

**Nicoletta CALZOLARI**  
 Istituto di Linguistica Computazionale  
 del CNR (Pisa, Italy)  
 Dipartimento di Linguistica (Università di Pisa)

### LES TENDANCES ACTUELLES DE LA LEXICOGRAPHIE ET DE LA LEXICOLOGIE ASSISTÉES PAR ORDINATEUR

Ces dernières années, on a assisté à une évolution de la linguistique automatique : un déplacement d'intérêt des aspects purement grammaticaux de la langue vers le lexique et, plus récemment, vers les grands corpus de textes s'est opéré. On est en présence d'une sorte d'évolution parallèle qui va des "toy-systems" (le prototype est le "block-world" de Winograd) au développement de systèmes experts (plus puissants, mais agissant dans un secteur limité, et donc avec un vocabulaire plus restreint) et, depuis peu, aux "very large NLP (*Natural Language Processing*) systems", tels les "machine-aided translation systems" ou les prototypes pour *Office Automation* où se manifestent l'exigence d'un vocabulaire à grandeur réelle et la nécessité d'une connaissance générale du monde. Ces deux tendances principales étant, tant du point de vue théorique que du point de vue de l'application, posées, il en a résulté que le traitement du lexique est devenu à la mode et celui des corpus textuels encore plus. Est apparue alors la nécessité d'avoir non seulement des lexiques informatisés très étendus ou des bases de données lexicales (les *Lexical DataBase* : LDB), mais aussi des lexiques où même les informations sémantiques seraient explicitées, c'est-à-dire des grandes bases de connaissances lexicales (les *Lexical Knowledge Base* : LKB).

L'évolution de la lexicographie et de la lexicologie assistées par ordinateur durant ces dernières années peut être ainsi retracée :

i) on part, dans les années 70, des *Machine Readable Dictionaries* (MRD), simples objets séquentiels, bien illustrés par les bandes de photocomposition,

ii) on passe, au début des années 80, par les LDB qui sont des objets plus structurés, doués d'un accès multidirectionnel aux données, interactifs par nature et comprenant souvent des taxinomies explicites ou des hiérarchies IS-A,

iii) on arrive, à la fin des années 80, aux LKB. Dans celles-ci, non seulement des liaisons IS-A, mais aussi beaucoup d'autres types de relations lexicales/sémantiques entre catégories conceptuelles sont formalisés ; de nouvelles directions d'accès aux données sont construites et des mécanismes d'inférence et de déduction sont incorporés (d'ordinaire sous forme de réseaux conceptuels). A présent, l'objectif prioritaire est la création de vastes réservoirs de connaissances linguistiques, sous forme de descriptions linguistiques aussi complètes que possibles et réutilisables, structurées dans une grande LKB ou dans plusieurs types de bases linguistiques reliées entre elles (bases de connaissances grammaticales, lexicales, textuelles).

Après avoir rappelé les tendances actuelles dans les différents domaines -c'est-à-dire la demande sur une grande échelle de systèmes NLP au sein de la communauté de la linguistique automatique ainsi que la reconnaissance de la capacité d'un système linguistique assisté par ordinateur de traiter des dizaines de milliers d'items lexicaux pour des applications dans le monde réel, joint au fait que la lexicographie (en tant que profession de *language industry*) a une tradition très longue, et que la création d'une LDB au contenu et à la dimension appropriés est très coûteuse et requiert beaucoup de temps (ce qui fait que le redoublement des efforts peut constituer un facteur aggravant)-, il faut souligner qu'un des mots-clé dans le domaine des LDB est devenu récemment le *réemploi (reusability)*.

*Réemploi* d'informations déjà existantes ou *réemploi* par rapport à des informations futures.

Dans le premier cas, on entend réemployer des informations lexicales implicitement ou explicitement présentes dans des ressources lexicales pré-existantes (par exemple, MRD, base de données (data base : DB) terminologiques, corpus de textes, etc.) pour aider à construire une LKB.

Dans le deuxième cas, on entend construire une LKB en mesure de permettre aux différents usagers (aussi bien aux différents systèmes de NPL qu'aux utilisateurs humains, lexicographes, traducteurs ou simples usagers de dictionnaires) d'obtenir -avec les interfaces appropriées- les informations en rapport avec leurs

différents objectifs.

En ce qui concerne la *reusability*, entendue dans son premier sens, elle est à l'origine, d'une certaine façon, de la naissance du projet ESPRIT, "Acquisition de la connaissance lexicale pour des systèmes d'élaboration du langage naturel" (AQUILEX), dans lequel sont engagés des groupes de chercheurs de Cambridge, d'Amsterdam, de Dublin, de Paris, de Barcelone et de Pise (coordinateurs). Son but principal est de développer des techniques et des méthodologies destinées à utiliser les MRD existants pour la construction de composantes lexicales à l'intention des systèmes NLP.

A partir de sources multiples MRD, on obtient surtout des informations lexicales, dans un contexte multilingue, en vue de créer une unique LKB multilingue. La base de connaissances aura alors sa source dans une structure conceptuelle/sémantique commune liée et qui, définissant les sens des mots individuels pour les langues examinées, sera assez riche pour pouvoir soutenir un modèle intensif de connaissance *profonde* d'élaboration du langage. Elle renfermera un vocabulaire général essentiel avec des informations phonologiques, morphologiques, syntaxiques et sémantico-pragmatiques associées, capables de s'étendre dans les composantes lexicales d'une grande variété de systèmes pratiques NLP [Boguraev *et alii*, 1988].

Si l'on examine le deuxième sens du terme *reusability*, on s'aperçoit qu'il est strictement lié à deux autres caractéristiques essentielles à une LDB.

La première caractéristique d'une LDB est d'être multifonctionnelle, et ceci essentiellement du point de vue de l'application. La LDB doit être un dépôt central de données qui peuvent être *réemployées* pour des buts différents et dans plusieurs applications soit procédurales, soit humaines, par l'intermédiaire de différentes interfaces. Le lexique est évidemment une composante essentielle de tout système NLP (pour analyser, produire, répondre à des questions par traduction automatique, pour récupérer les informations, pour la lemmatisation, l'intelligence artificielle, etc.) et la pratique courante consiste à construire une composante lexicale *ad hoc* pour chaque projet de langage naturel NLP. Il faudrait plutôt s'orienter vers de vastes lexiques (tant en extension qu'en profondeur de représentation), dans lesquels les informations seraient représentées de façon à être aisément interfacées grâce à des procédés d'application différents selon les exigences de l'application même. Ainsi la même série de données pourrait servir à des applications différentes. Chaque interface ne projetterait sur l'application spécifique que la représentation de données utile à des besoins

particuliers. A cet égard, un autre aspect essentiel d'une LDB est d'être aisément extensible : c'est dire que les chercheurs doivent pouvoir ajouter leurs propres informations en conformité avec le contenu effectif de la LDB.

