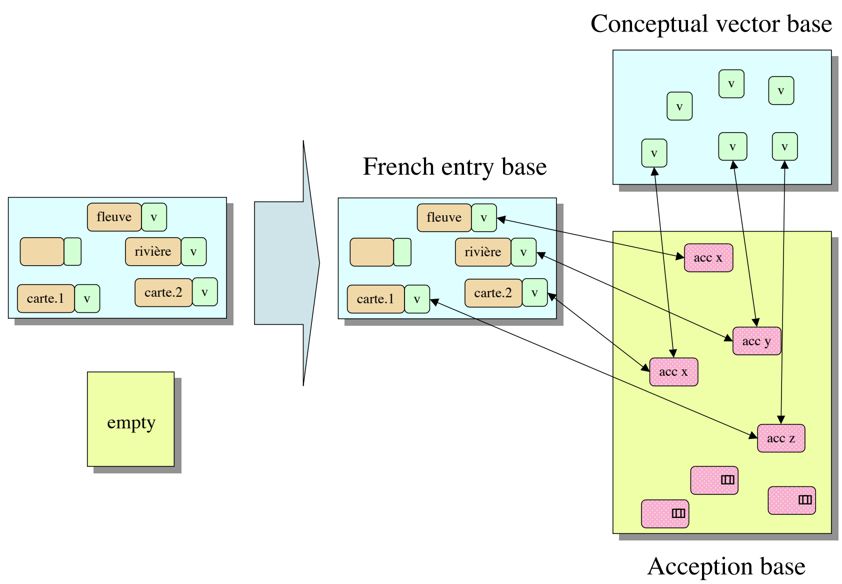

Fig. 5. Bootstrapping the populating process by promoting each word meaning to an acception.

3.3 Linking Translation Association to Meaning

We consider a bilingual dictionary Db from language A to language B and the bilingual dictionary $Db\sim$ form B to A. In either case, the (simplified) structure of a bilingual association is as follows:

$$Bd_w \equiv \langle cat, glose^*, equiv^+ \rangle$$

In the bilingual dictionary Db, the entry w_b has n subentries $w_{b,k}$ $(n > 0)$, each containing the following information: morphological data (like category Noun, Verb, Adjective, Adverb), some glosses, and equivalents. The glosses are by themselves words and are optional, but usually at least one gloss is given for each subentry, if the term is polysemous (or has different usage in different domains). There is always at least one equivalent. A typical example is:

demand $\equiv$

$1 : \langle VT, \{\textit{money, explanation, help}\}, \{\textit{exiger, réclamer}\}\rangle$
$2 : \langle VT, \{\textit{higher pay}\}, \{\textit{revendiquer, réclamer}\}\rangle$
$3 : \langle N, \{\textit{person}\}, \{\textit{demande}\}\rangle$
$4 : \langle N, \{\textit{duty, problem,situation}\}, \{\textit{revendication, réclamation}\}\rangle$
$5 : \langle N, \{\textit{for help,money}\}, \{\textit{demande}\}\rangle$

The structure of an entry in our vectorized dictionary Dv is as follows:

$$Bv_w \equiv \langle cat, def, vector\rangle$$

In a vectorized dictionary Dv, the entry w_v has n subentries $w_{v,k}$ ($n > 0$), each contains the following information: morphological data (like category Noun, Verb, Adjective, Adverb) a definition, and a vector. At this stage we purposely ignore the definition. The vector was beforehand computed from this definition.

For each $w_{b,k}$, we can compute its vector $V(w_{b,k})$. It is a two steps process: first we compute a context vector V_c from the glosses g_i if any, otherwise the context vector would be the empty vector. Then, we compute the weak contextualisation vector between the global term vector and the context vector.

$$V_c(w_{b,k}) = V(g_1) \oplus V(g_2) \oplus \cdots \oplus V(g_n)$$
$$V(w_{b,k}) = \gamma(V(w), V_c(w_{b,k}))$$

The bilingual subentry $w_{b,i}$ is then associated to the closest vectorized monolingual subentry $w_{v,j}$. The set is beforehand according to morphological information (we associate a noun to a noun, etc.) In other words, the vector $V(w_{v,j})$ should be the closest to $V(w_{b,i})$ among all monolingual subentry vectors.

