



HAL
open science

Étude d'un modèle computationnel pour la représentation du sens des mots par intégration des relations de contexte

Hyungsuk Ji

► **To cite this version:**

Hyungsuk Ji. Étude d'un modèle computationnel pour la représentation du sens des mots par intégration des relations de contexte. domain_stic.othe. Institut National Polytechnique de Grenoble - INPG, 2004. Français. NNT: . tel-00008384

HAL Id: tel-00008384

<https://theses.hal.science/tel-00008384>

Submitted on 7 Feb 2005

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.

Institut National Polytechnique de Grenoble

N ° attribué par la bibliothèque

| | | | | | | | | |

Thèse

Pour obtenir le grade de

Docteur de L'INPG

Spécialité : Sciences Cognitives

préparée au laboratoire Institut des Sciences Cognitives
dans le cadre de l'**École Doctorale Ingénierie pour la Santé, la
Cognition, l'Environnement**

Présentée et soutenue publiquement

par

Hyungsuk JI

le 16 novembre 2004

**Étude d'un modèle computationnel pour la
représentation du sens des mots par intégration des
relations de contexte**

JURY

M^{me} Hélène Paugam-Moisy, Université de Lyon II (Directeur de thèse)
M^{me} Sabine Ploux, Institut des Sciences Cognitives (Directeur de thèse)
M. Bruno Bachimont, Université de Technologie de Compiègne (Rapporteur)
M. Benoît Habert, Université de Paris X (Rapporteur)
M^{me} Catherine Garbay, Laboratoire TIMC-IMAG (Examineur)
M. Eric Wehrli, Université de Genève (Examineur)
M. Guy Denhière, Université de Provence (Examineur)

Résumé

Dans cette thèse nous présentons une approche théorique du concept et un modèle linguistico-informatique. Cette théorie, non définitionnelle, est fondée sur une représentation gaussienne du concept. Nous introduisons le terme « contexonyme », une formalisation de la relation de contexte entre les mots. Cette notion lie la théorie du concept au modèle informatique. Basé sur ces deux notions, notre modèle informatique apprend des contexonymes de manière automatique à partir de corpus de taille importante non annotés. Pour chaque mot donné, le modèle propose la liste de ses contexonymes et les organise par une méthode de classification hiérarchique. Les contexonymes ainsi obtenus reflètent des connaissances encyclopédiques ainsi que diverses caractéristiques langagières comme l'usage des mots ou encore les fines différences sémantiques entre synonymes. Les résultats sur des tests montrent que le modèle peut être utilisé pour des tâches de TAL ainsi que comme ressource lexicale dynamique.

Mots clefs : contexonyme, concept, traitement automatique des langues (TAL), corpus, contexte, représentation sémantique, mot lié contextuellement, apprentissage automatique.

Abstract

In this thesis, we present a theory of concepts and a related computational linguistic model. This non-definitinal theory proposes a view that considers concepts as a Gaussian representation. The term *contexonym*, formalized version for ‘contextually related words’, is introduced that relates this theory of concepts and computational model. Based on these two principles, our computational model learns automatically contexonyms from a very large untagged corpus. For a given word, the model proposes a set of its contexonyms and organizes them by a hierarchical clustering method. The contexonyms thus obtained reflect encyclopedic knowledge in addition to various linguistic features like word usage or subtle difference between near-synonyms. The results on different tests show that the model could be used for natural language processing (NLP) tasks as well as a dynamic lexical reference.

Keywords: contexonym, concepts, natural language processing (NLP), corpus, context, semantic representation, contextually related words, machine learning.

Remerciements

Je remercie tout d'abord M^{me} Sabine Ploux de m'avoir accueilli dans son équipe, d'avoir dirigé ma thèse et de m'avoir accordé une grande autonomie pendant ces dernières années.

Je tiens à remercier également ma directrice de thèse, M^{me} Hélène Paugam-Moisy qui m'a soutenu et qui m'a aidé à organiser mon travail lors de ma thèse.

Je suis très reconnaissant à M. Eric Wehrli qui m'a accueilli dans son laboratoire LATL à Genève.

Je voudrais remercier les membres de mon jury pour l'intérêt qu'ils ont témoigné à mon travail, particulièrement M. Benoît Habert et M. Bruno Bachimont qui m'ont donné des commentaires précieux.

Un grand merci à M^{me} Anne Reboul de m'avoir soutenu et de m'avoir encouragé.

Merci à mes amis au laboratoire et ailleurs qui m'ont encouragé et aidé pour finir ma thèse : Jean-Marc Blanc, Christelle Dodane, Flavie Martin, Emanuelle Reynaud, David Meunier, Bernard Jacquemin, Olivier Castéra, Valérie Buron, Chanwoong Lee et Sohyun Ahn, parmi les autres.

Enfin, mes remerciements profonds s'adressent à mes parents et mon épouse Eunkyong pour leur soutien inconditionnel et leur encouragement toujours chaleureux.

Table des matières

1	Introduction	1
2	Concept et sens	5
2.1	Etendue du concept	7
2.1.1	Entité mentale contre entité abstraite	10
2.2	Sens et référence	13
2.2.1	Les noms propres et leurs sens	13
2.2.2	Contenu du concept	16
2.3	Théorie du concept	24
2.3.1	Théories diverses	24
2.3.2	Théorie proposée	31
3	Sens et contexonyme	45
3.1	Langage formel et langue	48
3.2	Une critique relative aux propositions de Pustejovsky et Jackendoff	51
3.2.1	Le problème renversé	56
3.3	Polysémie contre homonymie	58
3.4	Proposition	60
3.4.1	Lien contextuel	62
3.4.2	Contexonyme	66

4	Les modèles informatiques	69
4.1	Référence lexicale	71
4.2	Les modèles d'organisation du sens des mots	74
4.2.1	Le modèle de Ploux	76
4.3	Les modèles utilisables dans une tâche de TAL	77
4.4	Les modèles capables d'apprendre	82
4.5	Proposition	84
5	Modèle automatique d'organisation des contexonymes (ACOM)	91
5.1	Procédure d'organisation du contexonyme	93
5.2	Caractéristique du modèle	101
6	Test sur des exemples	105
6.1	Test sur des mots choisis de manière aléatoire	105
6.2	Test sur les exemples d'Edmonds et Hirst	107
6.3	Test sur les exemples de Dagan et Itai	110
6.4	Test avec la méthode de fusion	110
6.5	Test sur les exemples de Hirsh et Tree	116
6.6	Test sur deux corpus différents	117
7	Discussion et conclusion	121
7.1	Une étude pluridisciplinaire	121
7.2	Modélisation du concept	124
7.3	Modèle sémantique	132
7.4	Modèle informatique	136
7.5	Conclusion et perspective	139
A	Résultats du test sur des mots choisis de manière aléatoire	143

TABLE DES MATIÈRES

ix

B Résultats du test sur les exemples d'Edmonds et Hirst

149

C Test sur les résultats de Hirsh et Tree

155

Table des figures

2.1	Profil épistémologique de la notion personnelle de masse.	32
5.1	Rerésentation du mot <i>maigre</i> et son équivalent anglais.	92
5.2	Certains contexonymes de <i>match</i>	98
5.3	Classification des contexonymes de <i>match</i>	99
5.4	Représentation de <i>conductor</i>	102
6.1	Fusion de <i>wins</i> et <i>match</i>	115
7.1	Représentation schématisée du concept C dans la langue A et du concept D dans la langue B. a, b, c et d sont les concepts de bases.	125
7.2	La région correspondant à <i>tickets</i> dans la représentation de <i>conduc- tor</i>	127

Liste des tableaux

5.1	Tableau d'un candidat du contexonyme.	95
5.2	Deuxième tableau du contexonyme.	96
6.1	Résultats du test sur les exemples d'Edmonds et Hirst.	108
6.2	Test sur les exemples de Dagan et Itai.	111
6.3	L'effet de fusion pour <i>match</i> et ses voisins.	113
6.4	Résultats du test sur les exemples de Hirsh et Tree.	116
6.5	Résultat du test pour deux corpus séparés (<i>Le Monde</i> et <i>L'Humanité</i>).	118
A.1	Résultats du test sur des mots choisis de manière aléatoire.	143
B.1	Résultats du test sur les exemples d'Edmonds et Hirst.	149
C.1	Les résultats du modèle ACOM sur les exemples de Hirsh et Tree.	155

Chapitre 1

Introduction

Depuis la machine de Turing, l'espoir de développer un système aussi intelligent que les êtres humains s'est répandu dans plusieurs disciplines. L'idée d'un système capable de comprendre le langage humain est un exemple de ce type de rêve. La naissance des sciences cognitives est aussi très étroitement liée à cette tendance. En l'absence de résultats impressionnants et en raison du développement à pas menus de la recherche sur le langage humain, l'intérêt s'est vite tourné vers l'étude des aptitudes humaines de plus bas niveau, à savoir les perceptions telles que la vision, l'audition, la motricité, etc. Certains auteurs déclarent même que l'intelligence se trouve non pas dans les fonctions de haut niveau comme le langage, mais plutôt dans les fonctions de bas niveau comme les perceptions.

On peut trouver plusieurs explications à la lenteur du développement d'un système capable de comprendre le langage humain. Une chose évidente est que cette lenteur n'est pas due à une lenteur de développement au sein des techniques matérielles. La vitesse de calcul d'un ordinateur, par exemple, double tous les 18 mois environ, ce qui constitue une vraie révolution. Certes, l'ordinateur le plus rapide du monde n'a pas encore atteint la vitesse de calcul du cerveau humain, mais cet écart est susceptible d'être réduit, voire même dépassé dans un délai très

court.

Cette constatation laisse donc penser que ce retard n'est pas dû au matériel, mais plutôt à la méthode, au modèle et à la théorie. De fait, il existe de nombreux débats scientifiques concernant les capacités les plus remarquables chez l'être humain comme le langage et la cognition. Toutefois l'un de ces éléments n'a pas été explorés de manière exhaustive : il s'agit des relations entre les concepts ou entre les sens des mots, relations tellement évidentes qu'elles paraissent même *banales*.

A partir d'un mot donné, nous sommes tous capables de produire les mots ou les concepts qui lui sont associés. Si l'on entend par exemple le mot *neige*, on a tendance à imaginer des mots comme *hiver, ski, froid, blanc*, etc. plutôt que *chaise, pomme de terre, opéra*, etc. Réciproquement, si l'on rencontre des mots anglais comme *final, wins, Agassi* suivis du mot anglais *match*, on interprétera ce dernier comme *match de sport* plutôt que comme *allumette*.

La tendance générale de ce type d'association de mots est aussi constatée dans des études de psychologie cognitive. Toutefois, pour la machine, ce type de relation entre les mots ou les concepts n'est pas évident. En effet, les systèmes de traduction automatique des langues les plus connus proposent le mot cible français *allumette* malgré la présence des mots *final, wins, Agassi*, etc. D'autres exemples de lien contextuel tel que *rédiger – article* et *tituber – ivre* mettent en évidence la différence entre l'homme et la machine.

Dans cette thèse, nous présentons une étude sur le concept et le sens des mots ainsi qu'un modèle informatico-linguistique qui pourrait répondre aux questions posées ci-dessus. Le sens des mots et le concept concernent plusieurs domaines de recherche et ce sont des sujets à la fois théorique et pratique. Cette nature pluridisciplinaire se retrouvera aussi dans notre étude.

Dans le chapitre deux, nous examinerons les études philosophiques menées sur le concept lexical et nous présenterons notre théorie du concept.

Ensuite, dans le chapitre trois, nous examinerons des théories du sens des mots puis nous introduirons la notion de *contexonyme* qui s'avérera être une formalisation de la notion de « mots pertinents liés de manière contextuelle ».

Dans le chapitre quatre, nous examinerons les modèles informatiques liés aux sujets du concept et du sens des mots sous quatre critères que nous avons établis pour les modèles computationnels.

Dans les chapitres cinq et six nous présenterons notre modèle informaticolinguistique baptisé « Modèle automatique d'organisation des contexonymes (Automatic Contexonym Organizing Model : ACOM). La discussion et les perspectives feront l'objet du chapitre sept.

Chapitre 2

Concept et sens

Parmi les caractéristiques qui séparent l'homme de l'animal, le langage humain est sans doute le plus marqué. L'existence d'une part d'environ six mille langues différentes sur la terre et le fait d'autre part que les gens réussissent à communiquer en utilisant leur langue malgré les ambiguïtés qui y sont présentes posent la question de savoir s'il existe une entité plus fondamentale que le langage humain.

En tant qu'élément de construction de la cognition, le concept est étroitement lié à cette question. En effet le concept est un élément fondamental pour plusieurs disciplines de recherche et, à partir du concept, de nombreux phénomènes peuvent s'expliquer et de nombreuses connaissances peuvent être construites. De ce fait, la notion de concept se positionne fréquemment au centre des débats dans nombre de disciplines comme les sciences cognitives, la philosophie, la psychologie et la linguistique ; le concept est aussi un élément central pour le traitement automatique des langues (TAL) et, plus généralement, en intelligence artificielle (IA). Le fait que, d'une part, il existe depuis toujours des controverses sur ce thème dans les domaines concernés et que, d'autre part, aucune machine intelligente n'ait jamais vu le jour malgré des démarches courageuses initialisées par Turing, exige

une révision fondamentale de la discussion portant sur le concept, non seulement en ce qui concerne la notion en elle-même, mais aussi en faveur de son application pratique.

Quant à la relation entre le concept et le langage humain, il existe des débats très vifs autour de ce sujet. D'un côté, il existe des théories selon lesquelles le concept est influencé par la langue que l'on utilise (Whorf 1956, Sapir 1929, Dennett 1996). D'un autre côté, d'autres théories réclament que le concept existe avant le langage humain et que le concept ne soit pas influencé par le langage général (Fodor 1975) ou bien une langue spécifique (Vygotsky 1930/1978, Vygotsky 1934/1986). De nombreux résultats expérimentaux sur ce sujet préconisent tantôt la première vue tantôt la seconde. Par exemple, les différentes représentations du concept chez les sujets qui utilisent différentes langues valident la première position tandis que la constatation de l'existence du concept chez les jeunes enfants n'ayant pas encore acquis le langage appuie la deuxième hypothèse.

D'autre part, l'étendue du concept varie très significativement selon les auteurs : alors que le concept de Putnam (1975b) exclut l'extension, celui de Fodor (1998) l'inclut. Par exemple, pour Putnam, le concept EAU n'inclut pas l'extension H₂O tandis que pour Fodor il le fait. En outre, certains auteurs comme Jackendoff supposent qu'il existe deux types de concepts — concept I et concept E (Jackendoff 1990).

Ainsi, comme le dit Frege (1892/1997), le sens du mot *concept* est tantôt psychologique, tantôt logique et parfois même un mélange des deux notions. De ce fait, avant de discuter du concept et de sa théorie, il est primordial de bien cerner la notion de concept puisqu'elle varie selon les auteurs et les disciplines concernées. Dans cette étude, nous nous concentrerons sur le concept lexical. Pour faciliter la lecture nous utiliserons le terme *concept* à la place du terme *concept lexical*.

2.1 Etendue du concept

L'étendue du concept varie de façon significative selon les points de vue des auteurs. Par exemple, alors que Descartes ne distingue pas explicitement le concept et l'idée, Frege (1891/1997, 1892/1997, 1892/1952) distingue explicitement ces deux notions. Chez Frege, tandis que l'idée est « subjective », le concept est « une fonction d'un argument, dont la valeur est toujours une valeur de vérité » (Frege 1891/1997). Ainsi, Frege rend le concept plus « objectif », laissant l'idée comme subjective séparant ainsi les deux notions. En effet, chez Frege, ainsi que chez Carnap, cette séparation est fondée sur la préconisation de l'entité abstraite contre l'entité mentale : en effet, pour eux le soutien à cette dernière est qualifié de « psychologisme ». C'est pour cela que Frege affirme que l'usage du concept est « purement logique » (Frege 1892/1997).