La deuxième caractéristique d'une LDB concerne le point de vue théorique, et réside dans le fait que cette base est poly-théorique, c'est-à-dire multifonctionnelle par rapport à différentes théories linguistiques. Comme nous l'avons déjà dit, beaucoup de recherches ont été effectuées jusqu'à présent à un niveau expérimental, c'est-à-dire avec des prototypes lexicaux de petites dimensions. De plus, on donnait traditionnellement une grande importance à la représentation, l'organisation et l'usage de la connaissance linguistique figée et réglée par des procédés linguistiques. Les données lexicales étaient jugées peu importantes ou, du moins, faciles à traiter. S'il reste indiscutable que différentes théories linguistiques et organisations associées par ordinateur ont des conséquences considérables sur la construction grammaticale, on a accordé une moindre attention aux conséquences sur le lexique. Quoi qu'il en soit, nous avons l'intuition que les lexiques destinés à différentes théories linguistiques peuvent contenir des informations identiques d'un certain point de vue, puisqu'elles décrivent les mêmes faits linguistiques. Mais il nous faut garantir la validité de cette intuition avant de commencer à rendre effectives les informations requises par les systèmes NLP dans une LDB.

Le fait d'être polythéorique crée des difficultés ; c'est pourquoi, une étude de faisabilité a été entreprise pour établir la possibilité d'atteindre un certain niveau de cohérence entre les différentes théories destinées à partager la même quantité d'informations lexicales, et si c'est possible, de déterminer jusqu'à quel niveau d'analyse linguistique on peut envisager une représentation *neutre* ou poly-théorique des propriétés linguistiques.

Un groupe de travail engageant d'éminents représentants des plus importantes *écoles linguistiques* actuelles a été constitué. Il doit étudier dans le détail la possibilité de représenter les informations linguistiques employées souvent dans les analyseurs et les générateurs (par exemple, les plus grandes catégories syntaxiques, la sous-catégorisation et l'intégration, les classes de verbes, les taxinomies nominales, etc.), de façon à pouvoir les réemployer dans les structures théoriques suivantes : organisation et liaisons, grammaire de structure de la phrase généralisée, grammaire fonctionnelle lexicale, grammaire relationnelle, grammaire systémique, grammaire catégorielle. Cette équipe travaillera sur plusieurs langues. Elle commencera par

examiner avec précision le traitement que les théories précédentes ont réservé à un échantillonnage représentatif de verbes anglais et italiens. Car, du moment qu'un lexique polythéorique paraît réalisable, il devrait être possible de réemployer les données lexicales dans la structure de différentes catégories linguistiques (par exemple, GB, LFG, GPSG, RG, etc.) ainsi que dans la pratique lexicographique, grâce à des interfaces appropriées qui traduisent les données dans la notation/représentation [Walker, Zampolli, Calzolari, 1987].

### Les possibilités de *réemploi* des données préexistantes sous forme de MRD

Beaucoup d'articles et de livres ont déjà traité ce sujet [cf. Amsler, Boguraev, Briscoe, Byrd, Calzolari, Nagao, Picchi, Walker, Zampolli, etc.] aussi, nous désirons mettre l'accent sur ce que l'on considère comme étant l'évolution naturelle du travail fait jusqu'à présent dans ce domaine, c'est-à-dire la possibilité d'une application procédurale de *toute la série* des informations sémantiques implicitement contenues dans les MRD.

Dans cette structure, le dictionnaire est considéré comme la source principale des connaissances générales de base et, aujourd'hui, beaucoup de projets se donnent pour objectif l'acquisition du sens des mots à partir des MRD et l'organisation de la connaissance d'une LKB. La méthode est inductive et la stratégie adoptée heuristique : par une généralisation progressive des éléments communs trouvés dans les définitions du langage naturel, la tendance est de formaliser la connaissance générale de base implicitement contenue dans les définitions du dictionnaire, essentiellement pour tenter d'en dégager les principes fondamentaux et les rapports sémantiques. Tout cela signifie qu'on va au-delà de l'origine et de l'organisation des taxinomies, dont la méthodologie d'acquisition est maintenant bien déterminée [Chodorow *et alii*, 1985 ; Calzolari, 1982, 1984].

Pour identifier le terme *genus*, il faut simplement examiner la première partie de la définition ; cela peut être effectué si l'on tient compte du fait que les définitions sont NP (=syntagme nominal) quand il s'agit de définir un Nom, VPs quand il s'agit de verbes et AdjPs quand il s'agit d'adjectifs. La procédure consiste donc à chercher les lettres initiales de NP, VP, AdjP qui sont respectivement N, V ou Adj. Ils représentent les termes *genus* et sont joints à ce qui doit être défini par une liaison IS-A.

Quand on réorganise un MRD dans une structure taxinomique

(seulement avec des hiérarchies IS-A explicites), on emploie le MRD comme source de connaissances, mais par un seul des moyens possibles d'acquérir à partir de celui-ci un concept (sous une forme inductive), c'est-à-dire en reliant ce concept à tous ses exemples ; tous les exemples de la même catégorie/classe sont extraits et reliés ensemble à leur hyperonyme immédiat.

Dans l'approche LKB, le dictionnaire est considéré comme le *mécanisme de classification* le plus efficace, c'est-à-dire comme une méthode empirique pour illustrer les concepts et les divers types de relations lexicales/sémantiques [Calzolari, Picchi, 1988].

Notre approche méthodologique est la suivante :

1) partir des définitions en texte libre, en langage naturel et de forme linéaire, constituées d'ordinaire d'un terme *genus* et d'une partie *differentia* ;

2) analyser leur structure et leur contenu du point de vue linguistique et informatique ;

3) les transformer et les réorganiser en formats équivalents du point de vue informatique composés de nœuds et de rapports.

Mais, en ce qui concerne l'aspect informatique, le point 3 est subdivisé suivant les étapes suivantes :

a) *analyser* l'entrée du dictionnaire, à savoir *analyser une bande dictionnaire*, c'est-à-dire reconnaître essentiellement dans l'entrée lexicale les différentes parties significatives ;

b) produire une entrée lexicale avec une structure en forme d'arbre ;

c) réaliser une analyse morphologique et une désambiguïsation des homographes, c'est-à-dire ajouter les définitions des catégories grammaticales (POS) ;

d) après ces étapes préalables, adopter une technique de production adaptée à une analyse syntaxique très simple qui sache reconnaître approximativement les NP et PP (*syntagme nominal* et *syntagme prépositionnel*).