At the end of this stage, for some meanings of a given word in the vectorized monolingual dictionary, we have a unique link to an entry on the bilingual dictionary. We can now try to associate the bilingual subentry to a monolingual entry of the target language. We do this process on the source (English) side and the target (French) side independently. Then the process of associating meanings to acceptions is undertaken and will use both equivalent dictionaries (SL to TL and TL to SL) and will make extensive use of conceptual vectors when decision making cannot rely only on equivalent set intersection.

3.4 Linking Meaning to Acception

The idea is to associate a meaning in the target language to a meaning in the source language. As a meaning is uniquely linked to an acception, the link toward this acception follows. Some definitions are needed here.

The call SLM the *Source Language Meaning* for which we try to find some acceptable acception (if any). Similarly, we call TLM a given *Target Language Meaning* and *acceptable TLM set*, the set of such meanings that could be acceptable equivalent for the SLM. We say that two vectors are *close* is their thematic distance is below a given threshold. The lower this threshold, the stricter we are

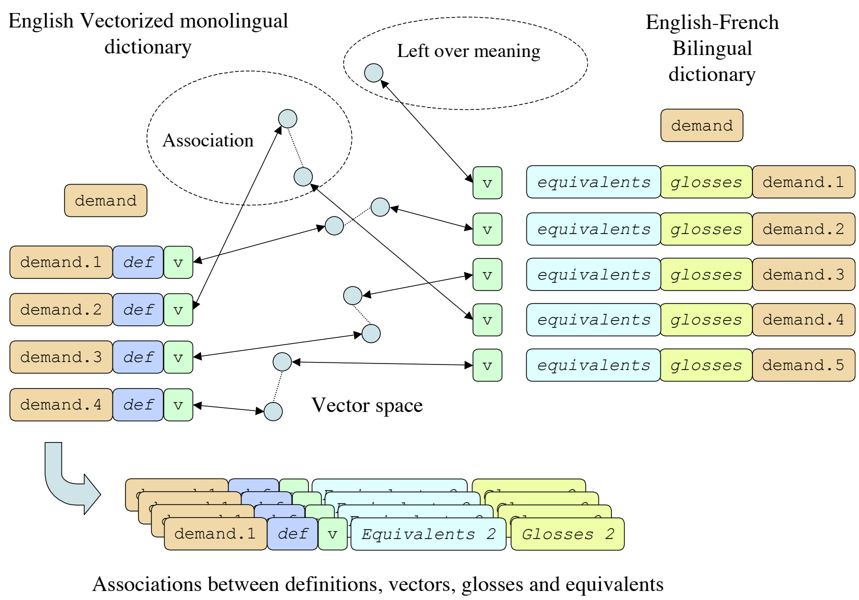

Fig. 6. Process of associating n bilingual entries to p monolingual entries. After a vector has been computer for each entry, association is done through vector distance. At most $\min(n, p)$ association are done. An association is done when the distance between the two vectors is lower than $\frac{\pi}{4}$ (empirical value).

for acception linked, but the less acception we would be able to automatically associate. An acceptable tradeoff for this threshold has been empirically found to be $\frac{\pi}{4}$. In fact, we discovered that this value could be dynamically adjusted during the populating process by minimizing some figures like: number of left-over meanings (meaning not associated to an acception), number of acceptions created (after the bootstrapping, we tend not to create to many new acception unless needed), size of refinement trees in the acception base (we tend to prefer short refinement trees). The basic situation are the following.

One meaning to one monosemic equivalent. In this case there is a direct selection between the SLM and the sole TLM (as there is no other choice). No vector operation is involved, unless for checking (cf upper Fig 7). If the vectors are not close, then a warning (toward lexicographs) is issued. The problem could originates, either from some errors in the equivalent, or from one (or both) vectors with inadequate activations.