Cette distinction est moins explicite chez Putnam puisqu'il qualifie le concept de « notion vague » ainsi que l'intension et préfère utiliser le terme d'*extension* qui est essentiellement la même notion que le concept frégeen. Putnam propose une modification de la définition traditionnelle de l'extension comme « l'ensemble des choses pour lesquelles le terme est valide (the set of things the term is true of) » : il propose d'utiliser *terme et un sens qui ont une extension* au lieu de *terme* tout court et *sous-ensemble flou* au lieu d'*ensemble* (Putnam 1975b). De ce fait, lorsque Putnam utilise le terme *concept*, ce terme signifie au fond l'*idée* de la terminologie de Frege et le terme *concept* chez Frege peut être interprété dans un sens plus large que celui de Putnam. Afin d'éviter la confusion, appelons l'*idée chez Frege* C_{fr} , le *concept chez Putnam* C_p et l'*extension chez Putnam* l'extension.

En adoptant la distinction entre langage E et langage I de Chomsky, Jackendoff distingue le concept E et le concept I (pensée) (Jackendoff 1990). Chomsky emploie langage E (externalized language, E-language) en ce sens que « la construction est comprise indépendamment des propriétés de l'esprit/cerveau » ;

en suivant la « notion de structure » de Jespersen, Chomsky introduit le langage I (internalized language, I-language) qui correspond à « quelques éléments de l'esprit de la personne qui connaît la langue acquise par l'apprenant et utilisée par le locuteur-interlocuteur » (Chomsky 1986). Comme l'admet Jackendoff, la terminologie du concept E et du concept I provient de la terminologie de Chomsky. En gros, le concept E correspond au concept chez Frege ou à l'extension chez Putnam et le concept I à l'idée chez Frege et au concept chez Putnam.

En ce sens, on peut remarquer l'analogie à ces deux concepts des notions de « contenu étroit », qui correspond à la représentation mentale du monde chez un individu indépendant de son environnement, et de « contenu large », qui exige la dépendance ou le rôle du monde extérieur (environnement).¹

Chez Fodor, cette séparation ne se rencontre pas, et il sera permis de juger que le concept chez Fodor comprend les deux parties. Appelons C_{fd} *le concept chez Fodor*.

Frege met le « sens » entre le C_{fr} et la référence. Il introduit le « mode de présentation » qui distingue deux *sens* différents ayant la même référence. Par exemple, si trois lignes a, b et c se rejoignent pour former un unique point d'intersection, il existe au moins deux *modes de présentation* tels que « point d'intersection de a et b » ou « point d'intersection b et c » ce qui désigne la même référence, soit le point d'intersection commun (Frege 1892/1952). La raison pour laquelle Frege sépare le C_{fr} du sens est que le sens, comme dans l'exemple ci-dessus, est d'après Frege, indépendant de l'individu et réside à l'extérieur des individus, ce qui en garantit sa nature partagée.

Quant au sens, Putnam propose quatre distinctions concernant la description

¹Il existe néanmoins une tension entre Jackendoff et Chomsky, car Chomsky est contre la distinction entre le contenu étroit et le contenu large (Chomsky 1995) et il n'est pas d'accord avec l'interprétation de Jackendoff de ses langage E et langage I (Jackendoff 1990).

du sens d'un mot. Par exemple, pour le mot *eau*, il propose le suivant :

1. marqueurs syntaxiques tels que *nom, concret*
2. marqueurs sémantiques tels que *chose naturelle, liquide*
3. stéréotype tel que *sans couleur, transparent, insipide, désaltérant, etc.*
4. extension tel que H_2O .

Ainsi, pour Putnam les descriptions 1, 2 et 3 sont des intensions de l'eau mais non la description 4. Cette séparation est introduite pour résoudre le problème de la contradiction avancé par Engels : bien que les poissons aient pour définition « respirer par des branchies », on a découvert des poissons qui possèdent des poumons, ce qui contredit cette définition (Engels 1895, Putnam 1975a). Une fois faite la séparation entre l'extension, comme « respirer par les branchies », et l'intension, d'après Putnam, il n'existe plus de contradiction.

Fodor n'accepte pas cette séparation et il considère le concept C_{fd} comme étant ce qui comprend les deux parties séparées par Frege et Putnam. Ainsi, C_{fd} est équivalent à : soit $C_{fr} + \text{sens}$, soit $C_p + \text{l'extension}$. Plus précisément, Fodor garde $C_{fr} + \text{sens}$ comme C_{fd} , mais il sépare le mode de présentation de C_{fd} . Cette explication suppose que le sens et le mode de présentation ne sont pas la même chose. Si les deux sont identiques, alors C_{fd} devient la même chose que C_{fr} , mais ce n'est pas le cas, comme nous le verrons plus tard.

Fodor n'accepte pas que le mode de présentation réside « en dehors de la tête ». Fodor dit que « si les modes de présentation sont capables d'individualiser les concepts et si les référents n'en sont pas capables, c'est parce que les modes de présentation *sont des objets mentaux* tandis que les référents n'en sont pas » (Fodor 1998). En effet, cette position provient du fait qu'il préconise la psychologie partagée. Fodor exclut de la psychologie cognitive l'approche de la capacité ou le rôle inférentiel comme Stich (1990), car une fois cette position prise, il faudra accepter une différence entre individus, ce qui contredit un principe partagé.

Fodor ne peut pas non plus intégrer le mode de présentation au sein du concept car, dans ce cas, la correspondance entre une référence unique et deux modes de présentation distincts est contradictoire avec le principe d'un état mental unique : autrement dit, une telle intégration permettrait l'émergence de l'individualisme. Afin d'éviter cette contradiction, Fodor identifie le mode de présentation à un véhicule dépourvu de sémantique. Cette solution donne néanmoins naissance à un autre problème que Fodor tente d'éviter : si ce véhicule est une chose différente de la sémantique et s'il distingue une différence sémantique entre les individus, ce véhicule n'est plus alors autre chose que la capacité ou le rôle de l'inférence, comme l'a fait remarquer Aydede (1998).

Fodor critique Frege comme disciple du platonisme à cause du fait que Frege met le sens à l'extérieur de la *tête*. D'une part, Fodor peut avoir raison, en ce sens que le sens n'existe pas en tant que tel et qu'il provient de nous-même. D'autre part, Fodor peut avoir tort en ce sens que le sens se forme d'une certaine manière, *indépendamment* des individus. Par *indépendamment* nous entendons l'effet global ou collectif qui résiste en dépit de la variété des individus et non une conséquence de l'externalisme.

2.1.1 Entité mentale contre entité abstraite

En fait, la différence de position vis-à-vis du concept est étroitement liée à la position que l'on adopte concernant la nature du concept, soit comme entité mentale soit comme entité abstraite.

Au début, le concept a été très étroitement lié au subjectivisme et à l'état mental mais cette position a été critiquée par Frege qui considère le concept comme une entité abstraite distincte d'une entité mentale. Cette position sur l'entité abstraite est critiquée de nos jours par les études en psychologie cognitive et en sciences cognitives. Ainsi, en première approximation, la position psychologique

ou cognitive préconise l'entité mentale tandis que la position logique privilégie l'entité abstraite.

Fondamentalement, cette différence a pour origine une différence de position vis-à-vis de l'interprétation du sujet et du monde. De ce fait, nous allons brièvement examiner la distinction historique entre le concept et l'objet ou bien entre l'idée et le monde.

Chaque débat que ce soit celui du dualisme contre le monisme, de l'idéalisme contre le matérialisme, de l'entité mentale contre l'entité abstraite, de l'externalisme contre l'internalisme (individualisme), a sa propre série d'arguments mais ces débats peuvent s'organiser selon certains critères qui relèvent d'un problème fondamental sous-jacent à ces débats. Pour ce faire, on peut isoler d'abord, au sein des débats historiques, une distinction entre deux types d'entités différentes : l'Idée, l'âme, la pensée et l'esprit, d'un côté et le corps, la référence, la chose, la nature et la matière, de l'autre. Bien que tous ces appariements ne soient pas tout à fait homogènes, il existe néanmoins des différences essentielles communes comme l'abstrait contre le concret, l'immatériel contre le matériel, etc. Nous adoptons cette distinction comme point de départ pour la discussion qui suit, mais sans aucune intension de justifier ou de favoriser le dualisme.

La différence de position vis-à-vis de la nature du concept, à savoir l'entité mentale contre l'entité abstraite, ou vis-à-vis de l'étendue du concept, est liée au fait que l'on distingue ou non la propriété ou bien la substance en philosophie.

La distinction entre le concept et le monde (appelons-la « Deux Entités ») se manifeste très clairement par le dualisme qui oppose ces deux entités. Dans la philosophie grecque, il existait un monisme comme le principe de Parménide qui déclare « tout est en un » (Palmer 1999). Platon distingue les Deux Entités de telle sorte que les Deux Entités deviennent isolées.

Chez Descartes la substance est divisée en deux, à savoir le corps et l'âme, et

cette dernière est considérée comme essentielle alors que Spinoza considère que la « moitié de l'âme » est analogue à la « moitié d'un animal » (Bennett 2001). La dichotomie chez Descartes s'applique aussi aux propriétés qui se divisent en « mentalité - idée » et « matérialité - extension ». Pour Descartes et Spinoza, il existe une barrière insurmontable entre l'idée et l'extension c'est-à-dire, que l'on ne peut pas déduire de conclusion de l'une en partant de l'autre (l'idée et l'extension).

Le dualisme cartésien a pour but encore plus explicite de séparer les Deux Entités qui avaient été réunies, injustement pour Descartes, par les stoïciens monistes. Le dualisme platonicien, de même que le dualisme cartésien, garde au sein de sa théorie un problème inévitable : si le concept et le monde sont ainsi séparés, la communication ou l'interaction entre ces deux entités est en principe peu probable, voire impossible. En effet, Kant a fait remarquer que la validité des règles qui sous-tendent les principes platoniciens « est détruite par les idées puisque la nature de l'Idée platonicienne, à savoir l'archétype des choses elles-mêmes, ne découle nullement des sensations, autrement dit il n'existe pas de moyen par lequel l'Idée puisse valider des règles de jugement (Kant 1781/1929). Kant s'oppose aussi à Descartes sur la question de l'idéalisme matériel en considérant « comme douteux et indémontrable » l'existence d'objets dans l'espace indépendamment de la présence humaine.

Bien que cette séparation soit en général considérée comme la caractéristique principale du dualisme, le dualisme appelle une autre question. Si, en apparence l'opposition entre dualisme et monisme réside dans la question de la séparation entre concept et monde, l'essentiel du débat concerne aussi la relation entre les Deux Entités. L'interprétation la plus naïve du monde ou du concept est de voir chaque entité en tant que telle sans envisager de liens entre elles. L'idée des éléments de base d'Empédocle ou d'Aristote correspond en effet à un effort de réunir

ces choses qui, à première vue, sont séparées. Dans cette perspective, le dualisme de Platon est perçu comme une tentative, non seulement de séparer les entités en question, mais aussi d'identifier les relations entre ces entités, différentes par nature, afin de les regrouper et de les réunir. Si l'on adopte ce point de vue, la différence entre Platon et Aristote n'est alors pas si grande : en effet pour Descartes, cette différence est tout simplement une position renversée (Descartes 1644/1953).

Sous l'angle de la relation entre les deux entités, celle d'idée et celle d'extension, le retour du dualisme chez Descartes est en effet une « négation de la négation », d'après la terminologie hégélienne. Autrement dit, il s'agit d'une amélioration plutôt que d'une simple répétition : alors que le dualisme platonicien est basé sur une construction libre de même que le vol d'une colombe est libre, pour reprendre la comparaison de Kant, le dualisme cartésien est basé sur une construction plus rigoureuse. A partir de Descartes, comme le remarque Hegel, la philosophie est devenue indépendante basée sur la raison et la conscience de soi (Hegel 1806/1995) et le dualisme de Descartes était une tentative pour repousser la limite de la métaphysique — bien qu'il ne l'ait pas suffisamment repoussée, de l'avis de Leibniz (Deleuze 1988).

2.2 Sens et référence

2.2.1 Les noms propres et leurs sens

Mill déclare que les noms propres ne sont pas connotatifs (Mill 1843/1988) :

Les noms propres ne sont pas connotatifs ; ils désignent les individus, mais ils n'affirment pas, n'impliquent pas des attributs appartenant à ces individus.

Mill avance cette idée en discutant du nom d'une ville qui s'appelle *Dartmouth*. A l'origine cette ville a été baptisée ainsi parce qu'elle est située à l'embouchure du fleuve *Dart*. Mais, même si l'embouchure est obstruée ou si cette rivière disparaît par un changement brutal de la nature, le nom de cette ville restera toujours *Dartmouth*. D'où la déclaration de Mill : « Les noms propres sont attachés aux objets eux-mêmes et ne dépendent pas de la permanence de tel ou tel attribut ». Ainsi, distingue-t-il la « signification du nom » et celle du nom propre.

Pour Mill, un nom comme *Darthmouth* n'est qu'une « marque insignifiante » et « nous ne disons rien de ces choses » (Mill 1843/1988). En revanche, pour Mill, les noms communs, comme ceux qui figurent dans la proposition « bâtie en marbre », ont « des marques significatives » qui pourront accompagner « un attribut ».

Il reste à discuter cependant s'il existe vraiment une différence entre « bâtie en marbre » et « bâtie à Dartmouth ». Pour Mill ce premier est « une sorte de livrée dont l'attribut revêt tous les objets auxquels il appartient ». Certes, une phrase telle que « Ils vont construire une tour dans le centre de Leeds » n'évoque pas des informations significatives — au moins pour les gens qui n'habitent pas en Angleterre : elle n'évoque pas autant d'information que « bâtie en marbre » qui, quant à elle fait penser aux images d'une ville couverte de marbre ainsi qu'à ses couleurs ou sa qualité de matériau. Mais si on entend la phrase « Ils vont construire une tour dans le centre de Paris », elle évoque des informations non négligeables. Par exemple, l'image de Paris avec cette nouvelle tour entourée par d'autres monuments est sûrement une information significative.

Contrairement à Mill qui s'arrête ici en confirmant que certains noms propres, comme *Socrate*, sont connotatifs, Kripke étend le trait non connotatif à ce type de noms propres et aux choses naturelles. A titre d'exemple, « Aristote » ne peut pas avoir de sens, d'après Kripke, car si jamais un sens comme « Aristote fut le

professeur d'Alexandre le Grand » s'avère faux, il faudra nier, par conséquent, l'existence même d'Aristote (Kripke 1972). Donc, pour éviter cette contradiction, Kripke pose qu'il est indispensable d'ôter la notion de sens aux noms propres. Wittgenstein avait déjà constaté ce genre de problème. Wittgenstein pose la question suivante (Wittgenstein 1953/1972) :

Consider this example. If one says 'Moses did not exist', this may mean various things. It may mean: the Israelites did not have a single leader when they withdrew from Egypt — or: their leader was not called Moses — or: there cannot have been anyone who accomplished all that the Bible relates of Moses — But when I make a statement about Moses, — am I always ready to substitute some one of those descriptions for 'Moses'? I shall perhaps say: by 'Moses', or at any rate, a good deal of it. But how much? Have I decided how much must be proved false for me to give up my proposition as false? Has the name 'Moses' got a fixed and unequivocal use for me in all possible cases?

En outre, contrairement à Mill, Kripke étend la portée de la dénotation des noms propres aux choses naturelles. Kripke déclare que les choses naturelles, comme les noms propres, n'ont pas de descriptions. Par exemple, Kripke dit que la référence d'un tigre ne changera pas en dépit d'une éventuelle découverte affirmant que les tigres ne possèdent que trois pattes. Le principe que nous venons d'introduire afin de montrer que les noms propres ont des descriptions /connotations / sens s'appliquera de la même manière à cette catégorie : si cette découverte est réservée à un nombre de spécialistes très restreint, elle n'a aucun effet pour le concept/mot « tigre » ; si la découverte est connue de tout le monde, la propriété « avoir quatre pattes » disparaîtra au fur et à mesure de la définition du concept.

Dans un sens, « tigre à quatre pattes » périt et un nouveau concept de tigre à savoir « tigre à trois pattes » émerge. En d'autres termes, même si l'on utilise le même mot pour « tigre », le concept de tigre évolue même si le changement est progressif et difficile à saisir. Dans la terminologie de Hegel, on passe d'un changement quantitatif à un changement qualitatif pour cette transition, le cas échéant, de tigre à quatre pattes vers tigre à trois pattes et ce dernier dénote une différence très claire.