L'instrument le plus efficace est donc un mécanisme *pattern-matching* alimenté par :

(i) les résultats obtenus en parcourant les données du dictionnaire dans la LDB (comme les exemples décrits en annexe le montreront)

de façon à saisir les mots et les associations de mots les plus intéressantes,

(ii) les calculs de fréquence des définitions, mots et syntagmes, et évidemment,

(iii) l'intuition du linguiste.

Illustrons par quelques exemples le procédé d'analyse des définitions. Dans les figures jointes en annexe (de la p. à la p.), on cherche à simuler les procédés de recherche de la LDB en italien et ceux de lecture du dictionnaire en quête de mots, de structures et de modèles particuliers. Il est possible de recueillir certaines données sémantiques que l'on peut ensuite chercher et trouver dans un MRD structuré d'une façon appropriée.

La fig. 1 (p. 132) montre une partie de la taxinomie pour le mot italien *libro* (livre), c'est-à-dire une série de mots définis comme *types de livres* (on les voit ici avec leurs définitions). Mais figurent, dans le dictionnaire, d'autres éléments concernant les livres. Il est en l'occurrence possible d'extraire tous les verbes italiens en rapport avec les livres (voir fig. 2, p. 132), ainsi que tous les adjectifs et les noms qui ont à voir avec les livres (fig. 3 et 4, p. 133). Nous reviendrons plus loin sur le vocable *livre* pour mettre en évidence le genre d'informations qui manquent dans les dictionnaires, mais que l'on peut trouver dans les textes.

Notre travail actuel est consacré également à la formalisation de cet autre type de relations -pas aussi simple que les relations taxinomiques- qui associe les mots (ou les mots et les concepts) et pour l'extraction desquelles il faut analyser et élaborer la définition en entier, et pas seulement sa partie *genus*.

Voici quelques exemples des types de relations que l'on peut extraire des MRD. Dans la fig. 5 (voir p. 134) on voit, classés par ordre alphabétique, les premiers de 300 mots, reliés dans notre LDB par une liaison taxinomique au terme *strumento* (instrument). On retrouve le mot *attrezzo* (outil), qui figure dans cette liste, dans la fig. 6 (voir p. 134), laquelle nous montre quelques hyponymes de ce dernier terme avec leurs définitions. A partir de ces définitions, il est assez simple d'obtenir des relations sémantiques du type *used for*, *used in*, *shape*, *made of*, etc. Ces relations sont extraites au moyen d'un procédé *pattern-matching* agissant sur la partie *differentia* des définitions ; les différentes façons dont chaque relation est lexicalisée dans les définitions sont associées à la classification-relation. La relation *used for*, par exemple, dérive de modèles lexicaux tels que *per*, *usato per*, *atto a*, *che serve a*, *utile a* (pour, employé pour, apte à,

qui sert à, utile à) ; ces modèles lexicaux acquièrent cette signification quand ils se trouvent dans des positions particulières dans la définition des hyponymes du mot *strumento*. Ils peuvent dans d'autres contextes acquérir des significations différentes. Le résultat de cette analyse de contenu, du point de vue de la définition, est structuré dans la partie du réseau conceptuel illustré par la fig. 7 (voir p. 135).

D'autres sortes de relations sémantiques, qui peuvent être extraites assez facilement à partir des définitions, seront illustrées par quelques exemples. L'une d'entre elles est la relation *set of*, qui peut être spécifiée davantage en ce qui concerne le caractère de ses membres : nous avons des exemples de mots qui dénotent des *set of persone* (gens, personnes) (voir la fig. 8, p. 135), *oggetti* (objets) (voir la fig. 9, p. 136), etc.

D'autres types de données utiles concernent les informations sur des restrictions sélectives pour des verbes ou des adjectifs et dérivent surtout du modèle lexical *detto di* (dit de) après lequel on trouve le type de substantifs dont un Adjectif ou un Verbe peut normalement être le prédicat. La fig. 10 (voir p. 136) donne des Adjectifs et des Verbes renvoyant à des noms de *persone* (gens, personnes) ; la fig. 11 (voir p. 137) relève des Adjectifs qui sont habituellement attribués à des noms de couleur, génériques ou spécifiques ; par exemple pour les spécifiques, on trouve *giallo* (jaune), *rosso* (rouge), etc.

Un type intéressant de données relationnelles qui peuvent être extraites de certaines sortes d'actions sont les informations concernant les mots du lexique correspondant aux lexicalisations des rôles thématiques spécifiques de l'action même. Deux exemples vont nous permettre d'expliquer ce que nous entendons par là. La fig. 12 (voir p. 137) montre le résultat de questions posées à la LDB-italienne concernant toutes les entrées rattachées à des définitions qui comportent la forme-mot *vende* (il vend) (pas dans la position *genus*). Le résultat de la recherche est le suivant : on trouve 242 entrées dont 221 sont des noms de personnes qui *tipicamente vendono* (typiquement vendent) quelque chose, c'est-à-dire d'agent typiques par rapport à l'action de vendre. Ces entrées représentent des *fillers* lexicalisés cas/rôle dans les structure/cas de *vendere* (vendre). Ceci est dû au modèle de définition utilisé, c'est-à-dire *chi vende* (qui vend). Au sujet de cet exemple, nous pouvons faire des remarques intéressantes.

Une première remarque est que le même résultat a été obtenu en réalisant une recherche semblable dans un dictionnaire anglais. Après

réalisant une recherche semblable dans un dictionnaire anglais. Après avoir vu l'exemple italien, le groupe IBM de Yorktown [Byrd, 1989] a répété l'expérience et a obtenu le même résultat pour les données anglaises. Cela signifie qu'il y a bien une corrélation entre les modèles de définition employés dans la pratique lexicographique, et ceci indépendamment de la langue traitée. De fait, cette analogie entre conventions lexicographiques apparaît dans beaucoup d'autres exemples et peut être exploitée pour la création de la LKB multilingue, objectif principal du projet ESPRIT cité plus haut.