One meaning equivalent to one polysemic equivalents. Here, we have to choose which are the equivalent meanings that could be acceptable (cf lower Fig 7). A first filter, is done one inverse equivalent, then among remaining meaning, only those which vectors are reasonably close.

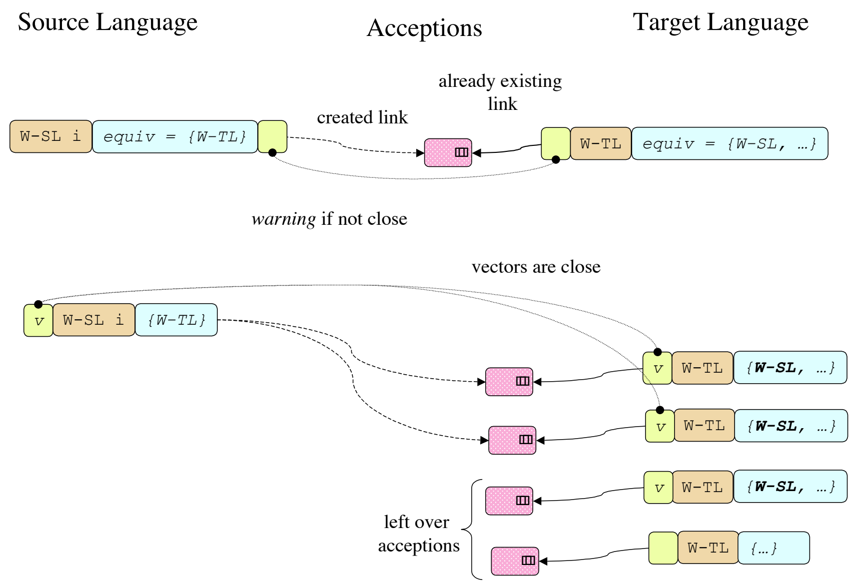

Fig. 7. *Upper part:* direct association between two monosemic equivalents. A warning may be issued if the distance between the two vectors is too wide. In this case, most probably, at least one of the vectors does not have proper concept activations. *Lower part:* one meaning with only one equivalent is to be associated to any subset of k meanings. Only the subset of the closest vectors is chosen.

One meaning to several polysemic equivalents. This case is only a generalization of the above, to several equivalents (cf Fig 8).

Error cases. The main error case is when the computed associated set is empty. It may happen if the information in the bilingual dictionaries are inconsistent: a SL w meaning has some equivalent that either (1) does not exist in the TL dictionary or (2) that actually exists but which every equivalent set does contain w. In that case, we issue a warning, but we still create an acception for the concerned meaning of w. The first case of error does really happen when the bootstrapping has been made with a reduced dictionary, and the linking is done with a larger dictionary. The second case is definitively an inconsistency between bilingual dictionaries.

The linking process creates links that may violate the above defined constraints, and a *Linking Cleanup* may be necessary.