2.2.2 Contenu du concept

Exemple de Kripke

Kripke introduit un monde imaginaire afin de réfuter l'idée que la référence dépend du sens. Par exemple, en ce qui concerne un monde imaginaire où les gens identifieraient l'objet et la couleur non pas à l'aide de la lumière mais à l'aide du son, Kripke écrit (1972) :

Would we say that in such a possible world, it was sound which was light, that these wave motions in the air were light? It seems to me that, given our concept of light, we should describe the situation differently. [...] The way we identified light *fixed a reference*.

Ainsi, Kripke maintient que le sens d'un terme ne peut pas être décidé, comme les noms propres, par une « connotation » ou un « sens de Frege ». A notre avis, dans ce monde imaginaire où le son a le rôle de la lumière, le son est en fait la lumière. Car, ce n'est pas la vérité scientifique qui forme le sens d'un terme mais c'est l'usage d'un terme qui forme son sens. C'est-à-dire, de même que les phrases « Je l'ai convaincu » et « Je l'ai persuadé » sont traduites en anglais par « *I persuaded him* », la phrase « Il fait sombre. Il n'y a pas de lumière » serait traduite

dans ce monde imaginaire par « Il fait sombre. Il n'y a pas de son ». Ce genre de problème sera discuté de façon plus exhaustive en utilisant d'autres exemples.

La manière d'examiner ces problèmes est aussi liée aux conditions de vérité, puisque l'argument de Kripke et de Putnam est fondé sur ces conditions. Prenons le concept « temps » pour aborder ce problème. Appelons Tc le temps du XIX^e siècle et Tr temps relatif connu par Einstein. Le fait que le temps soit absolu est vrai pour Tc et faux pour Tr. Puisqu'ils se distinguent ainsi, on peut baptiser Tr d'un nom quelconque afin de le distinguer du « temps (Tc) » : donnons à Tr le nom de « cogtens ».

Avant Einstein, il n'était même pas nécessaire de penser « cogtens », puisque le concept « cogtens (Tr) » n'existait pas. Après Einstein, cependant, deux choix s'offrent à nous. Ou bien on arrête d'utiliser le terme « temps » et on adopte le nouveau terme « cogtens », ou bien on garde le terme « temps ». Un exemple historique du premier choix est « roi » et « président ». On aurait pu garder le terme « roi » en changeant sa propriété de « roi : chef d'état au pouvoir souverain à vie par voie héréditaire » vers « roi : chef d'état élu par le peuple » mais un nouveau terme a été sélectionné. Pour le deuxième choix, on peut trouver de nombreux exemples comme « masse », « univers », « soleil », etc. En effet, c'est un problème de choix et non un problème prédéfini par principe.

Appliquons cette idée au problème d'Aristote. On sait qu'il existait un philosophe grec, disciple de Platon qui a écrit *Éthique à Nicomaque*, *Politique*, *Physique et Métaphysique*. Appelons « Ar » ce philosophe et « Ac » un homme qui a réellement vécu en Grèce au quatrième siècle. Jusqu'à aujourd'hui, « Ar » et « Ac » sont connus comme étant un même personnage, soit Aristote. Maintenant, supposons qu'une grande découverte relève que « Ar » et « Ac » ne sont pas la même personne. Dès lors, il existe deux choix : ou bien appeler « Ac » Aristote et donner à « Ar » un autre nom comme « Sapienteles », ou bien appeler « Ac »

anonyme et « Ar » Aristote. On imagine facilement que le deuxième choix sera accepté puisque cela nous évitera d'apprendre nouveau mot *Sapienteles*, de la même manière que pour les termes temps et *cogtens*.

Maintenant imaginons deux individus ordinaires dans un pays ayant une longue histoire. L'un s'appelle *Decipicuros* et l'autre *Gnarelius*. Un historien sait que *Decipicuros* a écrit *Praecepkia* et que *Gnarelius* n'a rien laissé. Puis, un jour, il découvre que *Decipicuros* n'a rien laissé et que c'est *Gnarelius* qui a écrit *Praecepkia*. Il existe toujours deux choix, de même que précédemment, mais on suppose facilement que cet historien choisira la première solution contrairement au cas d'Aristote exposé ci-dessus.

Pourquoi y a-t-il une différence ? La réponse est relativement simple : dans le cas d'Aristote, le sens d'Aristote a été répandu et est devenu partagé, contrairement au cas de *Decipicuros*. Autrement dit, dans le cas d'Aristote, nous avons déjà associé de nombreux sens à ce nom, mais pour *Decipicuros* cette association est négligeable. C'est pour cette raison que l'on garde les termes *temps*, *univers*, *soleil*, *masse*, etc., bien que leurs sens aient changé très significativement. Il faudra noter néanmoins que cette association n'est pas la condition nécessaire de la décision de garder ou de changer le terme. Si le changement du concept se propage très rapidement et globalement, comme ce fut le cas pendant la révolution française, un terme comme roi change, malgré l'association importante liée à ce terme/concept.

En effet, même s'il pense que les noms propres n'ont pas de sens, Ziff (1960) attribue une autre caractéristique aux mots ayant des traits historiques qui peuvent être décidés par des descriptions associées.

Nous ne pensons pas que ce choix se fasse de façon discrète et par rupture, mais plutôt de façon graduelle : le changement se fait au regard de la capacité ou du volume de sens.

Exemple de Putnam et Burge

Putnam a présenté une idée analogue. Il suppose un monde imaginaire qui est une copie parfaite de notre monde : on suppose qu'il existe sur la terre un homme qui s'appelle Oscar, et qu'il existe aussi un Oscar, sur une Terre Jumelle, qui réagit et pense la même chose que l'Oscar de notre monde. Un jour, on découvre que l'eau de la Terre Jumelle n'est pas composée de H₂O mais de XYZ. Putnam déclare que les deux entités différentes ici (H₂O et XYZ) peuvent correspondre, chez les deux Oscars à un état mental identique.

L'argument de Putnam vise en effet précisément à « détruire la théorie des représentations sémantiques innées ». Pour lui, lorsqu'on comprend un mot, on associe ce mot à un concept qui lui-même détermine la référence de ce mot. Or, un concept ne détermine pas, d'après Putnam, cette référence, comme c'est le cas pour la Terre Jumelle. Et comme le concept est dans la tête d'un individu, c'est-à-dire dans la représentation sémantique mentale, comme le disent les mentalistes, l'exemple de la Terre Jumelle infirme le mentalisme.

Face à cette critique, Fodor répond que les contenus du concept d'eau sont différents d'un monde à un autre, mais que la Terre Jumelle ne peut pas exister réellement.

En fait, un autre exemple similaire à la Terre Jumelle est plus facile à discuter. Putnam dit qu'il ne peut pas distinguer entre « *beech* (hêtre) » et « *elm* (orme) » et ceci montre la divergence entre l'état mental psychologique et la référence. Fodor répond à ce problème en disant que ce genre de chose est négligeable et que l'on peut consulter un expert pour distinguer ces deux arbres, de même qu'on peut utiliser du papier de tournesol pour savoir si un liquide est acide ou basique (Fodor 1994).

Deux problèmes découlent de l'argument de Fodor. D'abord, il faudra expliquer la différence entre le concept de l'expert vis-à-vis de *beech/elm* et celui des

gens ordinaires. S'il n'existe pas de différence, la différence entre l'expert et le non expert est réduite à une différence de capacité à procéder à des inférences ou à des déductions. Mais Fodor refuse depuis toujours de réduire le concept à telle capacité, car ceci conduira, selon Fodor, à un individualisme extrême. Donc, il faudra accepter logiquement donc qu'il existe une différence, concernant le concept *beech/elm*, entre l'expert et le non-expert. Dans ce cas, un autre problème apparaît : s'il existe une telle différence, elle s'exprime de la façon suivante : l'expert possède un concept qui distingue BEECH de ELM, tandis que le non-expert n'en a pas. Ainsi, il n'est pas possible pour des non experts de voir la différence entre ces arbres, à moins que celui-ci n'apprenne les concepts concernés. Le dilemme chez Fodor est le suivant :

Fodor suppose que le concept ELM de Jean sur la terre a un contenu différent du concept ELM de Jean sur la Terre Jumelle, sur laquelle *elm* et *beech* s'échangent. Cela contredit le raisonnement précédent qui suppose le même concept chez les non-experts.

Bref, en essayant d'éviter l'énigme de Putnam avec l'argument expert / non expert on tombe sur une contradiction interne. Néanmoins nous n'acceptons pas l'argument de Putnam qui nous fait séparer le monde et le concept, car cela nous semble relever d'un dualisme d'avant Hegel.

Quant à Chomsky, il refuse d'être considéré comme un mentaliste au même titre que Fodor, car, pour Chomsky, ce qui est inné ce sont les éléments à partir desquels sont construits les représentations sémantiques mentales et ce ne sont pas les représentations elles-mêmes.

Burge (1979) présente une expérience mentale sur le contenu mental du concept. Il s'agit du contenu mental d'une maladie, dans deux mondes parallèles. Dans le monde₁, le patient Jean₁ ayant l'arthrite₁ à la cheville et au poignet voit un jour l'affection arthrite₁ s'étendre à sa cuisse. Plus précisément, Jean₁ a tiré

cette conclusion par raisonnement, il l'a fait en constatant un symptôme ou une douleur similaire à celle de l'arthrite₁ sur sa cheville. Mais dans ce monde₁, de même que dans notre monde, cette conclusion est considérée comme une erreur puisque le terme l'arthrite₁ ne s'applique qu'aux maladies articulaires. Dans le monde₂, tout est identique sauf que le jugement de Jean₂ est considéré comme correct. C'est-à-dire que, dans le monde₂, contrairement au monde₁, les médecins eux-mêmes emploient le terme *arthrite*₂ pour une affection, même si elle est située hors des articulations.

Avec cette expérience mentale, Burge maintient que les contenus mentaux sont différents entre Jean₁ et Jean₂ parce que les sens de arthrite₁ et arthrite₂ sont différents. D'après Burge la différence du contenu mental provient de « l'extérieur », par exemple de l'environnement social, et elle est liée aussi au raisonnement au niveau objectif, à savoir métalinguistique. Plus précisément, Burge avance que « les expressions que le sujet utilise parfois fournissent le contenu de ses états mentaux, même s'il a [...] mal compris ».

Si l'on constate bien que le recouvrement du sens d'un mot est différent entre les langues, par exemple entre *persuade* en anglais et *persuader* en français, l'argument de Burge, qui met l'accent sur l'extérieur ou l'environnement social, devrait s'y appliquer de la même manière. En d'autres termes, si on suit l'argument de Burge, on devrait conclure que le contenu mental est différent selon les langues qu'on utilise.

A première vue, cette remarque semble être juste. Par exemple, le contenu mental d'un locuteur francophone par rapport au terme *cœur* semble différent au celui de *heart* chez les locuteurs anglophones, parce que « J'ai mal au cœur » ne signifie pas « My heart hurts » mais « I feel sick/nauseous ». Certes, cet exemple montre la différence du recouvrement sémantique entre les deux langues. Cependant, ceci ne garantit pas en soi la différence du contenu mental.

En effet, lorsque l'on dit « le concept », on présuppose qu'un certain attribut est lié à la notion du concept. Et si le concept que nous désignons n'est pas bien établi, le débat peut nous amener à une discussion moins constructive. Une des erreurs les plus fréquemment commises est de construire un lien machinal entre le mot et le concept lexical. De fait, il ne faudra pas confondre la différence entre les concepts et la différence entre les mots. Par exemple, comme l'ont fait remarquer Murphy et Au (Murphy 1996, Au 1988), l'existence de plusieurs mots pour *neige* chez les Eskimos ne garantit pas que les Eskimos pensent la neige différemment des autres.

De ce fait, il sera faux, à notre avis, de juger que les locuteurs francophones ont un contenu mental différent pour l'*organe cœur*. D'ailleurs, nous ne pensons pas que Burge ait tenté d'avancer ce type d'idée. L'essentiel dans l'argument de Burge, d'après ce que nous avons compris, est que le concept d'un terme comporte la signification d'une communauté. En vertu d'un argument plus clair, imaginons que arthrite₁ a une conséquence c₁ et arthrite₂ une conséquence c₂. Puisque arthrite₂ comprend arthrite₁, c₂ inclut elle aussi c₁. Maintenant imaginons qu'une affection articulaire a une conséquence bien plus grave qu'une affection non articulaire. Autrement dit, c₁ est plus grave que c₂, car c₂ qui est composé de l'ensemble de c₁ et c₂' (c₂ sans c₁) ce qui moyenne la gravité. Par conséquent, Jean₁ aura plus de souci que Jean₂.²

C'est pour cela qu'il est incontestable qu'il existe une différence de contenu mental entre deux mondes. Certes, on peut toujours imaginer un cas similaire que celui de arthrite₁ et arthrite₂. Nonobstant, la question de savoir que si cette différence de concept mental est fondamentale ou bien si c'est une différence superficielle reste à discuter. Nous examinerons ce problème plus loin dans ce

²Cet exemple est un peu plus avancé que celui de Burge mais le principe reste le même à notre avis.

chapitre.

Le problème est maintenant de savoir si ce genre de différence de contenu mental est possible dans le monde réel. C'est-à-dire : est-il possible d'avoir deux sociétés distinctes ayant des champs sémantiques aussi différents qu'entre les mondes imaginés 1 et 2 ? A notre avis, bien que le même mot s'emploie dans le monde₁ pour désigner deux maladies différentes, il peut y avoir une manière différente d'exprimer cette maladie. Cette hypothèse est basée sur le fait que des différences sémantiques très importantes peuvent entraîner une stratégie selon laquelle les locuteurs utilisent une façon différente de s'exprimer. Ce n'est pas un cas impossible en effet. Par exemple, en français, le mot *cœur* s'emploie pour deux problèmes de santé différents. L'un vaut pour l'estomac et l'autre pour le cœur même. S'il n'y a pas de moyen de distinguer les deux sens, il sera très difficile de communiquer. En effet en français l'expression « J'ai mal au cœur » et l'expression « J'ai un problème au cœur » résolvent cette difficulté. Donc, si l'on applique cette stratégie réelle aux mondes imaginaires, on peut imaginer qu'il peut y avoir un changement de ce type dans le monde₂ contrairement à ce qui se passe dans le monde₁.

Si on voit une différence entre le CŒUR et NAUSEOUS, il est difficile de dire que les différences de contenu mental entre arthrite₁ et arthrite₂ sont plus importantes que celles entre cœur et nauseous. En outre, le sens de *cœur* est bien plus large que celui de *nauseous*, comme on le voit dans une phrase comme « J'ai un problème au cœur » où *cœur* désigne *heart* et non *estomac*. De même, entre les phrases « Je l'ai convaincu » et « I persuaded him » les contenus seront différents si l'on suit l'argument de Burge, car le champ sémantique de *persuade* est plus large que celui de *convaincre*, ce qui correspond, d'après l'argument de Burge, à des environnements sociaux différents.

En effet, les arguments autour de ces sujets sont étroitement liés au sens des

mots aussi. Nous examinerons ce point dans le chapitre suivant.

Putnam a raison lorsqu'il dit qu'un même état psychologique peut correspondre à des référents différents. Fodor pense par contre que les deux concepts (celui de notre monde et celui de la Terre Jumelle) ont des contenus différents. En effet, cette différence vient du fait que le concept chez Putnam est plus restreint que le concept chez Fodor : ce premier distingue le concept de l'extension tandis que ce dernier ne les distingue pas.

2.3 Théorie du concept

Avant de présenter notre théorie du concept, examinons les différentes théories du concept selon des points de vues variés.

2.3.1 Théories diverses

Traditionnellement le concept a été abordé par la donnée de conditions nécessaires et suffisantes et la théorie qui recourt à cette approche est appelée théorie définitionnelle ou théorie classique. Dans cette théorie, le concept de « *bachelor* », par exemple, se définit comme « un homme non marié ». L'avantage de la théorie définitionnelle se trouve dans le fait qu'elle peut réduire un système complexe, comme un concept, à des relations logiques. Cette théorie a été étendue à plusieurs domaines de recherche, y compris à des domaines d'applications. En effet, les relations logiques permettent d'implémenter facilement un système informatique, comme un système expert, puisque le principe de base de ce dernier est basé lui aussi sur les relations logiques.