Une autre remarque concerne la co-occurrence dans ces définitions d'un verbe de ce type (*vendre*) avec un autre, *faire*, (lexicalisé en italien sous les formes *fabbricare, fare, preparare*, etc.). Beaucoup de ces noms d'agent se rapportent aussi à l'action de *fare* et, par conséquent, appartiennent à deux parties du réseau conceptuel résultant. On peut remarquer que le syntagme nominal (NP) qui suit le verbe marque le type d'objet qui est typiquement vendu (ou fait) par ces agents. Bien évidemment, il est possible d'obtenir le même genre d'informations concernant les noms des agents dans l'action de vendre si nous cherchons tous les substantifs dont le *genus terme* est le mot *venditore* (vendeur). A partir d'une telle interrogation nous pouvons trouver encore 131 substantifs *agents* (quelques-uns sont illustrés dans la fig. 13, voir p. 138). Ici encore quelques substantifs se rapportent à l'action *fare* (faire), tandis que le PP (syntagme prépositionnel) introduit par la préposition *di* (de) exprime l'objet qui est vendu.

Cet exemple montre que les mêmes informations exactement peuvent être récupérées en parcourant le dictionnaire de différentes manières, à partir de la connaissance de sa structure (en particulier de la structure interne des définitions). Dans la LKB finale, toutes ces données seront incorporées dans une seule partie du réseau, indépendamment des différentes manières de lexicaliser concepts et relations.

Au moyen d'une interrogation un peu différente, nous pouvons aisément récupérer également les noms des *lieux* où l'action de *vendere* se déroule normalement. La fig. 14 (voir p. 138) montre le résultat de la recherche des entrées dans les définitions où se trouve le mot *vendono* (ils vendent). Encore une fois, le fait que les noms de lieux puissent être repérés de cette façon est dû à cette *defining formula* employée par les lexicographes : *dovelin cui si vendono* (là où on les vend). Les 33 entrées trouvées partagent toutes ce modèle de définition : cette question est tout à fait sans *bruit*.

Nous pouvons noter que les termes *genus* sont représentés par le

nom générique *luogo* (lieu), ou bien par ses hyponymes, lesquels sont des noms génériques indiquant les lieux où l'on vend quelque chose : *negozio, bottega, bancarella* (magasin, boutique, étalage). Ceux-ci sont à leur tour hyperonymes des entrées définies. Ce type d'informations hiérarchisées est déjà formellement codifié dans les taxinomies emmagasinées dans la LDB.

Ce qui nous intéresse ici est la possibilité de formaliser et de rendre effectifs dans la LKB les autres types de relations sémantiques, telles que *location* et *thème*, par rapport aux actions de *vendre* et de *faire*. La relation *thème* -à savoir les objets typiquement vendus dans des lieux déterminés- est de nouveau exprimée par l'objet NP du verbe. Dans ce cas également, des données semblables sont récupérées en demandant les hyponymes de *negozio, bottega, etc.* Notre but est de formaliser toutes ces informations dans un réseau sémantique, ainsi que le montre le fragment représenté dans la fig. 15 (voir p. 139).

Les exemples décrits plus haut montrent que les possibilités de la LDB peuvent être employées profitablement pour analyser et obtenir des données linguistiques qu'il faut ensuite réorganiser et représenter dans la LKB.

Dans la LKB, ces types de concepts et de relations et les interdépendances entre les sens du mot sont expliqués en clair. Car, quand, dans cette base de connaissances, nous allons au-delà des taxinomies, plusieurs sortes d'associations utilement représentées dans un réseau conceptuel sont établies, et lors du passage d'un milieu *unilingue* à un milieu *multilingue*, des associations entre les différentes langues sont également réalisées. Ces associations (en ce qui concerne certaines parties des langues réductibles à une série commune de concepts et de relations) sont retenues à travers le réseau conceptuel commun construit en opérant, sur des langues différentes, mais à l'intérieur de la même *ligne de recherche*, c'est-à-dire en essayant de ranger dans le réseau sémantique :

- une *même* connaissance du monde,
- pour les *mêmes* objectifs (NLP, *text processing*, etc.)
- avec la *même* méthodologie,
- à partir du *même* type de sources (MRD),
- à l'intérieur du *même* type de représentation.

Le réseau sémantique commun deviendra donc le point de convergence des résultats réalisés à partir des stratégies d'acquisition de connaissances appliquées à un certain nombre de sources diverses mais homogènes, et le milieu multilingue constituera un banc d'essai

efficace pour évaluer cette stratégie de projet et l'utilisation d'une partie d'une LKB.

### Les possibilités de réemploi des dictionnaires bilingues

Non seulement les dictionnaires MR unilingues, mais aussi les MRD bilingues peuvent être profitablement utilisés en tant que sources d'informations lexicales pour la création de LDB et de LKB.

On peut traiter ces dictionnaires avec un double objectif car, d'une part, ils renferment d'intéressantes informations *unilingues*, et d'autre part, ils constituent une source de liaisons entre deux LDB unilingues [Calzolari, Picchi, 1986 ; Picchi, Peters, Calzolari, à paraître].

L'un de ces objectifs est d'intégrer les différents types d'informations normalement contenus dans les dictionnaires unilingues et bilingues, de manière à ce que le contenu de l'information de chaque élément figure dans le nouveau système.

Les dictionnaires bilingues renferment une plus grande quantité d'informations concernant les exemples d'usage, les expressions figées ou idiomatiques : ce type d'informations peut très bien être intégré dans le dictionnaire unilingue et rester aussi facile d'accès. Les entrées lexicales originales unilingues se voient ainsi augmentées des différents types d'informations provenant de l'entrée bilingue correspondante ; plusieurs différenciations de sens ou d'autres exemples encore, des informations syntaxiques, des collocations ou expressions idiomatiques, etc. y figureraient alors. On peut aussi renverser le point de vue et examiner les entrées bilingues dotées d'informations contenues d'ordinaire dans les entrées unilingues : dans la plupart des cas ce sont des définitions. Un des deux points de vue -virtuellement présents dans le système bilingue intégré- serait simplement activé et disponible à l'utilisateur dès son premier accès aux bases de données lexicales bilingues directes.

Il vaudra mieux garder dans une même structure les caractéristiques propres aux sources des dictionnaires unilingues et bilingues, ainsi que l'intégration des deux avec différentes représentations des données.