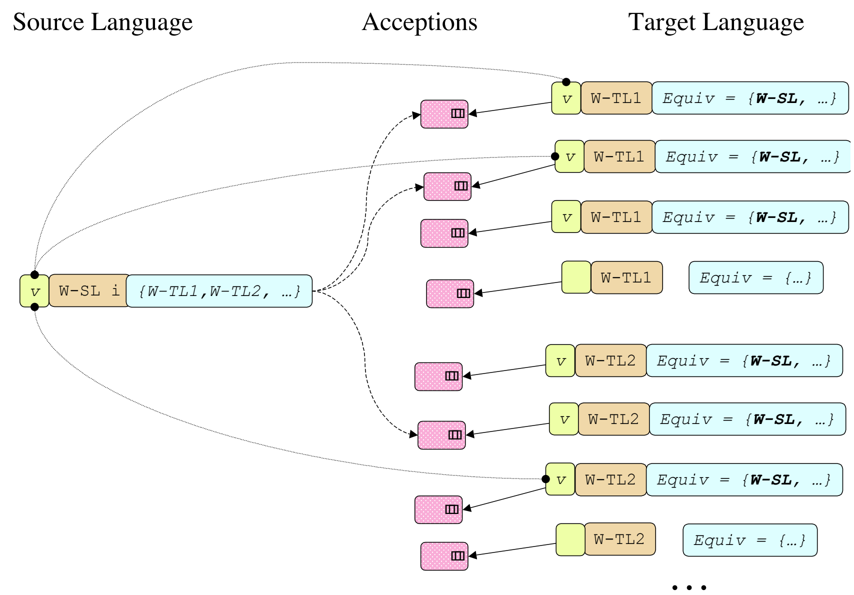

Fig. 8. Generalization of the above. One meaning with several associations should be associated to a subset of the meanings of each target association.

3.5 Linking Cleanup

This task is to check the link integrity. We have two basic cases: a given word meaning is linked to more than one acception, and a given acception is linked to more than one word meaning (cf fig 9).

1. **Multiple links**. We create an intermediate acception. The meaning is then linked only to this new acception (old links are deleted). Refinement links are created from the previous acceptions to the new one.
2. **Multiple meanings**. We have to choose which meaning should keep its link to the acception, the other one being deleted. Once again, the selection is done according to the closest vector.

In all generality, both situation could happen at the same time. The process is then undertaken iteratively with a priority to intermediate acception creation (cf Fig 10).

3.6 Extension

There are several left over aspects that still should be worth mentioning: vectors directly associated to acceptions, strengthvaluation on links, and the populating task viewed as a globally converging iterative process.

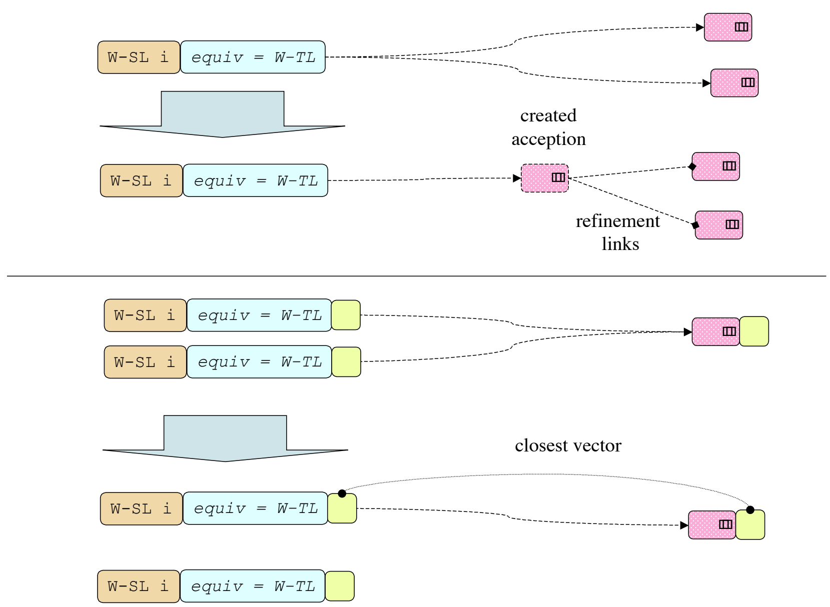

Fig. 9. *Upper part:* Multiple links case. An intermediate acception is created along with refinement links. *Lower part:* Multiple meanings case. Only one link should remain. The one which vectors are the closest is to be kept. The other meaning is orphan and should be attended later for either finding an other acceptable acception or creating an acception.

1. **Acception vector.** For each acception, we compute its own conceptual vector. These vectors are stored like entries of a monolingual dictionary, although this dictionary is constructed during the populating process. The vector of an acception is in its simplest form the mean of the vector of all linked meanings.
2. **Link strength.** Each created link has a strength (a value between 0 and 1) associated to it. The closer to 1, the harder the link is. The straightforward value of the link is the thematic distance between the vector of the link meaning or acceptions. However, some links may be softened (lowered) or hardened according to several criteria. One hardening criterion is when the Tl and SL terms are both monosemic and corresponding to each other. Other criteria rely on some added information related to lexical function applied to term meaning (cf Schwab paper on the subject: *Hardening of Acception Links Through vectorized Lexical Functions*).
3. **A global iterative process.** The populating process is done iteratively by autonomous agents. An agent explores the acception base and tries to evaluate links. For instance, a dangling acception (only one link) is to be