Cette théorie a cependant des difficultés à expliquer les cas qui, malgré leur conformité à la définition, ne peuvent pas être considérés comme des exemples d'un concept. Par exemple, bien qu'ils soient *bachelors* par définition, des céliba-

taires d'un certain âge, ou des couples qui ont vécu ensemble assez longtemps sans être mariés, ne peuvent pas être considérés comme *bachelors* ; Robinson Crusoe qui ne vit pas dans une société normale en est un autre contre-exemple (Fillmore 1982). Un autre exemple bien connu, qui met en cause la légitimité de la définition, est le concept de « jeu » présenté par Wittgenstein. Selon lui, il n'est pas possible d'avoir une définition générale qui satisfasse tous les exemples du jeu. Par exemple une définition comme *activité compétitive pour gagner* ne s'applique pas à tous les jeux, la ronde enfantine par exemple ; une définition comme *activité amusante* ne suffit pas, car le degré du trait amusant est très variable selon les jeux — voir la différence entre le jeu d'échec et le morpion, par exemple.

Ces exemples montrent le point faible, voire le défaut, de la théorie définitionnelle. Néanmoins il n'est pas suffisant d'infirmier la théorie définitionnelle en tant que telle, car il est possible de modifier les définitions de telle sorte qu'elles puissent rendre compte des exceptions. A titre d'exemple, en précisant la définition de *bachelor* comme « un jeune homme non marié vivant dans une de nos sociétés » au lieu de la définition traditionnelle « un homme non marié », la théorie définitionnelle pourrait résister à la critique mentionnée ci-dessus. De même, il est possible de *ignorer* la définition en faisant la combinaison des définitions possibles de « jeu » comme « activité d'une ou plusieurs personnes faisant une compétition dans la plupart des cas et qui sépare en général, mais pas toujours, le gagnant et le perdant à la fin. ». Afin d'éviter de faire entrer les activités violentes comme la *guerre* parmi les exemples de jeu, une modification peut être apportée à la définition : « qui exclut l'usage des armes ». Pour inclure la « roulette russe », une autre modification sera faite encore comme « à l'exception de la roulette russe », etc. En effet cette approche qui contient des exemples comme « guerre » ou « roulette russe » nuit gravement à l'approche de la relation logique, car cette manière de définir n'est pas très éloignée de l'énumération ou du classe-

ment des exemples, ce qui contraste avec l'approche logique. D'ailleurs, il reste à discuter si une telle définition est vraiment définitionnelle.

Malgré ce problème et sa régression éventuelle, la théorie définitionnelle pourrait survivre *en principe* grâce à cette modification mais c'est la théorie prototypique qui a posé comme question décisive la légitimité de la théorie définitionnelle. La théorie prototypique est une théorie qui suppose le prototype d'un concept donné, par exemple, un rouge-gorge pour le concept OISEAU.

Cette théorie présuppose logiquement à la fois des contre-prototypes, par exemple « pingouin » pour le concept oiseau, qui s'opposent au prototype, et des exemples entre ces deux extrêmes qui pourraient être ordonnés vis-à-vis du prototype et du contre-prototype. Par exemple, si l'on donne cinq points pour le rouge-gorge comme exemple typique d'oiseau et un point pour l'autruche, la poule pourra avoir trois points et le pingouin quatre points, etc. Ces différentes valeurs prototypiques entre les exemples d'un concept sont mesurables par la comparaison des temps d'accès, de récupération, d'apprentissage ou bien par les points d'évaluation de la typicalité fournis par les sujets. De nombreuses expériences de ce genre en psychologie cognitive garantissent la validité de la théorie prototypique (Mervis et Rosch 1981, McCloskey et Glucksberg 1978, Rips, Shoben et Smith 1973, Rosch 1973b, Rosch 1975, Tversky 1977, Kahneman et Tversky 1973, Collins et Loftus 1975).

L'effet prototypique pose plusieurs questions à l'égard du concept : reflète-il la structure complexe du concept ? Est-il lié à la représentation mentale du concept ? Comment la structure, la représentation se forme-t-elle au sein du concept ? Ou bien s'agit-il d'un tout autre lien indépendant de la nature du concept ?

Osherson et Smith (1981) ont étudié cette dernière question. Ils supposent qu'il existe un noyau du concept et ils insistent sur le fait que l'effet prototypique provient du processus d'identification. Par exemple, le concept *femme* contient

l'information de l'existence d'un système reproductif et le processus d'identification contient les informations sur la forme du corps, la longueur des cheveux, la hauteur de la voix, etc. Ainsi la théorie définitionnelle rend compte du noyau et la théorie prototypique rend compte du processus d'identification. Par conséquent, selon leur proposition, l'effet prototypique se réduirait à un processus et la nature du concept elle-même n'aurait aucun lien avec l'effet prototypique.

Bien qu'ayant suivi une voie différente, Lakoff (1987a, 1987b) partage l'essentiel de leur point de vue, dans la mesure où il n'affirme pas l'existence de structure ou de représentation complexe du concept.

Nous pensons que la proposition de Osherson et Smith vis-à-vis du concept est un recul théorique. Car au lieu de proposer une nouvelle théorie qui non seulement surmonte la théorie prototypique mais aussi la théorie définitionnelle, ils ont choisi de ressusciter la théorie définitionnelle qui était déjà infirmée par la théorie prototypique.

En effet, l'expérience psychologique consiste à mesurer une sorte de quantité ou de qualité afin de valider ou d'infirmier une théorie, un modèle ou une hypothèse sur l'état mental. Or, si une théorie concernant l'état mental n'explique ou ne prévoit pas les résultats des expériences psychologiques cognitives, cette théorie doit ou bien être rejetée ou bien être classée comme indépendante de l'état mental.

De nombreux résultats d'expériences en psychologie cognitive ont confirmé la validité de la théorie prototypique et ont infirmé la théorie définitionnelle. Face à cet échec, la théorie définitionnelle peut continuer à exister en tant que théorie indépendante de l'état mental. Mais la proposition de Osherson et Smith est en fait une tentative pour faire revivre une théorie qui a été prouvée non adéquate pour expliquer l'état mental. Certes, ils ne préconisent pas exclusivement une théorie définitionnelle. Cependant, le fait que leur théorie est tout à fait compatible avec la théorie définitionnelle nous permet d'avancer ce que nous avons dit.

En effet, la théorie ptolémaïque peut résister contre la théorie copernicienne en fignant les mouvements des planètes mais la signification d'une telle tentative est maintenant claire. La théorie, chez Osherson et Smith, reconstruit ce type de problème. En effet ce qu'apportent le travail de Wittgenstein et la théorie prototypique est qu'il est impossible d'expliquer le jugement humain au moyen d'une théorie définitionnelle du concept et que leurs résultats sont incompatibles avec la théorie définitionnelle.

Par leur expérience concernant l'effet prototypique, Armstrong et al. (1983) ont obtenu des réponses graduelles des sujets sur le jugement d'un bon exemple pour les nombres pairs ou impairs. Par exemple 4 et 3 sont jugés être de bons exemples de nombres pair et impair tandis que 30 et 15 sont considérés comme de mauvais exemples ; et cette tendance a été observée pour des consignes variées. Ces résultats montrent en effet que l'effet prototypique peut s'appliquer à des concepts très larges au-delà des concepts d'objets concrets. Cependant, pour Armstrong et al. ces résultats infirment la validité de la théorie prototypique. Leur argument est le suivant : l'effet prototypique relève de la structure complexe du concept ; or, le concept NOMBRE PAIR est bien défini, exact et non flou, autrement dit sa structure est simple ; le fait qu'une telle structure soit simple montre que l'effet prototypique se contredit.

Face à cette critique, Lakoff défend la théorie prototypique en disant que l'effet prototypique n'a pas de lien avec la structure conceptuelle ou avec la représentation conceptuelle mais qu'il provient de l'interprétation du modèle cognitif. Le modèle cognitif au sens de Lakoff est un modèle qui examine l'appariement entre les différentes conditions du modèle et la connaissance du monde. Par exemple, l'effet prototypique pour les nombres pairs/impairs est expliqué par la condition du système de numération qui n'utilise que dix chiffres pour exprimer un nombre, ainsi ce système de numération privilégie un nombre à un chiffre à un nombre de

grande taille.

Ainsi, malgré leur différents points de vue concernant la théorie prototypique, Osherson et Smith, Armstrong et al. et Lakoff partagent le même point de vue à propos de la structure et de la représentation du concept. C'est-à-dire qu'un concept bien défini et exact n'a pas de structure ni de représentation complexe.

Une approche plus directe, qui nie la structure ou la représentation complexe du concept, est l'approche atomiste. Au lieu d'éviter ce problème en niant le lien entre la structure, la représentation et le concept, les atomistes déclarent explicitement que la structure et la représentation mentale du concept ne sont pas complexes. Pour eux, le concept est comme un atome qui ne se divise plus. La négation de la structure conduit logiquement à la simplification de l'apprentissage du concept. C'est de là que vient leur thèse du concept inné et de l'approche nativiste.

La théorie atomiste refuse à la fois l'approche définitionnelle et l'approche statistique. Par exemple, Fodor refuse la structure complexe d'un concept et par la suite il refuse aussi la notion de polysémie. Pour Fodor la polysémie *keep* signifie justement « *keep* » et rien d'autre.³ Fodor dit explicitement que la plupart des mots ne sont pas décomposables et il nie que le concept lexical ait une structure interne. C'est pour cela que, pour les atomistes, les concepts correspondent aux mots et la plupart des concepts lexicaux sont innés.

D'après Laurence et Margolis (2002, 1999), en modifiant l'argument de Fodor, l'argument de la théorie atomiste se présente ainsi :

1. A part des miracles ou super-science futuriste, tous les concepts sont innés ou construits à partir de primitives.
2. S'ils sont construits à partir de primitives, ils ont des structures.
3. Les concepts lexicaux n'ont pas de structure.

³« [T]here's no better way to say what 'keep' means than to say that it means *keep* » (Fodor 1998 : 55).

4. Donc, les concepts lexicaux ne sont pas construits à partir de primitives

5. Par conséquent, les concepts lexicaux sont innés.

De ce fait, si l'on prouve que les concepts lexicaux ont des structures, l'argument des partisans de l'atomisme s'effondre.

La différence entre Laurence et Margolis d'une part et Fodor d'autre part consiste dans le fait que ces premiers affirment que les concepts primitifs peuvent être appris alors que Fodor le nie.

Block pense que le contenu d'un concept est déterminé par ses relations aux autres concepts dans un système représentationnel appelé « conceptual role semantics ».

Block refuse d'accepter les concepts primitifs, car d'après lui les concepts sont reliés entre eux. Ainsi, contrairement à l'approche traditionnelle, selon Block les concepts ne sont pas appris un par un.

Nous pensons que le problème essentiel de la théorie atomiste est qu'elle présente un piège linguistique.

Si le concept « *keep* » est atomique, il n'y a pas de raison de refuser d'accepter « *persuade* », qui est censé être moins polysémique que « *keep* », comme étant atomique. De la même manière, il faudra accepter aussi, en français, que « persuader » et « convaincre » sont tous les deux atomiques. Il faudra accepter ensuite que deux concepts réunis « PERSUADER + CONVAINCRE » forment un tout non atomique c'est-à-dire complexe, car une structure atomique ne peut pas en principe avoir deux éléments ainsi séparables et distincts. Le problème est que le concept équivalent à l'anglais PERSUADE est PERSUADER + CONVAINCRE. De là découle la contradiction : un même concept est tantôt atomique en anglais tantôt complexe en français. Il n'y a qu'une solution pour éviter cette contradiction : « persuade » est atomique et « persuader » et « convaincre » sont *sous-atomiques*. Toutefois, comme le montre le terme lui-même, *sous-atomique* est contradictoire. Ce type

de problème n'est pas réservé au français non plus, car il existe des exemples inverses où deux mots en anglais correspondent à un mot français comme « *mutton* » et « *sheep* » vers « mouton ».

Comme discuté plus haut, la correspondance entre un concept et un sous-ensemble du concept n'est pas exclue comme par exemple la différence entre une vingtaine de sortes de neige et une seule neige. Selon le point de vue, cette correspondance se fera soit d'une unité atomique vers des sous-unités, soit d'une unité complexe vers une unité atomique. C'est un sujet à discuter mais il faudra noter qu'aucune langue ne peut éviter d'avoir un concept complexe ou sous-atomique. : soit « *mutton* » et « *sheep* » sont atomiques et « *persuade* » est complexe ; soit « *persuade* » est atomique et « *mutton* » et « *sheep* » sont sous-atomique. Dans tous les cas, cela n'est pas compatible avec la théorie atomique qui refuse à la fois les structures complexes et les divisions des unités atomiques.

On peut se demander si les défenseurs de la théorie atomiste pourraient accepter la notion de concept atomique indépendant de la langue. Mais il reste la question de savoir si on doit séparer le sens du mot de l'intérieur ou fusionner les mots. Et surtout de savoir quel critère appliquer.

Notre position vis-à-vis du concept est qu'il n'existe pas de correspondance directe entre un concept et le sens total du mot. Un mot peut être considéré comme une composition de concepts ou un concept peut être représenté par un ensemble des mots. Par ensemble de mots nous entendons aussi les locutions et les mots composés, et pas seulement les mots isolés. Par exemple, « pomme de terre » ou « tomber dans les pommes » constituent un concept et non plusieurs concepts.

2.3.2 Théorie proposée

Bachelard distingue les notions du concept MASSE selon cinq philosophies : réalisme naïf, empirisme clair et positiviste, rationalisme newtonien ou kantien,

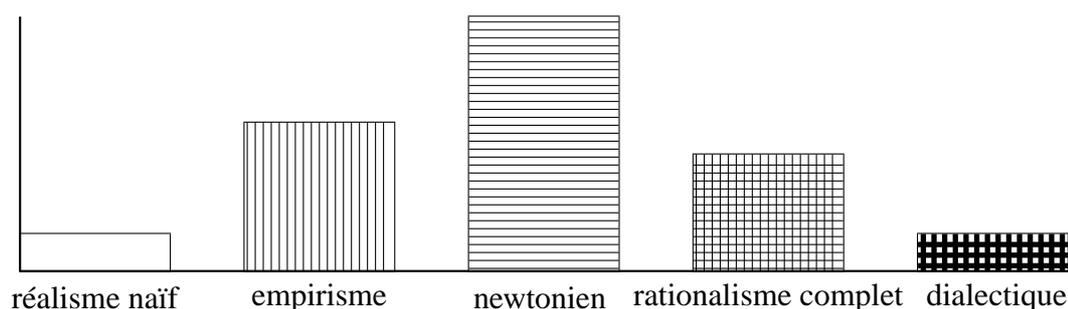


FIG. 2.1 – Profil épistémologique de la notion personnelle de masse — adapté de Bachelard (1940). L'ordonnée représente la fréquence d'usage effectif de la notion et l'importance relative.

rationalisme complet, rationalisme dialectique (Bachelard 1940). Ces cinq philosophies correspondent respectivement à la quantité de matière, la masse en tant que loi, la masse de la relativité et la masse négative chez Dirac. Comme schématisé dans la figure 2.1, Bachelard constate que, même si nous vivons à une époque post-diracienne, notre notion de masse en reste en général au rationalisme classique, suivi de l'empirisme clair et positiviste, du rationalisme complet, du réalisme naïf et du rationalisme dialectique. Ainsi, la notion de masse la plus correcte du point de vue scientifique est la moins représentative dans notre représentation mentale. Bachelard avance la raison pour laquelle est observée cette disparité des concepts :

On reconnaît, sur notre schéma, l'importance attribuée à la notion rationaliste de masse, notion formée dans une éducation mathématique classique et développée une longue pratique de l'enseignement de la Physique élémentaire. En fait, dans la majorité des cas, la notion de masse se présente pour nous dans l'orientation du rationalisme classique. Pour nous, en tant que notion claire, la notion de masse est surtout une notion rationnelle. [...] nous sommes dominés par la

tendance simplement rationnelle. Notre rationalisme simple entrave notre rationalisme complet et surtout notre rationalisme dialectique.

La remarque de Burge sur le concept mental de CONTRAT qui ne dépend pas forcément de la définition d'un dictionnaire — par exemple contrat en tant que *contrat écrit sur papier* contrairement à la définition dans le dictionnaire qui englobe aussi les *contrats non écrits* — est comparable à cette constatation de Bachelard.