La conception générale du système LDB bilingue telle qu'on l'a imaginée est représentée dans la fig. 16 (voir p. 139). En ce qui concerne les dictionnaires bilingues, la méthode utilisée consiste à *réemployer* les données disponibles sous une *machine-readable form*

en analysant et en transformant les informations déjà contenues dans les dictionnaires ordinaires. La procédure d'élaboration du MRD bilingue est assez proche de celle décrite ci-dessus pour les dictionnaires unilingues (c'est-à-dire analyse de l'entrée lexicale, projet d'une nouvelle structure, réorganisation assistée par l'ordinateur, etc.). Après cette partie préalable, il faut revenir à la lecture de la LDB bilingue, se servant des éléments structuraux déjà formalisés, afin de découvrir des caractéristiques et des structures non visibles immédiatement dans le dictionnaire imprimé, mais utiles pour une exploitation ultérieure du dictionnaire-machine.

Suivant ce qui vient d'être dit concernant les premières phases du traitement des données dans le dictionnaire bilingue, il n'y aura aucune différence si on prend comme point de départ l'une ou l'autre des deux langues. Dans un certain sens, il n'y aurait plus une langue *source* et une langue *cible*, étant donné que les procédés de consultation et d'accès sont indépendants et neutres par rapport à la direction (l'objet deviendrait bidirectionnel). Aussi les renvois bidirectionnels seront-ils automatiquement produits pour les informations contenues à chaque niveau de signifié comme indicateurs sémantiques, c'est-à-dire les synonymes/hyperonymes ou les indicateurs contextuels.

La partie du dictionnaire bilingue que nous sommes en train de traiter et qui peut être partiellement formulée dans ses différents signifiés est le domaine des *indicateurs sémantiques*. Ceux-ci nous amènent à choisir une traduction plutôt qu'une autre. Le problème rencontré consiste en ce que les contraintes sont de nature différente ; en effet ces indicateurs sont i) synonymes ou hyperonymes de l'entrée, ii) sont indicateurs contextuels au titre de sujets ou objets typiques de verbes, de substantifs typiques dont un adjectif peut être le prédicat, etc. Il est possible de semi-automatiser le processus de désambiguïsation des différentes valeurs, après avoir analysé les diverses possibilités et projeté une typologie de ce qui peut paraître dans ce domaine.

Une autre alternative est représentée par l'emploi des bases de données lexicales unilingues comme instrument destiné à étendre les informations fournies par le mot individuel à toute la série de mots auxquels il se rapporte effectivement. Par exemple, l'entrée *vivido* a plusieurs traductions selon les indicateurs contextuels qui se rapportent au sujet (noté entre parenthèses dans l'exemple ci-dessous) :

*vivido* ----- (*colori*) vif, voyant

Dans certains cas, les restrictions sémantiques génériques sur l'objet éventuel peuvent être interprétées comme une caractéristique sémantique et peuvent être étendues d'une façon procédurale du répertoire lexical unilingue à tous les hyponymes possibles (au moment de la question) de façon à pouvoir choisir la traduction appropriée dans tout contexte où se trouverait un nom spécifique de *colore* (couleur) (tout cela est déjà réalisable dans notre LDB unilingue). Aussi les informations qui peuvent être formalisées au niveau sémantique dans un dictionnaire unilingue -lequel sert à distinguer entre plusieurs sens du mot- devraient être, au début, du même type que celles fournies dans les dictionnaires bilingues sous forme d'*indicateurs sémantiques* ou de *conditions sélectives* pour forcer le choix d'une traduction particulière.

D'une façon analogue, on peut opérer dans d'autres domaines pour rendre explicites des informations cachées ou pour introduire de nouvelles informations à partir des indications de structure ou de contenu.

Après la réorganisation du MRD bilingue en une LDB bien structurée, nous sommes en présence d'une tâche difficile : utiliser les données de cette dernière pour construire des liaisons entre deux LDB unilingues.

La difficulté réside évidemment dans l'ambiguïté des mots employés soit comme entrées soit comme traductions. On ne peut jamais connaître les sens d'un mot employé dans une situation particulière. On va essayer de résoudre ce problème, autant que possible, dans le projet ESPRIT, surtout par l'utilisation d'indicateurs sémantiques dans la LDB bilingue et de taxinomies et autres informations conceptuelles figurant dans la LDB unilingue.

Etablir des liaisons entre les différents sens d'un mot dans les dictionnaires unilingues et les différentes traductions dans un dictionnaire bilingue représente un des problèmes les plus intéressants concernant la connexion entre ces différentes sortes de dictionnaires. Du moment qu'en matière de traduction le problème consiste dans le choix approprié entre les différents sens de mots ambigus du point de vue lexical, il est, à notre avis, absolument nécessaire pour un système de *machine translation* ou de *machine-assisted translation* d'être connecté à une base de données linguistiques, c'est-à-dire à une source d'informations lexicales organisées sous forme de répertoire lexical par taxinomies multidimensionnelles, où la possibilité de désambiguïser les modèles lexicaux est au moins semi-automatisée.

L'une des principales utilisations du système devrait être celle de *machine-aided translation* (MAT), en tant qu'aide efficace pour les

*workstation*, dans laquelle on fournirait l'accès à plusieurs types de dictionnaires ainsi qu'à d'autres ressources lexicales, et où l'efficacité et les fonctions des bases de données lexicales et des bases de données textuelles seraient exploitées au maximum.

Un système bilingue, comme celui représenté dans la fig. 16 (voir p. 139), devrait avoir pour prétention :

- d'être un instrument pour les lexicographes ;
- d'être un instrument pour les études lexicologico-contrastives ;
- d'être un moyen pour améliorer les LDB unilingues ;
- d'être une aide pour construire des dictionnaires de type *machine translation* ;
- d'être un instrument pour enseigner les langues ;
- de constituer un dictionnaire assisté par ordinateur pour des usagers *ordinaires*.

A notre avis, l'un des avantages principaux d'une LDB bilingue est lié au type totalement différent de lecture de données qu'elle offre — lecture possible aussi bien grâce à l'accès multiple à ses données que grâce à ses liaisons avec la LDB unilingue. En particulier, il est possible de créer non seulement des connexions entre les couples de mots dans la L1 et la L2, comme dans le dictionnaire imprimé, mais surtout entre les groupes de familles de mots reliés sémantiquement — ce que nous considérons comme une caractéristique essentielle pour un véritable dictionnaire bilingue et pour tous les objectifs dont nous avons fait la liste ci-dessus.