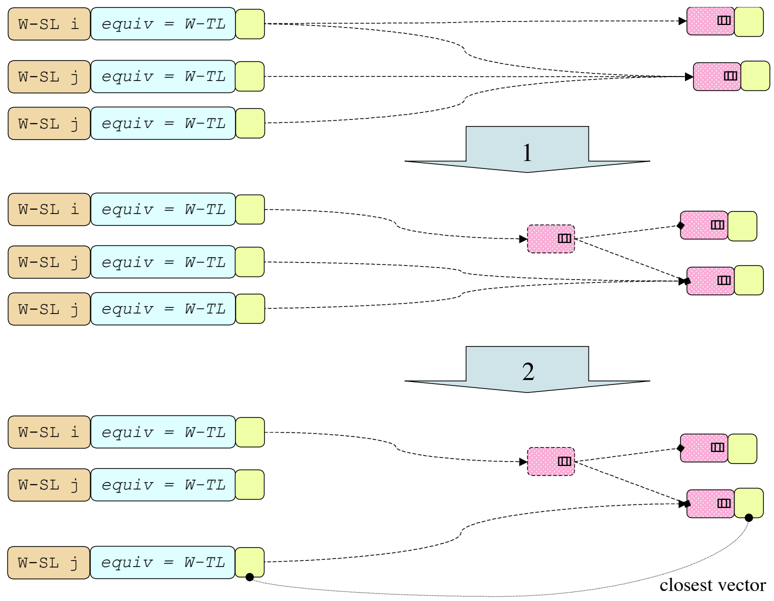

Fig. 10. Generalization of the above cases. Intermediate acception creation is always done first. Then follows link selection.

reconciliated to some TL entries. Orphan entries should also be taken care of: either a proper acception is found or (as least resort) a new dangling acception is created. This process is globally converging although in the meaning some back and forth hesitation (agents may create and delete a same set of objects repeatedly) does occur.

With acception vectors and link strength, it is possible to devise *walk through* strategies aiming at strengthening (or weakening) some links and also to formulate some warning about both dictionary and vector quality. Generally speaking, a warning is issued if information about translation equivalents and vectors do not agree. When in doubt, strategies involved always try to be conservative (better to have excess of links and acceptions than orphaned entries or acceptions).

4 Conclusion

This paper has presented a model of thematic representation using the formalism of conceptual vectors and discussed the effect of using it as part of decision taking in populating acception based lexical database.

The major applications of conceptual vectors are thematic analysis of texts, construction of large lexical databases and word sense disambiguation. Nevertheless, they can participate in lexical database building as an adequate tool for a first rough-hewing. Manually inserted lexicographic information (hard links) are then used by vector based autonomous agents to induce new acception links from lexical resources). In case of inconsistency, such agent can emit warnings.

The experiments we have conducted so far lead us to the following observations: the various parameter settings (like close distance threshold, adjustment parameters for link strength, etc.) are very difficult to guess. Most probably, they are not constant among the entire database. It is especially true, when we consider that the lexical density vary among domains. It is a much better strategy to let the system self-adapt.

These conclusions lead us to a more general one: populating automatically an acception base is feasible with quite simple rules but should be viewed as an iterative process. It is not to be considered that the acception base would ever be completed. Instead, we would expect (and have actually observed) that some lexical and acception regions would stabilize quite quickly (in 2 or 3 iterations).

References

[*Chauché*, 90] Jacques Chauché, *Détermination sémantique en analyse structurelle : une expérience basée sur une définition de distance.* TAL Information, 31/1, pp 17-24, 1990.

[*Chauché* and E. Sandford, 96] Jacques Chauché, *Détermination sémantique en analyse structurelle : une expérience basée sur une définition de distance.* Actes de MIDDIM-96, Le Col de Porte, France, August, 1996, pp 56-66.