Bien que Bachelard propose cette différence épistémologique au niveau personnel, ce genre de raisonnement pourrait, à notre avis, s'appliquer plus généralement. Nous proposons ainsi une théorie du concept qui tient compte de ce type de la relativité épistémologique. Nous présenterons notre théorie du concept par cinq caractéristiques principales. Commençons par la caractéristique représentative de la théorie.

1. a. Le concept est la représentation gaussienne des relations sémantiques, stéréotypiques et contextuelles plutôt qu'une définition ou un sous-ensemble flou.

Il n'est pas si facile parfois de trouver un terme/concept à partir d'une définition seule comme « Activité physique ou mentale purement gratuite, qui n'a, dans la conscience de la personne qui s'y livre, d'autre but que le plaisir qu'elle procure »⁴. Si ceci montre une difficulté de l'approche définitionnelle, la critique de Wittgenstein met en évidence l'impossibilité de cette approche. En outre, la théorie définitionnelle ne peut pas expliquer l'effet prototypique qui s'observe pourtant de façon quasi-universelle — même pour des concepts « moins concrets » comme les NOMBRES PAIRS OU IMPAIRS. L'approche utilisant un sous-ensemble flou qui rend plus flexible la frontière rigide de la définition souffre aussi de ce problème,

⁴La définition de *jeu* dans *Le Petit Robert*.

car elle est essentiellement fondée sur l'approche définitionnelle et elle ne peut pas exister sans définition préalable.

Contrairement à la théorie définitionnelle ou à la théorie atomiste du concept, même un concept considéré comme très atomique, comme par exemple les noms propres, n'est pas en fait atomique mais vêtu de relations riches. De ce fait, nous proposons une théorie du concept qui considère le concept en tant qu'une représentation des relations sémantiques, stéréotypiques et contextuelles. Par exemple, les quatre types de sens chez Putnam constituent un concept.⁵ Le fait que les sens stéréotypiques ou prototypiques constituent le concept permet à notre théorie du concept de rendre compte de l'effet prototypique. En fait, la représentation gaussienne de notre théorie implique l'existence de caractéristiques génériques ou représentatives d'un concept donné comparables à celles des prototypes. Néanmoins, contrairement à la théorie prototypique, dans notre théorie, l'effet prototypique n'est qu'une des caractéristiques du concept. En fait, notre théorie du concept garde une distance surtout avec l'interprétation dominante de la théorie prototypique qui en ferait une théorie d'exemplaires. D'une certaine manière, notre théorie n'est pas très éloignée du point de vue de Block qui refuse un apprentissage un à un et qui préconise des connaissances reliées.

- 1. b.** Le centre de cette représentation reflète les connaissances générales et l'extrême les connaissances potentielles. Lorsque l'on dit le *concept*, cela désigne dans la plupart des cas *le centre de la représentation*.

L'analogie de la représentation centrale du concept avec l'effet prototypique est un des aspects de la représentation de la théorie. De manière plus globale, le centre de la représentation d'un concept reflète des connaissances générales. Par exemple, les sens TRANSPARENT, LIQUIDE et BOIRE du concept EAU peuvent

⁵La relation avec l'extension sera discutée plus loin.

être plus représentatifs que DÉSALTÉRANT, MINÉRAL ; et H_2O sera encore moins représentatif et se situera à l'extrême de la courbe gaussienne de même que MASSE DIRACIENNE se situe à l'extrémité dans le profil épistémologique.

Le point crucial pour voir la différence entre représentation centrale et représentation marginale consiste à distinguer la différence entre connaissance réelle et connaissance potentielle. Si on considère un concept technique comme KO dans le jeu de go, il est évident qu'il existe des gens pour lesquels le concept KO est littéralement *vide*. Pourtant, le concept KO existe, ainsi que le terme *ko*. Cependant cette ignorance n'est pas une ignorance absolue comme *ding an sich* chez Kant. Ainsi le fait qu'il soit possible d'acquérir le concept KO pour n'importe qui a priori transforme cette ignorance en connaissance potentielle.

En ce sens, Oscar₁ et Oscar₂ ont la même connaissance réelle mais ont une connaissance potentielle différente. En d'autres termes, ils ont le même concept mais ils pourront avoir des concepts différents s'ils deviennent intéressés par la nature de l'eau voire s'ils deviennent chimistes ou physiciens.

Il est important de remarquer que nous ne sommes pas des robots qui font des mises à jour automatiques dès qu'apparaît une nouvelle découverte ou une nouvelle connaissance. De plus, même si c'était le cas, la représentation chez le robot ne contredirait pas une découverte nouvelle sur la nature de l'eau. Fodor pourrait avoir raison si et seulement si les connaissances absolues étaient intégrées dans le cerveau du robot depuis toujours. Mais il est impossible d'imaginer un robot qui connaisse *a priori* la nature de l'univers qui pourrait être découverte par exemple dans plusieurs millions d'années.

Le fait d'avoir distingué une différence entre ces degrés de connaissance nous a permis de classer le concept. Le même principe s'applique à la distinction entre le C_p et l'extension.

2. Le concept est formé de la même manière selon les indivi-

du. Le critère *partagé* n'est pas toujours constaté mais il est maintenu de manière globale.

Le critère « partagé » du concept veut dire qu'il existe un point commun entre les individus qui assure la communication entre eux. Toutefois, des exemples comme *ko* et *masse diracienne* mettent en évidence la différence du contenu du concept ou de la représentation du concept.

Une théorie du concept doit répondre à ce type de problème. Nous proposons une réponse qui est liée au principe 1.b.

D'abord, dans notre théorie du concept, le concept est formé de manière identique pour chacun des individus, c'est-à-dire que le même principe s'applique pour tout le monde sans exception. Cette déclaration dénonce l'individualisme extrême et préconise le caractéristique partagée du concept. Elle est aussi conforme à la constatation qu'il existe des catégories identiques pour la couleur malgré les différences entre langues (Rosch 1973a, Delgado 2004). Cette identité du concept correspond au principe 1.b. selon lequel le *concept* désigne la plupart des cas la partie centrale du concept. Le critère partagé est ainsi assuré par notre théorie car la partie centrale est quasiment identique entre les individus. En ce sens, la représentation du concept dans notre théorie peut être interprétée tantôt comme personnelle tantôt comme globale.

La différence du sens entre des langues différentes par exemple entre *river* et *fleuve* ne renvoie pas à une différence insurmontable, car le concept *brook+river* et le concept *ruisseau+rivière+fleuve* peuvent être considérés comme identique. De même, si nous regroupons tous les concepts concernant les maladies nous aurons une représentation mentale dont la différence au contenu mental devient significativement réduite. Dans cette représentation, une partie représente une maladie mortelle, une autre une légère maladie, etc. Pour cette raison, même s'il existe une différence entre les contenus du concept dans certains cas, nous pensons que,

à la base, les concepts les plus généraux ne sont pas différents. En ce sens, cette déclaration affirme l'existence, entre les êtres humains, d'un fondement commun qui puisse se manifester au niveau des capacités intellectuelles, de la communication et de l'échange d'émotions communes malgré les différences langagières et culturelles.

En deuxième lieu, l'application de ce même principe ne garantit pas exactement la même représentation du concept pour chacun des individus. Par exemple, contrairement à la plupart des gens, les spécialistes connaissent le concept KO et le concept MASSE DIRACIENNE de manière concrète. Dans ce cas, il faudra dire qu'il existe les différents contenus du concept entre les individus. Il n'est pas contradictoire néanmoins avec la déclaration précédente concernant le critère partagé. De même que chaque individu possède une vue différente malgré une identité physiologique, cette différence du contenu du concept peut co-exister avec le critère partagé.

Si on applique ce principe à l'énigme de la Terre Jumelle, Oscar₁ et Oscar₂ ont le même concept d'eau, tandis que le physicien Pierre₁ a un concept d'eau différent de celui d'Oscar₁ ou d'Oscar₂. Et c'est pour cette raison nous ne sommes pas d'accord avec Fodor qui affirme une différence du contenu du concept entre Oscar₁ et Oscar₂.

En ce sens, notre théorie du concept peut inclure l'extension mais elle est différente de la position fodorienne. Afin de mieux comprendre cette différence, un résumé et une proposition sur la nature du concept et son ontologie seront indispensables.

- 3. a.** Le concept est à la fois l'entité mentale et l'entité abstraite. Cette caractéristique du concept prend contre-pied les positions extrêmes qui préconisent exclusivement l'une ou l'autre des deux entités (mentale ou abstraite). En ce sens, notre théo-

rie du concept nie la dichotomie entre entité mentale et entité abstraite.

De même que le dualisme de Descartes a été nécessaire pour progresser face au monisme stoïcien, la dichotomie de l'entité mentale et de l'entité abstraite a joué un rôle non négligeable face aux approches monistes extrêmes apparues de part et d'autre comme l'internalisme et l'externalisme. Le fait de considérer le concept comme une entité abstraite, comme Frege, a l'avantage d'établir des relations entre les concepts et de ne pas être enfermé dans un subjectivisme extrême. Par exemple, le mode de représentation de Frege explique raisonnablement le fait que différents sens aient la même référence. Quant à l'entité mentale, elle défend correctement la compatibilité entre les humains, par le soutien de l'identité de l'entité mentale chez l'homme, sans quoi il serait difficile d'établir un lien psychologique partagé.

Si l'on considère un robot qui apprend des concepts, la portée de ces deux approches devient plus claire. Imaginons un robot équipé d'une capacité à représenter les concepts en tant qu'entités mentales. Aussi performant soit ce robot, s'il n'a pas accès à des relations correctement établies, il aura des difficultés à apprendre des choses. Réciproquement, si un robot n'a pas la capacité d'organiser et de représenter des relations établies par l'approche définitionnelle par exemple, il ne pourra pas représenter correctement ces relations.

De même que Kant et Hegel notamment ont permis de progresser en surmontant le dualisme de Descartes, la dichotomie entre entité mentale et entité abstraite mérite d'être révisée voire dépassée pour qu'un progrès soit possible.

Avec l'approche logique ou l'entité abstraite, il est relativement facile d'établir des relations entre les concepts. Par contre, les théories de l'entité mentale ont plutôt pour intérêt d'expliquer comment chaque individu réussit à communiquer avec les autres. Pour ce faire, il reste deux solutions : soit supposer une entité

mentale compatible, similaire voire identique entre individus, soit imaginer une capacité ou un processus qui assure la communication entre eux. Pour l'entité mentaliste, la deuxième solution est inacceptable car cette solution pourrait nier l'existence de l'identité de la représentation mentale : si l'on accepte que quelque chose, hors de la représentation mentale comme la capacité d'inférence, assure la communication, ceci est compatible avec la position qui nie la représentation mentale partagée. La supposition d'une identité pour l'entité mentale rencontre cependant des problèmes sévères, par exemple l'énigme de Putnam qui tente de démonter les prémisses préalables de la théorie de l'entité mentale. Comme nous l'avons montré plus haut, les mentalistes convaincus comme Fodor rencontrent une contradiction sévère dans l'explication des énigmes de Putnam notamment celle de *beech/elm*.

Quant à l'entité abstraite, elle se heurte au problème de l'incapacité à expliquer l'énigme de Putnam et les résultats de la psychologie cognitive, l'effet prototypique par exemple. Une solution possible pour les partisans de l'entité abstraite est, comme l'a fait Putnam, de séparer le concept et l'extension, mais cette séparation nuit gravement à la caractéristique du concept comme entité abstraite.

Le point central qui est de réconcilier ou de surmonter la dichotomie entre entité abstraite et entité mentale repose donc sur une supposition qui doit être expliquée avec une théorie générale.

Examinons maintenant l'étendue du concept :

- 3. b.** Le concept peut englober une extension⁶ et peut ne pas en englober. Notre théorie du concept dénonce donc à la fois la séparation systématique entre concept et extension chez Putnam et la fusion systématique du concept et de l'extension chez Fodor.

⁶Au sens de Putnam.

D'abord nous estimons que la tentative de Putnam de distinguer le concept (C_p) et l'extension en faisant l'expérience mentale de la Terre Jumelle montre de fait qu'il existe une différence entre le C_p et l'extension. Nous sommes d'accord jusqu'ici mais la voie se divise maintenant en trois. Avec cette constatation Putnam tire la conclusion selon laquelle la représentation mentale est fautive. Quant à Fodor, il affirme que le contenu du concept de l'eau entre deux mondes est différent. Chacun de son côté critique des positions comme l'externalisme extrême et l'internalisme extrême. Mais nous estimons que les deux approches posent également des problèmes. A notre avis, ce que Putnam attaque n'est pas la représentation mentale en général mais la représentation mentale fodorienne. En effet l'argument de Fodor est incompatible avec celui de Putnam, car le concept chez Fodor comprend l'extension comme nous avons examiné plus haut. Nous pensons que cette position de Fodor est une sorte de réponse à la question suivante : si la représentation mentale n'inclut pas d'extension tel que « L'eau est H_2O », ainsi si elle n'est pas dans la tête, où est la place pour telle extension ?

En effet, en attribuant la capacité de distinguer *elm* et *beech* aux experts, Fodor a été sur le point de résoudre cette énigme mais il s'en est détourné en identifiant les experts au papier tournesol.

Comme discuté plus haut, nous refusons la dichotomie entre entité mentale et entité abstraite. Nous refusons aussi la distinction entre concept et extension. Ceci ne veut pas dire que nous englobons l'extension dans tous les cas comme Fodor. Au lieu de faire entrer l'extension et de dire que le contenu de l'eau entre les deux mondes (la Terre Jumelle et la terre) est différent, nous supposons un concept de distribution gaussienne comme Bachelard. La plupart des cas, l'extension de l'eau est exclue, et n'apparaît que dans le cas exceptionnel reflété par la région extrême de la courbe gaussienne. En ce sens, l'étendue du concept de notre théorie pourrait être dite dynamique en général mais fixée dans un groupe spécifique.

Pour l'exemple de l'eau, ces groupes distinguent les gens ordinaires, les scientifiques etc. et pour l'exemple de « Confucius » ces groupes seront les asiatiques et les non-asiatiques y compris les spécialistes et les gens ordinaires.

Cette approche n'a aucun lien avec l'individualisme, car cette étendue du concept suppose une portée moyenne. En effet le résultat de la psychologie n'est d'autre chose que la moyenne des résultats.

4. Le concept peut et peut ne pas avoir de structure interne. Dans la plupart des cas, le concept a une structure complexe.

Nous supposons que la structure du concept est en général complexe. En ce sens, notre position est diagonalement différente de l'approche atomiste. Le fait qu'un mot polysémique comme *conduit*, *insensible* ou *chaud* possède plusieurs sens distincts assure en effet l'existence d'une structure interne du concept. L'organisation du concept que nous envisageons n'a pas non plus d'analogie avec le sens des mots classés de manière traditionnelle ou lexicographique. En effet, le classement lexicographique a une affinité très forte avec la théorie définitionnelle du concept. Comme nous l'avons examiné, la théorie définitionnelle du concept n'est pas très apte à expliquer les caractéristiques du concept. De ce fait, nous proposons une théorie du concept dont les caractéristiques sont reflétées dans une représentation du concept.

Nous gardons nos distances vis-à-vis d'une approche de type atomiste comme celle de Laurence et Margolis qui expliquent l'acquisition du concept à partir d'un concept primitif vers un concept complexe.

Mais nous envisageons des atomes au sein d'un concept au lieu d'identifier un concept à un atome. De ce fait ou bien plusieurs atomes constituent un concept ou bien un seul atome peut constituer un concept.

La structure du concept reflète l'identité du concept. Par conséquent, si deux concepts possèdent la même structure/représentation, ces deux concepts sont consi-

dérés comme identiques. Par exemple, le concept général de l'eau entre la terre et la Terre Jumelle est identique car, comme dit plus haut, le concept général de l'eau exclut l'extension (H_2O).

Les sens d'un mot polysémique correspondent aux *concepts de base* dans notre théorie du concept. La structure complexe du concept permet de classer les différences dans un concept de manière illimitée, mais un concept de base comme COMPÉTITION DE SPORT peut se distinguer du concept MATCH. Le concept de base est différent du concept atomique car ce premier permet une structure interne complexe. Pour cet aspect, un modèle plus concret lié à cette théorie sera présenté dans les chapitres suivants.