#### Les possibilités de *réemploi* des corpus textuels et leur intégration dans les LKB.

Nous avons vu que les MRD sont de remarquables sources d'informations lexicales et sémantiques, mais qu'on ne peut malheureusement pas y trouver tout ce qu'il faut connaître sur le lexique. Il y a des informations très importantes qui manquent tout à fait, ou ne sont ni très valables, ni très fiables, ni encore aisément repérables. Pour ce qui concerne cette sorte d'informations, il faut recourir à différents types de sources [Calzolari, 1989a]. Notamment, certains types de données ne sont accessibles qu'à la suite d'une recherche théorique de faits lexicaux et l'on ne peut évaluer leur source dans le travail typique des linguistes, fondé surtout sur l'introspection et l'intuition du locuteur natif. Dans cette étude, nous ne traiterons pas ces données, mais il faut être conscient de leur existence.

existence.

Il faut souligner qu'il existe beaucoup de types de données qui peuvent être profitablement obtenus -plus ou moins directement- en traitant des corpus très étendus de données textuelles. Les résultats de ce traitement ont été analysés et évalués par le linguiste et/ou par le lexicographe, mais il est important de se rendre compte que pour certains types de phénomènes linguistiques l'étude effectuée à travers l'analyse du corpus est *préférée* à l'introspection : les classements et les expressions figées en sont les exemples typiques. On peut proposer la liste, non exhaustive, des informations lexicales pour lesquelles on peut repérer des données dans les corpus textuels -avec différents degrés de difficulté et à plusieurs niveaux d'exhaustivité :

- données de fréquence (au niveau du mot, de la forme-mot, des sens du mot, des associations de mots, etc. ) ;
- sous-catégorisation ;
- collocations, expressions figées, expressions idiomatiques ;
- rôles thématiques, valence ;
- restrictions sémantiques sur les sujets ;
- sujet, objet, modificateur, etc., typiques (ils sont différents des rôles thématiques, du moment qu'ils sont leurs *fillers* ; en un certain sens, les informations sont les mêmes, mais fournies *par exemple*) ;
- informations aspectuelles ;
- noms propres.

Prenons par exemple, le verbe *dividere* (partager), et examinons sa fréquence et ses contextes dans notre corpus d'environ 10 millions de mots. A partir d'un total de 840 concordances, on obtient les modèles syntaxiques les plus fréquents suivants:

|          |                          |     |
|----------|--------------------------|-----|
| dividere | NP in NP                 | 268 |
| "        | NP                       | 175 |
| "        | NP ( <i>tra</i> ) NP, NP | 80  |
| "        | NP <i>con</i> NP         | 78  |
|          |                          | 601 |

Les 239 contextes restants sont, quant à eux, répartis dans environ 10 autres structures de sous-catégorisation.

Si nous analysons les contextes, on voit que chaque structure de sous-catégorisation peut être très souvent reliée à un ou à plusieurs sens du mot, de sorte que nous pensons pouvoir employer ces structures comme une aide utile à la désambiguïsation du sens. En

analysant les concordances, on peut alors obtenir des données concernant :

- a) les structures syntaxiques ;
- b) leur ordre de fréquence et, par conséquent, leur pertinence respective pour l'utilisateur ;
- c) les co-occurrences avec d'autres mots et classes de mots (au niveau syntaxique et sémantique) ;
- d) les principaux sens du mot ;
- e) la corrélation entre les sens du mot et les structures syntaxiques.

Il faut remarquer qu'il est essentiel de prêter attention aux divers types de textes, et qu'un bon équilibre dans un corpus de référence est nécessaire car les données de fréquence (à n'importe quel niveau, lexical, syntaxique, sémantique, de collocation, etc. ) peuvent être très différentes selon les types de textes.

Examinons de nouveau le mot *libro* (livre) pour ce qui concerne cette fois un autre type d'informations fourni par les textes. Si nous considérons les verbes liés à *libri* dans le dictionnaire italien, nous pouvons remarquer que ni *leggere* (lire), ni *scrivere* (écrire) ou *pubblicare* (publier), etc., ne se trouvent parmi ces verbes. Une remarque analogue avait, en outre, déjà été faite à propos des dictionnaires anglais [Boguraev *et alii* 1989]), et, comme ceci n'est pas le fait du hasard, on y voit le signe évident de la similitude entre dictionnaires de langue différente. Parmi les définitions de ces verbes, nous trouvons alors d'ordinaire des mots plus génériques liés aux *choses imprimées* telles que *scrittura, parole, segni, lettere, scritto, opera, volume, giornale* (écriture, mots, signes, lettres, texte, œuvre, volume, journal). Le mot *libro* (livre) figure dans quelques exemples, mais la liaison ne pourrait être établie que d'une façon indirecte, étant donné que le mot *libro* est défini à partir de mots tels que *volume, opera, scritti, stampati ...*, c'est-à-dire les mots mêmes qui paraissent dans les définitions de ces verbes.

En revanche, dans notre corpus de textes, ces verbes *leggere, scrivere, pubblicare* sont directement reliés à *libro*. En effet, à partir des 3222 index analytiques du lemme *libro*, on les retrouve dans les contextes de ce terme :

|                   |     |
|-------------------|-----|
| <i>leggere</i>    | 187 |
| <i>scrivere</i>   | 196 |
| <i>pubblicare</i> | 107 |

C'est l'analyse des grands corpus textuels qui permet de repérer ce type d'informations sur la collocation et c'est pourquoi nous utilisons des instruments statistiques/quantitatifs pour faciliter l'obtention semi-automatique de ce type de données et de bien d'autres à partir de notre corpus [Bindi, Calzolari, à paraître].

Quand on analyse un vaste corpus avec des millions de mots dans leur contexte, il est nécessaire, en un certain sens, de découvrir et de décrire :

- les usages qui ne sont pas décrits dans les dictionnaires commerciaux ;
- les fréquences relatives des différents sens des mots, et des différents modèles/structures syntaxiques ;
- et, surtout, les indications grammaticales/syntaxiques grâce auxquelles la désambiguïisation sémantique peut être partiellement obtenue ; étant donné qu'en présence de différents constituants syntaxiques, d'ordinaire, le sens du mot change, mais qu'à l'inverse, nous n'avons pas nécessairement un seul sens pour une même structure syntaxique.

Lorsqu'on recueille ce type de données pour un nombre déterminé de mots, bien souvent, on s'aperçoit que si l'on se conformait à l'usage effectif de la langue, ces données devraient être réorganisées d'une façon différente de la disposition adoptée actuellement par les dictionnaires standard. Aussi, afin d'automatiser la recherche de ce type d'informations directement à partir de corpus, il faut, avant tout, être en mesure de marquer le corpus selon les différentes catégories grammaticales (POS). Pour cette tâche, il existe déjà beaucoup de systèmes [cf. par exemple, Hindle, 1989 ; Webster, Marcus, 1989] et il devrait être possible, même sans un analyseur complet, d'appliquer au corpus du texte des procédures *pattern-matching* (comme celles que nous utilisons actuellement avec les définitions du dictionnaire). Ces procédures *pattern-matching* devraient être explicitement activées pour l'obtention du type de données que nous recherchons (c'est-à-dire phrases prépositionnelles, complétives, propositions infinitives, etc.).