[*Deerwester et al*, 90] Deerwester S. et S. Dumais, T. Landauer, G. Furnas, R. Harshman, *Indexing by latent semantic anlysis.* In Journal of the American Society of Information science, 1990, 416(6), pp 391-407.

[*Lafourcade et Prince*, 2001] Lafourcade M. et V. Prince *Synonymy and conceptual vectors.* Proc. of NLPRS'2001, Tokyo, Japan, August 2001, pp 127-134.

[*Lafourcade*, 2001] Lafourcade M . *Lexical sorting and lexical transfer by conceptual vectors.* Proc. of the First International Workshop on MultiMedia Annotation (Tokyo, Janvier 2001) 6 p.

[*Larousse*, 2001] Larousse. *Le Petit Larousse Illustré 2001.* Larousse, 2001.

[*Larousse*, 1992] Larousse. *Thésaurus Larousse - des idées aux mots, des mots aux idées.* Larousse, ISBN 2-03-320-148-1, 1992.

[*Lehmann et Martin-Berthet*, 98] Lehmann A. et Martin-Berthet F. *Introduction à la lexicologie. Sémantique et morphologie*, Paris, Dunod (Lettres Sup), 1998.

[*Lyons*, 1977] Lyons J. *Semantics* Cambridge : Cambridge University Press, 1977.

[*Mangeot*, 2001] . Mangeot-Lerebours M. *Environnements centralisés et distribués pour lexicographes et lexicologues en contexte multilingues* Phd thesis, Université Joseph Fourier, 2001.

[*Mel'čuk and al*, 95] Mel'čuk I., Clas A.et Polguère A. *Introduction à la lexicologie explicative et combinatoire.* , éditions Duculot, 1995.

[*Morin*, 1999] Morin, E. *Extraction de liens sémantiques entre termes à partir de corpus techniques.* Thèse de doctorat de l'Université de Nantes, 1999.

[*Muehleisen*, 1997] Muehleisen V. L. *Antonymy and semantic range in english.* Northwestern university, PhD, 1997.

[*Palmer, 1976*] Palmer, F.R. *Semantics : a new introduction* Cambridge University Press, 1976.

[*Polguère, 2001*] Polguére A. *Notions de base en lexicologie* Observatoire de linguistique sens-texte, 2001.

[*Quah all, 2001*] C. K. Quah and F. Bond and T. Yamazaki *Design and Construction of a machine-tractable Malay-English Lexicon.* Proc of AsiaLex, Seoul, 2001, pp 200-205.

[*Robert*, 2000] Robert *Le Nouveau Petit Robert, dictionnaire alphabétique et analogique de la langue française.* Ed. Robert, 2000.

[*Rodget*, 1852] Rodget P. *Thesaurus of English Words and Phrases.* Longman, London, 1852.

[*Riloff*, 1995] Riloff E. and J. Shepherd *A corpus-based bootstrapping algorithm for Semi-Automated semantic lexicon construction.* In. Natural Language Engineering 5/2, 1995, pp. 147-156.

[*Salton et MacGill*, 1983] Salton G. et MacGill M.J. *Introduction to modern Information Retrieval* McGraw-Hill, New-York, 1983.

[*Schwab*, 2001] Schwab D. *Vecteurs conceptuels et fonctions lexicales : application à l'antonymie*, mémoire de DEA d'informatiquer, Université Montpellier II, LIRMM, juin 2000, 60 p.

[*Schwab and al*, 2002] Schwab D., Lafourcade M. et Prince V. *Antonymy and Conceptual Vectors*, COLING 2002 processings.

[*Sérasset and Mangeot*, 2001] Sérasset G., Mangeot M. *Papillon lexical databases project: monolingual dictionnaries & interlingual links*, NLPRS 2001 processings.

[*Vèronis*, 92] J. Vèronis and N. Ide. *A feature-based model for lexical databases.* In Proc. of the Fourteenth International Conference on Computational Linguistic COLING'92, 1992, Nantes, France, 8 p.