5. La compositionnalité du concept dépend de la représentation relationnelle du concept. La compositionnalité peut être expliquée par le principe de déplacement et de fusion.

Imaginons un concept comme CHIEN D'UN MILLIONNAIRE QUI A TUÉ SON MAÎTRE, similaire à l'exemple de Fodor et Lepore (Fodor et Lepore 1996, Laurence et Margolis 1999). La théorie prototypique du concept est critiquée par le fait qu'il n'existe pas de prototype pour tel concept. En fait, pour la théorie prototypique d'exemplaires, cette constatation est évidente et, à première vue, notre théorie du concept semble souffrir aussi de ce type de critique. Cependant, l'observation plus détaillée révèle l'importance de la représentation relationnelle même pour ce genre d'exemples. D'abord, le concept CHIEN D'UN MILLIONNAIRE peut avoir une représentation qui marginalise, dans la représentation du concept, des CHIENS SALES DE RUE ; de même, le concept du CHIEN QUI A TUÉ marginalise aussi les PETITS CHIENS. Or, il est tout à fait possible que *chien d'un millionnaire* soit en réalité plus sale qu'un chien de rue, que *chien qui a tué* est de fait un très petit chien. Cette possibilité d'erreur ne dénonce pas cependant la validité d'une théorie représentationnelle. Au contraire, ce type de théorie met en évidence sa

similarité avec le concept humain.

En outre, la théorie prototypique est critiquée par son absence de capacité de compositionnalité. Par exemple, le prototype de PET et celui de FISH ne donnent pas de clef pour le concept composé PET FISH, car *poisson rouge* n'est ni le prototype de PET, ni le prototype de FISH. Notre théorie du concept en tant que représentation gaussienne peut avoir une solution à cette question. La représentation de *poisson rouge* peut être située loin du centre de la représentation du concept PET. Cependant, la présence du terme *fish* peut déplacer la courbe de sorte que la représentation de *poisson rouge* puisse être plus représentative, ou bien la fusion de ces deux concepts PET et FISH peut engendrer une représentation très différente de leur représentation originale. La démonstration de ce type de chose sera présentée lors de la discussion de la désambiguïsation des sens du modèle informatique basé sur notre théorie du concept.

Chapitre 3

Sens et contexonyme

Le concept, caractère très distinctif chez l'homme comme je l'ai examiné dans le chapitre précédent, n'est pas *tangible* en tant que tel et ne s'exprime que par le biais intermédiaire de la langue. En ce sens, la langue est un caractère typiquement humain et c'est celui qui distingue le plus explicitement les hommes des animaux. L'apparition de la première langue humaine remonte à au moins 50 mille ans et la variété des langues se manifeste par l'existence de plus de six mille langues (Song 2001) différentes sur la terre aujourd'hui.

Si l'on accepte que la capacité à former des concepts nous est commune et qu'elle se manifeste à travers les différentes cultures et les différentes communautés de langues, il faudra admettre aussi qu'il existe une équivalence sémantique entre les différentes langues. Cette équivalence ne présuppose pas néanmoins une correspondance systématique mot à mot.

En effet, il est fréquemment constaté qu'un mot dans une langue correspond souvent à un ensemble de mots dans une autre langue et réciproquement. Par exemple, le mot anglais *language* correspond à l'ensemble des deux mots français *langage* et *langue* et non à l'un des deux pris isolément. S'il n'existait que ce type de discordance entre les mots, à savoir une correspondance parfaite asymétrique

comme « mouton = sheep + mutton », établir l'équivalence sémantique des mots ne poserait pas de gros problèmes malgré l'arbitraire de cette correspondance. En effet il suffirait dans ce cas de construire une liste complète des équivalences sémantiques entre les langues comme par exemple celles-ci : { language = langage + langue }, { persuade = persuade + convaincre }, { pousser = push + propel }, etc.

Mais en réalité la correspondance sémantique des mots entre les langues est bien plus compliquée. De fait, on constate plus fréquemment que l'équivalence sémantique des mots entre les langues — loin d'être parfaite — est partielle. Par équivalence partielle, nous entendons la relation selon laquelle un mot d'une langue ne correspond à aucun mot ou aucun ensemble de mots dans une autre langue et le fait que seuls les sous-ensembles du sens d'un mot d'une langue correspondent à ceux de l'autre langue. Par exemple, pour le mot français *conducteur* il n'existe pas de mot anglais qui serait un équivalent global ; ce n'est qu'un sous-ensemble du sens du mot qui présente une équivalence : le sens « qui permet le passage d'un courant électrique » peut être traduit par *conductor*, le sens « personne qui conduit une voiture » par *driver*, etc.

D'une certaine manière, ce côté arbitraire de la correspondance sémantique entre les langues dénote un arbitraire sémantique au sein d'une langue quelconque. Par exemple, le mot *prendre* a 43 sens différents selon le dictionnaire *Le Robert* et il est difficile d'imaginer que cette correspondance entre le sens et le mot puisse être logique et universelle.

Le fait qu'un mot désigne plusieurs sens peut s'expliquer par les contraintes imposées à une langue. Le système phonétique d'une langue est construit sur des contraintes physiologiques. Il existe également une limite supérieure pour la longueur des unités langagières, et par exemple pour la longueur des mots. Etant donnée ces contraintes, il est donc inévitable qu'il y ait des ambiguïtés sur le sens des mots dans une langue donnée. Outre ces contraintes, il en existe encore une

autre : contrairement au langage artificiel, il existe des contraintes diachroniques. Autrement dit, la langue se forme dans le temps ce qui rend la sémantique d'une langue plus ou moins arbitraire, en contrastant avec le langage artificiel dont la sémantique est bien définie au départ et, le cas échéant, lors de sa modification. Cette dernière contrainte et les autres expliqueraient les différences de découpages sémantiques entre les langues comme *persuade* / *persuader*+*convaincre*.

Un examen plus précis sur la langue montre toutefois que ces contraintes ne sont pas les causes principales qui enferment une langue dans son ambiguïté. Au contraire c'est plutôt la langue même qui *préfère* la contrainte. Les contraintes seraient non pas les causes mais les résultats de la nature même de la langue. Le fait que les langues utilisent une petite partie du nombre limité des phonèmes possibles justifie l'argument selon lequel une langue pose des limites plus fortes que les contraintes possibles. En d'autres termes, la limite réelle du système phonétique d'une langue est plus restrictive que sa limite potentielle, à savoir sa limite physiologique. En ce qui concerne les mots, le nombre de mots utilisés dans la vie quotidienne est très limité par rapport au nombre total de mots existant dans une langue. En ce sens, les utilisateurs d'une langue ou le système de la langue préfèrent utiliser un nombre assez limité de mots. Ces deux faits dévoilent qu'il existe une stratégie économique dans le système du langage : utiliser un nombre limité de mots et un nombre encore bien plus limité de phonèmes.

Le fait que cette stratégie soit générale à travers les différentes langues suggère que l'avantage de cette stratégie récompense largement la difficulté issue de l'imprécision et de l'ambiguïté. Vu sous un autre angle, ce problème n'est pas un vrai problème chez l'homme contrairement à ce qui se passe avec une machine et il existe d'autres façons de penser la sémantique des mots que l'approche logique.

3.1 Langage formel et langue

Au moins depuis Aristote, il existe une tendance à juger le degré de rigueur d'un argument en utilisant la logique. Cette tendance s'accélère, avec le développement d'une mathématique et d'une logique formelles, à tel point qu'au siècle dernier la langue a été considérée comme « imparfaite » par Frege (1892/1952) et « imprécise et ambiguë » par Russel (1923). Certes, la langue est très différente de la logique formelle : de nombreuses caractéristiques de la langue se distinguent de celles de la logique. Le problème de l'ambiguïté, par exemple, est un des problèmes bien connus. Les exemples de Kripke et Putnam (chaleur, lumière, eau, etc.) présentés dans le chapitre précédent montrent les divergences d'un même sens/mot envers les différentes références et les exemples de Frege (l'étoile du matin/soir) associe la même référence à des sens différents.

En ce sens, l'argument qui maintient que la langue est imparfaite implique en effet que la logique puisse exprimer ces relations de manière plus précise et efficace que la langue. En effet, Russell a dit que c'est le rôle du philosophe de trouver un langage logique idéal. Cette thèse reste néanmoins à discuter, car une définition fine et exacte comme la logique ne garantit pas toujours que des avantages. Par exemple, bien qu'il soit possible d'éliminer le problème de l'ambiguïté de la polysémie ou de l'homonymie en faisant correspondre différents termes à chaque sens différent d'un mot, le nombre de termes nécessaires pour communiquer se multiplierait et créerait ainsi des problèmes nouveaux aux utilisateurs. Cette solution nécessiterait plus de ressources dans le lexique mental (le nombre d'entrées étant augmenté), plus de temps pour le processus d'accès au lexique et de récupération des éléments du lexique mental, etc.

Outre ces problèmes, langue et langage formel diffèrent également dans la façon de traiter l'information. Le fait qu'il existe plusieurs sens différents d'une part et qu'il n'existe pratiquement pas de problème d'ambiguïté lors de la communi-

cation normale entre les utilisateurs d'une langue montre que l'homme a, contrairement à la machine, une grande capacité à discerner les sens d'un mot dans un contexte donné. Cette capacité de choisir un sens parmi des sens différents est égale à la capacité de choisir un mot parmi des mots différents qui représentent les sens correspondants. A titre d'exemple, prenons ces mots suivants :

- (1) *aborder, éprouver, exiger, recevoir, conquérir, inscrire, s'adjoindre, absorber, durcir, attacher*

Pour un locuteur faire le choix d'un mot parmi la liste des mots (1) pour exprimer quelque chose n'est ni compliqué ni difficile. Il en va de même du côté interlocuteur pour saisir le sens d'un mot énoncé parmi (1). Considérons maintenant un mot M dont les sens sont les suivants :

- (2) M : ABORDER, ÉPROUVER, EXIGER, RECEVOIR, CONQUÉRIR, INSCRIRE, S'ADJOINDRE, ABSORBER, DURCIR, ATTACHER,

Tout d'abord, il n'est pas facile de deviner ce mot M qui est pourtant un mot français. De même, il sera difficile de choisir un sens parmi d'autres dans (2) si ce mot M est employé dans un énoncé. C'est en effet le fameux problème de la désambiguïsation des mots. Sans aucun contexte donné, on rencontre ce genre d'embarras exactement comme les machines. Maintenant, considérons ces phrases :

- (3) a. Ta « conversation » *exigera* beaucoup de temps ?
b. Il a *reçu* le ballon en pleine figure.
c. L'autre le regarda, et pensa que c'était toujours ça de *conquis* sur l'ennemi.
d. Mayonnaise, crème, gelée qui *durcit*.
e. Aliment qui a *attaché* au fond de la casserole.

Il ne sera pas difficile de saisir le sens des mots en italiques. La question est : si l'on remplace tous les mots en italiques par M, pourrions-nous saisir le sens du mot

M aussi facilement que les cas précédents ? La machine aura sûrement une grande difficulté, quel que soit le mot réel qui remplacerait le mot M. En revanche, pour l'homme, si l'on substitue M par *prendre*, il n'y aura aucune difficulté d'identifier et de saisir le sens, même si la liste des sens (2) n'a pas été présentée.

Par ailleurs, sous cet angle, cette capacité permet à l'homme de rétablir les informations perdues ou non-transmises, en rendant le système de la langue plus efficace qu'un système totalement désambiguïsé. D'une certaine manière, cette capacité est liée à une faculté particulière autre que la sélection automatique dans la liste des sens d'un mot. Cette capacité permettrait à l'homme de choisir correctement le mot même s'il ne l'a pas réellement aperçu dans une phrase. Par exemple, considérons ces phrases :

Jean a réservé une place de train hier mais il s'est levé trop tard ce matin. Il n'a pas pris son petit déjeuner et il a couru jusqu'à la gare. Il est arrivé à temps et il a (...) son train heureusement.

On peut facilement supposer que le mot manquant dans ces phrases est *pris*. La probabilité augmentera si l'interlocuteur a entendu des sons similaires tels que /bri/ ou /pRy/. Or, cette tâche est loin d'être évidente pour une machine. On note ici une différence essentielle entre l'homme et la machine et entre la langue et le langage formel.¹ Cette capacité de l'homme, remarquable par rapport à celle de la machine, suggère que la nature « imparfaite » et « imprécise » de la langue est liée à cette capacité. Ce lien peut être pensé de deux manières différentes.

D'abord, cette capacité peut être considérée comme le résultat de la langue. C'est-à-dire que l'homme peut développer cette capacité parce qu'il est tout à fait

¹On pourra opposer à cet argument concernant la comparaison homme/machine qu'une approche statistique comme le modèle de Markov peut permettre de deviner un mot dans ce type de situation. Il est vrai que ce genre de modèle statistique simule bien mieux la performance humaine que le modèle symbolique et en ce sens il est plus proche de l'homme. Cependant ici *la machine* désigne un système conventionnel basé sur la logique symbolique.

habitué à choisir le sens spécifique d'un mot parmi plusieurs sens possibles qui sont posés par le système de la langue.

Sous un autre angle, cette capacité pourrait être la cause du système de la langue. La levée des ambiguïtés permettrait à une langue d'utiliser une stratégie plus économique, comme l'usage des moindres mots possibles, contrairement à un système de langage formel qui exige autant de mots que d'entités (sens totaux).

Ce qui est commun aux les deux cas c'est que la nature « imparfaite » et « imprécise » de la langue est un avantage et non un défaut comme le pense Frege ou Russell.

Cet argument sera discuté, de manière plus détaillée, plus loin dans ce chapitre.

3.2 Une critique relative aux propositions de Pustejovsky et Jackendoff

La capacité langagière chez l'homme, qui contraste avec des réalisations comme les dictionnaires ou des systèmes d'I.A., demande une théorie ou un modèle qui puissent rendre compte de ces différences. La théorie du lexique génératif de Pustejovsky est une de ces tentatives. Cette théorie a pour objectif explicite de *prévoir* l'usage créatif d'un mot, ce qui est impossible dans une approche traditionnelle telle que l'énumération des sens.

La théorie du lexique génératif tente de surmonter l'approche énumérative du sens des mots telle que *bien fonctionner* pour *good* dans *good umbrella*, *bien faire des actes* pour *good teacher* et *savoureux* pour *good meal*, etc. D'après Pustejovsky, ces différents sens du mot *good* étant fixés par les noms que ce mot accompagne, les différents sens de *good* égaliseraient le nombre des types distincts sur lesquels cet adjectif s'applique dans la langue (Pustejovsky 1995). De là découle, l'impossibilité d'utiliser, dans ce but, l'approche traditionnelle du sens du

mot à savoir l'énumération des sens. Afin de résoudre ce problème, Pustejovsky propose la théorie du lexique génératif selon laquelle le sens de *good* par exemple, est engendré par le « telos » du nom modifié. De même, d'après Pustejovsky, c'est le *telos* de *garage* qui permet d'utiliser (ou bien de rendre compréhensible) l'expression *fast garage* où le sens de *fast* n'est pas couvert par l'énumération des sens des dictionnaires (le sens qui fait appel à la longueur du temps au lieu de « se déplacer rapidement », « faire un acte rapidement », etc.).

L'incapacité de rendre compte de ce type de sens créatif au sein de l'approche du lexique d'énumération du sens est, d'après Pustejovsky étroitement liée à la rigidité de la frontière du sens lexical dans l'approche traditionnelle. Pustejovsky maintient que le chevauchement des différents sens de CHANGEMENT D'ÉTAT et de CRÉATION de *bake* dans les phrases « John baked the potatoes » et « Mary baked a cake » n'est pas respecté dans l'approche traditionnelle. De même, le sens OUVERTURE et OBJET PHYSIQUE de *window* dans les phrases « John crawled through the window » et « The window is closed » se croisent et le sens OBJET et ENDROIT pour le mot *noisy* dans « a noisy car » et « a noisy room » se recouvrent, ceci n'est pourtant pas représenté par l'approche énumérative du sens des mots.

En ce qui concerne la limite de l'approche par l'énumération, nous partageons le même point de vue que Pustejovsky mais nous estimons que sa solution comporte d'autres problèmes : même si cette théorie refuse d'énumérer les sens d'un mot et essaie de garder ses distances par rapport à l'approche définitionnelle, elle nous paraît aussi être une prisonnière de ce dernier.