Cette stratégie de recherche des indications syntaxiques (et de collocations) pour la désambiguïisation sémantique (à employer pour des traductions différentes du même mot) est maintenant évaluée dans

des traductions différentes du même mot) est maintenant évaluée dans un projet-pilote que nous mettons en œuvre dans un contexte multilingue.

**La *Workstation* du lexicographe en tant que modèle d'intégration d'instruments et de données à partir de différents milieux et connaissances**

L'importance d'une collaboration entre les chercheurs qui opèrent dans les domaines de CL/NLP et de LLC/TP (comme nous l'avons déjà dit dans la première partie de cette étude) est évidente si nous considérons qu'il est nécessaire de traiter les grands corpus textuels pour obtenir de meilleures LKB. Le projet de ces grandes LKB intégrées peut vraiment devenir l'objectif de projets de coopération, dans lesquels les données *typiques*, les instruments, les procédures, la connaissance, la capacité, les résultats, etc., des deux secteurs de CL/NLP et de LLC/TP *doivent* opérer parallèlement et coopérer pour interagir entre eux.

Pour obtenir au moins quelques-uns des résultats décrits jusqu'à présent, il faut poser les exigences suivantes :

- projet et utilisation d'instruments efficaces ;
- grandes séries de données lexicales et textuelles ;
- systèmes très modulaires ;
- possibilité de partager les ressources, les données et les procédures ;
- une très grande coopération entre les recherches traditionnellement différentes ou les communautés industrielles.

Un modèle du genre d'intégration auquel nous avons pensé est illustré par la station de travail du lexicographe (LW) que nous projetons à Pise [Calzolari, Picchi, Zampolli, 1987]. Elle est conçue comme un système très modulaire, où sont intégrés les différents types de données et de procédures. Au niveau des données, la LW renferme -ou renfermera- une base de données textuelles, une ou plusieurs bases de données lexicales unilingues, une anthologie d'informations taxinomiques, des bases de données lexicales

bilingues, un corpus de référence, tandis qu'au niveau des procédures, elle comporte un instrument morphologique, des analyseurs de dictionnaire, un repère-hyponyme, un système de recherche d'informations, un programme de lemmatisation, une procédure *pattern-matching* pour les définitions du dictionnaire, un instrument de révision, etc.

Cette série de composantes, complexe et variée, reflète notre vision de l'exigence d'une intégration et d'une interaction entre données et moyens traditionnellement pertinents, et concernant CL ou LLC seulement. Il semble donc important de réaliser une coopération efficace entre plusieurs groupes de chercheurs (on entend *groupe* comme *domaine de recherche*), dans le but de lier des mondes qui jusqu'à présent n'ont pas été tellement mis en relation, surtout, à vrai dire, dans la tradition américaine.

### Bibliography

AHLWEDE (T.), EVENS (M.)

1988, "Parsing vs Text Processing in the Analysis of Dictionary Definitions", *Proceedings of the 26th Annual Meeting of the Association of Computational Linguistics*, Buffalo, New-York, p. 217-224.

ALSHAWI (H.)

1989, *Analyzing the Dictionary Definitions*, p.153-170, in [(B.) BOGURAEV, (E.J.)BRISCOE, eds.].

AMSLER (R.A.)

1981, "A Taxonomy for English Nouns and Verbs", *Proceedings of the 19th Annual Meeting of the Association for Computational Linguistics*, Stanford, California, p. 133-138.

ATKINS (B.T.)

1987, "The Uses of Large Text Databases, Semantic ID Tags : Corpus Evidence for Dictionary Senses", *Third Annual Conference of the University of Waterloo Centre for the "New Oxford English Dictionary"*, Waterloo, Canada, p.17-36.

ATKINS (B.T.), KEGL (J.), LEVIN (B.)

1986, "Explicit and Implicit Information in Dictionaries", in *Proceedings of the Conference on Advances in Lexicology*, Waterloo.

BINDI (R.), CALZOLARI (N.)

forthcoming, "Statistical analysis of a large textual Italian Corpus in search of lexical information", presented for EURALEX 1990, Malaga.

BOGURAEV (B.), BRISCOE (E.J.), eds.

1989, *Computational Lexicography for Natural Language Processing*, London, Longman.

BOGURAEV (B.), BRISCOE (E.J.), CALZOLARI (N.) *et al.*

1988, "Acquisition of Lexical Knowledge for Natural Language Processing Systems (AQUILEX)", *Technical Annex, ESPRIT Basic Research Action, N°3030*, Cambridge.

BOGURAEV (B.), BYRD (R.), KLAVANS (J.) *et al.*

1989, "From Structural Analysis of lexical Resources to Semantics in a Lexical Knowledge Database", in *Proceedings of the First International Lexical Acquisition Workshop*, Detroit (Michigan).

BROWN (P.), COCKE (J.), DELLA PIETRA (S.) *et al.*

1988, "A Statistical Approach to Language Translation", *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest.

BYRD (R.J.)

1989, "Discovering Relationships among Word Senses", *Dictionaries in the Electronic Age, Fifth Annual Conference of the University of Waterloo Centre for the "New Oxford English Dictionary"*, Oxford.

BYRD (R.J.), CALZOLARI (N.), CHODOROW (M.) *et al.*

1987, "Tools and methods for Computational Lexicology", *Computational Linguistics*, vol. 13, n° 3-4, p. 219-240.

CALZOLARI (N.)

1982, "Towards the organization of lexical definitions on a database structure", *COLING82*, ed. by E. Hajicova, Prague, Charles University, p.61-64.

1984, "Detecting Patterns in a lexical Database", *Proceedings of the 10th International Conference on Computational Linguistics*, Stanford, California, p. 170-173.

1988, "The Dictionary and the Thesaurus can be combined", in *Relational Models of the Lexicon (Studies in Natural Language Processing series)*, ed. by M. Evens, Cambridge (Mass.), Cambridge Univ. Press, 1988, p. 75-96.

1989a, "Lexical Databases and Text Corpora : perspectives of integration for a lexical Knowledge Base", in *Proceedings of the First International Lexical Acquisition Workshop*, Detroit (Michigan), n°28.

1989b, "Computer-aided Lexicography : Dictionaries and word Databases"; *Computational Linguistics*, ed. by I.S. Batori, W. Lenders, W. Putschke, Berlin, Walter de Gruyter, p. 510-519.

*forthcoming*, "Structure and Access in an Automated Lexicon and Related Issues", [in WALKER (D.), ZAMPOLLI (A.), CALZOLARI (N.), eds].

CALZOLARI (N.), PICCHI (E.)

1986, "A Project for a Bilingual Lexical Database System", *Advances in Lexicology, Second Annual Conference of the UW Centre for the "New Oxford English Dictionary"*, Waterloo, Ontario, p. 79-82.

1988, "Acquisition of Semantic Information from an on-Line Dictionary", *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, p. 87-92.

CALZOLARI (N.), PICCHI (E.), ZAMPOLLI (A.)

1987, "The Use of Computers in Lexicography and Lexicology", in *The Dictionary and the language learner*, ed. by A. Cowie, *Lexicographica Series Maior*, 17, Tübingen, Niemayer, p. 55-77.

CHODOROW (M.S.), BYRD (R.J.), HEIDORN (G.E.)

1985, "Extracting Semantic Hierarchies from a Large Online Dictionary", *Proceedings of the Association for Computational Linguistics*, Chicago, Illinois, p. 299-304.

CHURCH (K.W.)

1988, "A Stochastic Parts Program and Noun Phrase Parser for unrestricted Text", *ACL, Second conference on Applied Natural Language Processing*, p. 136-143.

CHURCH (K.W.), HANKS (P.)

1989, "Word Associations Norms, Mutual Information and Lexicography", *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, British Columbia, p. 76-83.

FOX (E.), NUTTER (T.), AHLWEDE (T.) *et al.*

1988, "Building a large Thesaurus for Information Retrieval", *Proceedings of the Second Conference on Applied Natural Language Processing*, Austin, Texas, p. 101-108.

GOETSCHALCKX (J.), ROLLING (L.), eds.

1982, *Lexicography in the Electronic age*, Amsterdam, North-Holland.

[Gruppo di Pisa]

1979, "Il Dizionario di Macchina dell'Italiano", in *Linguaggi e Formalizzazioni*, ed. by D. Gambarara, F. Lo Piparo, G. Ruggiero, Roma, Bulzoni.

HINDLE (D.)

1988, "Acquiring Disambiguation Rules from Text", *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Morristown (N.J.), p. 118-125.

Japanese Electronic Dictionary Research Institute

1988, *Electronic Dictionary Project*, Tokyo.

KATZ (B.), LEVIN (B.)

1988, "Exploiting Lexical Regularities in Designing Natural Language Systems", *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, p. 316-323.

KAY (M.)

1983, "The Dictionary of the Future and the Future of the Dictionary", in A. Zampolli, A. Campelli, eds., p. 161-174.

KLAVANS (J.L.)

1988, "Building a Computational Lexicon using Machine Readable Dictionaries", paper presented at *The Third Congress of the European Association for Lexicography*, Budapest.

NAGAO (M.), NAKAMURA (J.)

1988, "Extraction of Semantic Information from an Ordinary English Dictionary and its Evaluation", *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, p. 459-464.

NEFF (M.), BOGURAEV (B.)

1989, "Dictionaries, Dictionary Grammars and Dictionary Entry Parsing", *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver (B.C.), p. 91-101.

PICCHI (E.), CALZOLARI (N.)

1986, "Textual Perspectives through an Automatized Lexicon", in *Méthodes quantitatives et informatiques dans l'étude des textes*, Genève, Slatkine, p.705-715.

PICCHI (E.), PETERS (C.), CALZOLARI (N.)

1988, 1990, "A Tool for the Second Language Learner : organizing Bilingual Dictionary Data in interactive Workstation", in *Proceedings of the XX ALLC Conference*, Jerusalem.

PUSTEJOVSKY (J.)

1989, "Current issues in Computational Lexical Semantics", invited lecture, *Proceedings of the Fourth Conference of the European Chapter of the ACL*, Manchester, England, XVII-XXV.

QUEMADA (B.)

1961, *Introduction aux Actes du Colloque international sur la Mécanisation des Recherches Lexicologiques*, Besançon, p. 13-18.

SMADJA (F.)

1989, "Macrocoding the Lexicon with co-occurrence Knowledge", paper presented at the *First Lexical Acquisition Workshop*, Detroit.

TALMY (L.)

1985, "Lexicalization Patterns : Semantic structure in lexical Forms", in T. Shopen, ed., *Language Typology and Syntactic Description : grammatical Categories and the Lexicon*, Cambridge University Press.

THOMPSON (H.)

1989, *Linguistic Corpora for the Language Industry* (Background paper).

VAN der STEEN (G.J.)

1982, "A Treatment of Queries in Large Text Corpora", in S. Johansson, ed., *Computer corpora in English Language Research*, Norwegian Computing Centre for the Humanities, Bergen, p. 49-65.

VOSSEN (P.), MEIJS (W.), den BROEDER (M.)

1989, "Meaning and Structure in Dictionary Definitions", p.171-192, in [BOGURAEV (B.), BRISCOE (E.J.), eds.].

WALKER (D.), ZAMPOLLI (A.), CALZOLARI (N.), eds.

1987a, *Towards a polytheoretical Lexical Database*, Pisa, ILC.

1987b, Special issue of *The Journal of Computational Linguistics*, vol. 13, n° 3-4, p. 193.

*forthcoming*, *Automating the Lexicon : Research and Practice in a Multilingual Environment*, OUP.

WEBSTER (M.), MARCUS (M.)

1989, "Automatic Acquisition of the Lexical Semantics of Verbs from Sentence Frames", in *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, British Columbia, p. 177-184.

WILKS (Y.), FASS (D.), GUO (C.-M.) *et al.*

1989, "A Tractable Machine Dictionary as a Resource for Computational Semantics", p.193-228, in [BOGURAEV (B.), BRISCOE (E.J.), eds.].

ZAMPOLLI (A.)

1968, *Projet pour un lexique electronique de l'italien*, in Busa, ed, p. 109-126.

1983, "Lexicological and lexicographical Activities at the Instituto di Linguistica Computazionale", in A. Zampolli, Cappelli, eds, 1983, p. 237-278.

1986, "Multifunctional Lexical Databases", *Encrages*, n°16, p. 56-65.