D'abord, le « telos » d'un mot est une sorte d'approche définitionnelle dans la mesure où le *telos* impose de manière impérative à un mot une propriété constante. Afin d'expliquer les différents sens de *bake* dans *bake a cake/potato* par exemple, Pustejovsky oppose le *telos* de *potato* et celui de *cake* par la présence ou non du trait « objet fabriqué ». Bien que cette manière de discerner les mots soit différente

3.2 Une critique relative aux propositions de Pustejovsky et Jackendoff 53

de la définition, elle attribue néanmoins à un mot un caractère constant et défini. En ce sens, cette façon d'aborder le sens du mot partage une stratégie commune avec la théorie définitionnelle du concept qui définit par exemple *bachelor* comme « homme non marié ». Comme c'était le cas pour la définition de *bachelor*, il suffit de quelques contre-exemples pour infirmer la méthode (par exemple, voir Fodor 1998).

D'ailleurs, la manière dont un adjectif ou un verbe se combinent avec un nom en engendrant un sens particulier est en réalité *post hoc* plutôt qu'une vraie prédiction. On peut imaginer des expressions telle que *fast stomach* ou *fast rice* qui sont conformes à la théorie du lexique génératif : puisque *stomach* a pour telos *digérer*, *fast stomach* peut être interprété comme *estomac qui digère rapidement* et, de la même manière, *fast rice* comme *riz qui pousse rapidement*. Cependant, il est tout à fait possible aussi d'utiliser, éventuellement, *fast rice* pour désigner *riz qui se cuisine rapidement*. Le sens que *fast rice* prendra éventuellement parmi les deux sens n'est pas déterminé ou prévu par l'application des règles fixées par Pustejovsky. Autrement dit, les exemples pris par Pustejovsky pour la théorie du lexique génératif peuvent être *post hoc*.

Attribuer cet échec éventuel uniquement au telos des noms (*rice* dans l'exemple ci-dessus) s'avère faux si la comparaison est menée entre des langues différentes. Plus précisément, si cette théorie n'est pas capable d'expliquer la raison pour laquelle une expression équivalente dans une autre langue devient incompréhensible ou bien considérée comme mal formée, le telos d'un mot ne peut plus être l'unique responsable. A titre d'exemple, prenons le mot anglais *wife* et le mot équivalent français *femme*.² Tandis que *his good wife* est un bon exemple pour la théorie du

²*Épouse+femme* semble à première vue équivalent à *wife* mais comme l'usage du terme *épouse* est très réservé plus spécifique que l'usage du terme *femme*, il suffira pour le moment de prendre *femme* comme mot équivalent.

lexique génératif, l'expression équivalente en français *sa bonne femme* ne l'est pas.³ Il faudra supposer dans ce cas qu'il existe une différence soit entre *wife* et *femme* soit entre *good* et *bonne*. Ce dernier cas est exclu en principe chez Pustejovsky car l'essence de sa théorie consiste en la capacité d'expliquer les sens différents de manière générative. Par conséquent, il ne reste qu'à différencier *wife* et *femme* afin de maintenir la cohérence de la théorie. Mais il paraît difficile de distinguer parmi leurs différences celles qui permettraient d'expliquer clairement leur usage convenable en présence des mots *good/bon*.⁴ Bien que cette tentative de distinguer deux mots réussisse, cette théorie se confrontera à une critique encore plus sévère. Car, cette distinction extrêmement subtile est plus descriptive et plus comparative que la linguistique descriptive et comparative même en général, alors que la théorie du lexique génératif a pour but de surmonter ce genre de problème de la linguistique traditionnelle.

Outre la caractéristique multivoque des sens du mot dans la sémantique de la langue, une autre caractéristique sémantique à savoir le sens prototypique ou l'usage stéréotypique d'un mot a mis en cause l'approche définitionnelle traditionnelle. Jackendoff tente de répondre à ce problème en utilisant le mot *climb* examiné par Fillmore (1982). Jackendoff explique l'usage stéréotypique de *climb* avec les conditions de satisfaction. D'après lui, le mot *climb* implique deux conditions : celle d'« un individu qui se déplace vers le haut (condition 1) » et celle d'« un individu qui se déplace en s'agrippant (alpinisme) (condition 2) ». Ces condi-

³En français *sa bonne femme* est, contrairement à *his good wife*, une expression très péjorative. Les expressions *bon ami* et *good friend* présentent aussi une différence : la première évoque parfois le sens PETIT AMI contrairement à la dernière.

⁴Ceci est plus clair dans l'exemple *good friend / bon ami* où il est difficile de trouver une différence entre *friend* et *ami*. Le fait que le concept équivalent à FEMME est WIFE + WOMAN ne modifie pas cet argument, car le sens de *femme* est fixé à WIFE et non WOMAN dans l'expression *sa bonne femme*.

tions sont utilisées pour le jugement de la typicalité de l'usage du mot : l'usage de *climb* dans la phrase « Bill climbed (up) the mountain. » est stéréotypique puisqu'il satisfait à la fois la condition 1 et la condition 2 ; par contre, l'usage de ce mot dans la phrase « Bill climbed down the mountain » ou dans la phrase « The snake climbed (up) the tree » est considéré comme marginal puisqu'il ne satisfait qu'une seule condition (condition 2 et condition 1 respectivement). De même, l'emploi du mot *climb* n'est pas correct dans la phrase « ?*The snake climbed down the tree », puisqu'il viole les deux conditions.

Ainsi, Jackendoff suppose une typicalité du sens d'un mot par le nombre des conditions. Considérons maintenant une langue où il existe un mot M_1 dont le sens serait « le mouvement d'un alpiniste qui grimpe une montagne », un mot M_2 pour « le mouvement d'un serpent qui gravit une montagne » et un mot M_3 « le mouvement d'un alpiniste qui descend une montagne ». Imaginons maintenant le concept C qui comprend les sens de M_1 , M_2 et M_3 . Ce concept C est donc équivalent au concept *climb*. Le mot qui sera utilisé dans la phrase équivalente à la phrase « Bill climbed (up) the mountain. » ne peut néanmoins pas être stéréotypique du concept C , puisqu'il existe trois mots différents pour le concept C . Certes, il est possible, dans cette langue, que M_1 soit le plus prototypique parmi M_1 , M_2 et M_3 , au regard du concept C . Mais le critère selon lequel M_1 devient prototypique du concept C est radicalement différent du critère de Jackendoff qui suppose une analyse des conditions. En effet, la manière selon laquelle Jackendoff évalue la typicalité de l'usage d'un mot consiste en trois étapes. D'abord il faudra analyser le sens d'un *mot* et non du *concept*, par la suite, il faudra établir les conditions et finalement il faudra revenir sur les mots en utilisant ces conditions. Ainsi, cette théorie, contrairement à la théorie du lexique génératif, ne peut aucunement prétendre qu'elle prévoit les sens des mots.

D'ailleurs, le problème de base de cette théorie est qu'il n'existe pas de condi-

tion pondérée. C'est-à-dire que pour Jackendoff, le sens stéréotypique est proportionnel à la participation simultanée des différents sens/usages du mot qu'il a séparé. Ceci présuppose donc une contribution équivalente de chaque sens. Toutefois, pour un mot donné, le sens dominant pourrait être une représentation qui attribue un poids plus important à une valeur stéréotypique. A titre d'exemple, imaginons un mot dont les conditions peuvent être classées en trois critères comme *climb* mais dont seul le mot conforme à la seule condition 1 possède une représentation plus stéréotypique. Dans ce cas, l'argument de Jackendoff est erroné, car cette situation viole la supposition initiale.

Une approche plus moderne, comme la théorie du lexique génératif ou la théorie de Jackendoff, tente de surmonter la limite de cette approche énumérative. C'est pour cette raison qu'il est nécessaire d'avoir une théorie qui peut élucider l'usage des mots d'une manière générale et universelle sans recourir à la spécificité descriptive.

3.2.1 Le problème renversé

En général l'approche de Pustejovsky et celle de Jackendoff, de même que de nombreuses approches sémantiques, partagent un point commun. C'est d'essayer d'analyser de manière exhaustive les différents aspects ou les différentes natures du sens des mots. Cependant, malgré leurs théories profondes, le lien censé être direct pour l'amélioration de la tâche du traitement automatique des langues (TAL) n'a pas été atteint. Et pour cela, des auteurs comme Wilks critiquent sévèrement Jackendoff.

Toutefois, la simple absence de relation directe entre la théorie et le modèle pratique n'est pas une condition suffisante pour invalider une théorie. Comme l'a défendu Jackendoff face à cette critique féroce, le décalage entre la théorie et l'application est acceptable dans la mesure où il n'est pas prouvé que la théorie

est fausse.

A part l'impossibilité d'application de la théorie de Pustejovsky et Jackendoff, nous voulons indiquer une limite sous-jacente à leur approches. Pustejovsky classe en quatre niveaux de représentations (structure d'argument, structure d'événement, structure qualia, structure d'héritage lexical), système génératif (*type coercion, selective binding, co-composition*). Ces niveaux sont encore divisés en quatre aspects essentiels comme : *constitutive, formal, telic et agentive*. En utilisant une règle de préférence, une structure d'argument, des rôles thématique et les fonctions conceptuelles et spatiales, etc. Jackendoff distingue différents aspects des mots de fonction comme *with, for, in, to, from, etc.*

Nous pensons que ce genre d'outils détaillés pour analyser le sens inverse le vrai problème et nous proposons d'examiner la sémantique des mots en inversant le point de vue.

La théorie du lexique génératif a pour but, parmi d'autres, d'expliquer l'usage original d'un mot donné. Cependant, l'usage original ne permet pas n'importe quelle composition de mots, comme nous venons de le voir. D'ailleurs, on ne peut pas toujours prévoir dans quel sens, au sein d'une expression, un mot va s'engager, de même qu'il n'est pas possible de prévoir dans quelle direction tombe un crayon posé verticalement, comme l'explique la théorie du chaos.

Nous supposons l'existence de concepts tels que BOIRE, SUCER, MANGER, DÉVORER, MARCHER, TITUBER, COURIR, SPRINT (anglais), IRRESOLVABLE (anglais), etc. Ces concepts n'existent toutefois pas en tant que tels mais ils sont liés au monde (événement, action, situation, etc.) ou aux relations entre les concepts.

De manière plus rigoureuse, ces concepts de base peuvent encore être décomposés. Ce sous-concept de base correspond en effet à un élément pour lequel il n'existe plus la moindre ambiguïté. En revanche, un mot de la langue correspond souvent à un ensemble de concepts de base et bien moins fréquemment au concept

de base lui-même.

Considérer un mot comme un ensemble de concepts de base suppose la création des sens du mot par les différentes relations (par rapport au monde, entre les mots). Nous supposons de plus que ces liens sont arbitraires. Par exemple, le verbe qui décrit l'habillement (vêtements, chapeau, chaussettes) peut être ou bien décrit par des mots différents, comme en coréen, ou bien par un mot unique, comme en anglais et français.

Ce point de vue sur le sens des mots nous décharge de la nécessité de construire des liens logiques entre les sens d'un mot. En effet, puisque les différents sens d'un mot sont, de ce point de vue, les résultats d'une pure coïncidence, cette tentative de reconstruction des sens devient non essentielle. Comme nous l'avons dit plusieurs fois, cette tentative de construction du sens vient du piège langagier. Par exemple, bien que le lien entre *paper* comme PAPIER et *paper* comme ARTICLE soit tenu pour fixe et sûr pour certains auteurs anglophones, ce lien est beaucoup moins évident en français puisqu'il existe deux mots indépendants. De fait cette observation met donc en cause la nature de la polysémie puisque *paper* est un des exemples les plus typiques de la polysémie.

3.3 Polysémie contre homonymie

La dichotomie polysémie/homonymie distingue les mots selon le degré de relation sémantique des sens. Dans la lexicographie cette distinction est souvent liée à l'ontologie des mots. Si le graphème a une seule étymologie, les différents sens de ce graphème sont considérés comme des variations du sens d'un mot — ce graphème est alors polysémique même si les sens sont relativement différents ; sinon les différents sens d'un même graphème sont considérés comme homonymiques. Le partage d'un même graphème par les différents sens est alors pure coïnci-

dence. Ainsi, dans un dictionnaire normal les entrées respectent cette convention : différentes entrées pour les homonymes et une même entrée pour les termes polysémiques.⁵

Selon ce point de vue, *conducteur* est polysémique puisque les différents sens de ce mot partagent la même origine latine *conductor* ce qui est d'ailleurs reflété par une même entrée dans le dictionnaire du français *Le Robert*. Toutefois, certains auteurs mettent cette tradition en question en constatant que le lien historique ne reflète plus les usages ordinaires actuels (Palmer 1981). Ainsi, cette dichotomie polysémie/homonymie n'est pas acceptée de façon unanime. Plusieurs preuves en psycho-linguistique montrent aussi qu'il n'existe pas de distinction absolue entre polysémie et homonymie (Klein et Murphy 2001). En effet, il n'est pas toujours facile de dire que les liens du sens, pour un mot polysémique, sont plus significatifs que ceux de mots différents. Par exemple, le sens PAPIER et ARTICLE de *paper* seront moins reliés dans le contexte de la publication que *paper* et *publish*.

L'admission de la notion de polysémie — c'est-à-dire polysémie en tant que mot ayant différents sens au lieu de plusieurs mots séparés — présuppose que les mots différents, quant à eux, possèdent des sens plus éloignés entre eux par rapport à la polysémie. Considérons un mot *M* qui possède deux sens S_1 et S_2 . Si les sens S_1 et S_2 ne partagent aucun lien ni aucune origine historique, *M* est classé dans la catégorie de l'homonymie et peut se décrire comme un ensemble de deux mots M_1 et M_2 possédant respectivement les sens S_1 et S_2 . Dans ce cas, M_1 et M_2 sont des mots différents, même s'ils partagent le même graphème et la même phonétique. En revanche, s'il existe un lien entre S_1 et S_2 de nature soit sémantique soit historique, *M* reste un mot unique. Cette distinction présuppose que les mots

⁵Certains dictionnaires adoptent différentes approches, notamment le *Dictionnaire du français contemporain (DFC)* et Jean Dubois, Françoise Dubois-Charlier *Dictionnaire de la langue française* qui présentent les entrées des mots de manière homonymique.

différents sont plus éloignés entre eux qu'un seul mot, même s'il existe des sens très différents pour ce mot. Par exemple, l'affinité du mot M et d'un autre mot W — ou bien plus précisément la similitude de sens entre le mot M et le mot W — est plus éloignée que celle de S_1 et S_2 . Or, il est difficile d'admettre ce point de vue si on compare le sens « qui n'a pas de sensibilité morale ; qui n'a pas ou a peu d'émotions » de *insensible* soit avec l'autre sens de *insensible* tel que « inanimé » soit avec d'autres mots tels que *froid* : dans ce contexte, le sens entre *insensible* et *froid* est plus proche que le sens entre « qui n'a pas de sensibilité morale » et « qui n'a pas de sensibilité physique ».

La comparaison entre les langues met plus clairement en évidence ce problème de dichotomie polysémie/homonymie. Les sens PAPIER et ARTICLE du mot *paper* sont considérés comme liés en anglais puisque ce mot est souvent utilisé en tant qu'exemple de polysémie, qui se distingue de l'homonymie (à savoir les mots différents). Or, en français ce lien étroit entre PAPIER et ARTICLE n'est pas explicité car contrairement à l'anglais, il existe deux mots différents en français : *papier* et *article*. Ainsi, le lien qui est considéré comme étroitement lié dans une langue est moins accessible dans une autre.

3.4 Proposition

Ceci met en cause non seulement la pertinence de la dichotomie polysémie/homonymie mais révèle aussi la nature de la polysémie. Le fait que l'on accepte l'homonymie pour différents mots nie en effet le lien logique entre le sens et graphème ou phonétique. Nous venons de montrer que le lien entre différents mots n'est pas toujours moins fort que les liens internes d'un mot polysémique. En combinant ces deux faits on pourrait être amené à en déduire que la polysémie pourrait être considérée comme homonymique en associant des mots différents

à des sens différents. Ce point de vue renvoie à la théorie du concept que nous avons proposée dans le chapitre précédent et qui consiste à supprimer la frontière extérieure et intérieure d'un mot pour extraire les concepts de base .

Le principe de base qui consiste à enlever la frontière intérieure d'un mot, notamment polysémique, est supposé garantir l'indépendance des sens distincts d'un mot. Ceci ne veut pas dire que les sens sont indépendants mais plutôt que le lien doit être examiné après avoir séparé ces sens et les liens entre mots. L'essence de cette idée est donc de considérer un mot comme un ensemble de sens au lieu de le considérer comme un amalgame inséparable.

Cette question est liée également à la représentation de la polysémie/homonymie qui se subdivise par un sens noyau ou des sens séparés. La position du sens « noyau » suppose que le sens « noyau » d'un mot polysémique et les autres sens de ce mot sont expliqués par les liens qu'ils ont avec ce sens « noyau ». En revanche, la position des sens séparés refuse cette notion de sens « noyau » et explique les différents sens d'un mot par la multi-présence des sens plus ou moins indépendants. Les preuves de chaque théorie se retrouvent donc par les expériences des effets d'amorçage psycho-linguistique qui mesurent les temps de réaction et les réponses correctes des sujets. Certaines expériences montrent qu'il n'existe pas de différence de traitement entre la polysémie et l'homonymie suggérant que, pour les représentations mentales, polysémie et homonymie ne sont pas très différentes (Nunberg 1979, Caramazza et Grober 1976, Lehrer 1990, Klein et Murphy 2001).

D'ailleurs il existe aussi des débats sur la distinction entre sémantique et connaissance encyclopédique. Par exemple, tandis qu'il semble clair qu'il existe un lien sémantique entre *philosophie* et *philosophe* et un lien encyclopédique entre *philosophe* et *Aristote*, la nature du lien entre H_2O et *eau* est assez problématique. Il peut être considéré comme un lien soit sémantique soit encyclopédique.

Répondre à cette question conduit à mettre en question une dichotomie sémantique/encyclopédique. Si le lien entre H₂O et *eau* est sémantique puisque l'eau est définie comme « composée de molécules de H₂O », le lien entre *philosophe* et *Aristote* pourrait être expliqué de la même manière : puisqu'Aristote est défini comme « Aristote est le philosophe grec qui a écrit *Éthique à Nicomaque*, *Politique*, *Physique* et *Métaphysique* », le lien sémantique pourra être établi par cette définition.

Nous partageons ce point de vue, à savoir qu'il n'existe pas de différence significative et représentationnelle entre polysémie et homonymie et entre la connaissance sémantique et la connaissance encyclopédique. En effet, il existe une chose essentielle sous-jacente à tout ce qui est présenté jusqu'ici — comme la priorité des conditions nécessaires ou suffisantes / usage, dichotomie du lien sémantique/encyclopédique, polysémie/homonymie : ce sont les *relations entre les mots*.

3.4.1 Lien contextuel

Le commentaire de Firth « You shall know a word by the company it keeps »⁶ décrit métaphoriquement la nature du sens des mots qui est décidé par les mots qu'il *fréquente*. En effet, c'est Malinowski qui a présenté pour la première fois le terme *contexte de situation*. Il a fait remarquer qu'un mot n'a pas de sens en-dehors d'une situation spécifique chez les tribus Trobriand. Par exemple, le mot *bois* prend un sens spécifique PAGAIE POUR LES CANOËS si et seulement si le discours est adressé dans une situation liée au canoë (Malinowski 1923).

Comme il a été démontré ci-dessus, de nombreuses théories ne peuvent pas prévoir l'usage des mots mais plutôt elles les décrivent d'une certaine manière. Le problème courant de ce genre de théories est qu'elles sont loin d'atteindre

⁶Cette phrase parodie le proverbe anglais « A man is known by the company he keeps. » (« Dis-moi qui tu hantes, je te dirai qui tu es »).

la généralité et l'exhaustivité. Ces théories présentent certains exemples afin de valider leurs propositions mais elles sont vulnérables et facilement infirmées par des contre-exemples. De plus, le point définitivement faible de ce genre de théories est qu'il ne semble pas envisageable d'implémenter une application automatique basée sur ces théories. Par exemple, afin de rendre compte des conditions chez Jackendoff, un travail supplémentaire est indispensable ; afin d'utiliser la structure proposée dans la théorie de lexique génératif, un même travail est nécessaire.

Pour répondre à ces deux questions — une théorie générale et prête à être appliquée — nous proposons une théorie qui peut *expliquer* et *simuler* les relations entre les mots. Nous nous concentrerons d'abord sur le problème de la construction des relations entre les mots. En effet, les erreurs chez Jackendoff ou chez Pustejovsky ne viennent pas du fait qu'ils ont négligé les relations réelles entre les mots mais plutôt de celui qu'ils ont privilégié en développant leurs théories *certaines* relations entre les mots et non *toutes* les relations.

Afin d'éviter ce genre de piège, nous prenons une voie cartésienne — la méthode mais pas l'ontologie — c'est-à-dire, nous voulons nous priver des notions linguistiques généralement utilisées : synonymie, antonymie, polysémie, homonymie, sémantique/encyclopédique, générativité, condition nécessaire et suffisante, etc.

Maintenant concentrons-nous sur la construction des relations entre les mots de manière fiable. Avant d'aborder ce sujet directement, il sera indispensable de décider quelle relation on doit choisir entre les relations de mots, unités minimales de sens ou unités minimales de concept, etc. En chinois, les mots sont en général monosyllabiques et en indien d'Amérique du Nord les mots ressemblent parfois à la construction d'une phrase. Ceci montre que même le terme *mot* n'a pas de signification unifiée pour toutes les langues.

Le terme *mot* n'est pas convenable même dans une langue donnée. Par exem-

ple, tandis que *garçon* est un mot, *pomme de terre* est composé de trois mots. En ce sens, un *lexème* est une unité linguistique plus rigoureuse au sein des études sémantiques : *pomme de terre* est considéré comme un *paralexème* ou un *syntagme lexicalisé* au lieu de trois lexèmes. Cependant, le problème reste toujours si l'on considère un mot comme *porte-bonheur* : en effet il n'y a pas de différence essentielle entre *pomme de terre* et *porte-bonheur*, mais ce premier est considéré comme paralexème alors que ce dernier est un lexème.

La notion de lexème regroupe la forme fléchie *faux* du verbe *faillir* et le participe *failli* parce qu'ils appartiennent au même lexème. Il est utile néanmoins de garder la notion de mot qui distingue *faux* de *failli*. Ce que nous voulons dire dans cette comparaison, c'est qu'il existe une différence entre *aime/aimé* et *faillir/failli* : parce que la forme *faux* est rarement utilisé, il peut être supposé que *failli* entraîne *faux* beaucoup moins systématiquement que *aimé* n'entraîne *aime* et il sera souhaitable, dans certains cas, de ne pas faire un amalgame entre ces deux mots *faux* et *failli*.

Pour cette raison, nous utilisons des unités de type *mot* au lieu de *lexème* dans la discussion suivante.

Nous pensons que des relations entre mots peuvent être construites si l'on observe l'usage des mots. Par exemple, une expression figée comme *au fur et à mesure* garantit que *fur* est co-occurent de *mesure* si c'est le cas dans suffisamment de textes écrits ou de paroles observés car, à moins qu'une expression ait disparue, elle doit être utilisée soit à oral, soit dans les textes. De même, la cooccurrence entre *tomber/tombé* et *pommes* sera relevée par rapport à l'anglais *fall* et *apples* puisqu'il n'existe pas en anglais d'expression équivalente à *tomber dans les pommes*. Ce genre de relation peut s'appliquer au-delà du cas des expressions figées comme par exemple pour *pomme* et *rouge*, *boire* et *alcool* ou *payer* et *chèque*. On peut se poser la question de savoir si cette relation peut s'étendre aux

relations stéréotypiques comme *ivrogne* et *tituber/titubé* ou *neige* et *froid*. Nous pensons qu'en principe ce genre de relation peut être aussi saisi.

Si le lien entre *ivrogne* et *tituber* est évident, cette relation va être répétée dans des conversations ou dans des textes comme les fictions, les romans, etc. Réciproquement, cette relation observée dans des textes ou des conversations nous influence pour constituer la relation.

En effet la relation entre *ivrogne* et *tituber* reflète une relation de contexte et ces deux mots sont des mots liés de manière contextuelle. Nous pensons que l'essence d'un mot se trouve dans cette relation. En effet, il est impossible d'examiner un mot sans tenir compte des mots qui l'accompagnent dans un contexte donné. Cette relation fournit des liens plus riches. Le synonyme ou l'antonyme sont considérés comme des cas spécifiques de mots liés de cette manière.

Une critique peut être soulevée concernant ce rôle contextuel notamment pour la désambiguïsation. En effet, Hirst (1987) a fait remarquer un piège lié à la dépendance contextuelle par cette phrase :

(4) The lawyer stopped at the bar for a drink.

En effet en présence du mot *avocat* le mot ambigu *bar* a de fortes chances d'avoir le sens BARRE et non BISTROT, mais dans cet exemple ce n'est pas le cas.

Nous pensons à deux arguments concernant ce type de problème.

D'une part, bien que l'inférence tirée par la relation entre *lawyer* et *bar* soit erronée, le fait que l'on puisse identifier le sens de *bar* à l'aide du mot *drink* montre en effet que le lien contextuel entre *bar* et *drink* est utile.

D'autre part, si la phrase (4), provoque d'une certaine manière une interprétation inattendue, il serait souhaitable d'avoir un système qui peut imiter ce type d'interprétation. Une phrase comme la suivante déclenche cet effet de manière un peu plus forte :

(5) In the school, the teacher examined the pupil of the fainted student.

Un cas plus marqué que les phrases (4) et (5) sera le suivant :

(6) Seven days make one weak.

Face à cet énoncé, un système de reconnaissance de la parole, qui a entendu le son /wi :k/, peut choisir un mot entre *week* et *weak*. Si un système A choisit le mot *week* en faveur de *seven days*, il tombe dans un piège. Toutefois, on ne peut pas conclure qu'un autre système (système B) qui choisit le mot *weak* sans hésitation est plus intelligent. Car, l'intension du locuteur qui a énoncé la phrase (6), vise justement d'abord une interprétation du type de celle choisie par le système A suivie par une autre interprétation du type de celle choisie par le système B. En d'autres termes, l'interprétation du système A dans un tel cas ne devrait pas être abandonnée. Il sera plus évident si l'on compare l'énoncé (6) avec cette phrase :

(7) Hard work makes one weak.

Pour cet énoncé (7), contrairement au cas précédent, le fait qu'un système de reconnaissance de la parole se trompe en choisissant le mot *week* n'a aucune valeur intéressante. Ceci montre donc qu'un système qui se trompe *du fait du contexte* n'est pas quelque chose à éviter mais plutôt *conseillé* pour obtenir un système intelligent. Dans une certaine manière cet effet attendu par le locuteur vis-à-vis de son interlocuteur n'est pas très loin de l'effet perlocutoire étudié par Austin (1962/1970)

3.4.2 Contexonyme

Nous appelons *contexonymes* d'un mot les mots qui lui sont liés de manière contextuelle. Par exemple, *froid*, *hiver* et *blanc* sont les contexonymes de *neige* mais *bureau*, *cinéma*, *délicieux*, etc. ne sont pas des contexonymes de ce mot. Contrairement au synonyme qui a une définition concrète comme « un mot qui

peut remplacer un autre mot dans une phrase sans en changer la valeur de vérité », le contexonyme n'a pas ce genre de définition. Toutefois, afin de ne pas être piégé dans un jugement subjectif, il est inévitable d'établir un critère objectif qui décide d'inclure ou d'exclure un mot comme contexonyme ou non, pour un mot cible donné. Nous proposerons un critère dans le chapitre 5. Pour le moment, concentrons-nous sur la nature du contexonyme.

Si le contexonyme reflète les relations contextuelles d'un mot, il reflète plusieurs natures de ces relations y compris l'usage, la connotation et la nuance, le registre etc. Par exemple, pour un mot péjoratif, ses contexonymes reflèteront sa connotation ; pour un mot soutenu, familier, ses contexonymes reflèteront son registre, etc.

Les contexonymes sont plus dynamiques et plus flous que les synonymes ou les antonymes. Tandis que la définition du synonyme pose une contrainte très rigoureuse à satisfaire, il n'existe pas de telle contrainte à satisfaire pour les contexonymes, laissant ainsi plus de liberté. De plus, les contexonymes évoluent rapidement par rapport aux synonymes ou aux antonymes. Une des raisons de cette évolution rapide est directement liée à la propriété précédente, à savoir dynamisme et flou. Les relations complexes entre les contexonymes en sont une autre raison. Ce type de changement au cours du temps a été constaté par l'étude de la comparaison des corpus issus de différentes époques (Habert, Illouz et Folch à paraître).

De fait, comme l'a fait remarquer Bréal (1897/1982), le sens des mots devient tantôt restreint tantôt élargi. Pour la restriction du sens Bréal constate que « [Les] mots [sont] à sens général pour commencer, et ensuite réduits à une certaine catégorie d'objets ». Par exemple, le latin *tegmen* (marquer quelque chose) et ses proches *tegmentum*, *tectura*, *tegmentum*, *tectorium*, *teges*, *toga* se sont diversifiés, comme *toit* par exemple, qui ne s'emploie que pour les bâtiments contrairement à *tegmen*.

Pour l'élargissement du sens des mots, Bréal avance : « Les idées générales que l'humanité a acquises dans le cours des siècles n'auraient pu recevoir de nom sans cet élargissement du sens ». Par exemple, le *temps* était à l'origine « la température, la chaleur » et a obtenu le sens abstrait de durée.

En principe, nous pensons que les contexonymes devront refléter ce changement du sens.

Comme le contexonyme ne peut pas être défini de manière vérifiable, nous adaptons une stratégie inverse. C'est-à-dire que, au lieu de discerner les contexonymes au préalable de manière sémantique, comme c'est le cas pour les synonymes, nous choisissons les contexonymes *a priori* et nous établissons les liens sémantiques par la suite.

Notre hypothèse est que les contexonymes d'un mot peuvent être découverts si un corpus adéquat est utilisé. Par adéquat, nous entendons un corpus adapté à l'objectif de la recherche. S'il s'agit du sens général d'un mot, un corpus bien équilibré est adéquat. S'il s'agit du sens spécial d'un mot tel qu'un terme technique ou bien un usage scientifique ou encore un usage courant utilisé dans le langage parlé, un corpus adapté sera nécessaire.

Chapitre 4

Les modèles informatiques

Pour certaines tâches du traitement automatique des langues (TAL), une connaissance plus étendue que le seul sens canonique d'un mot est capitale. A titre d'exemple, voici le texte d'un journal¹ :

The final was Hewitt's first and Sampras' 17th, but the less experienced 20-year-old Australian was much more energetic. After consecutive wins against former champions Pat Rafter, Andre Agassi and Marat Safin, Sampras appeared to have nothing left for his second match in barely 24 hours.

Les mots ambigus *the final* et *match*, devraient être traduits en français par *la finale* et *match* respectivement, mais les systèmes les plus connus dans le domaine de la traduction automatique proposent comme traduction les mots *le final* ou *allumette*. Voici les deux traductions obtenues par les deux systèmes (les accentuations sont absentes dans leurs résultats originaux) :

**Le final* était Hewitt's premier et Sampras' 17e, mais l'a éprouvé moins 20-ans Australien était beaucoup plus énergique. Après consé-

¹Texte de CNN.

cutif gagne contre Rafter ancien de *Caresse* de champion, Andre Agassi et Marat Safin, Sampras a paru d'avoir rien ne part pour sa deuxième *allumette* dans à peine 24 heures.²

*La finale était Hewitt premier et Sampras 17ème, mais moins l'Australien 20-year-old expérimenté était beaucoup plus énergique. Après que les victoires consécutives contre d'anciens champions *tapotent le comble*, l'Andre Agassi et le Marat Safin, Sampras a semblé à gauche n'avoir rien pour sa deuxième *allumette* en à peine 24 heures.³

Ces deux traductions deviennent d'autant plus incompréhensibles par la présence des traductions erronés des mots clefs comme *match* et *Pat Rafter*. De manière similaire, la machine a des difficultés à résoudre le problème de l'anaphore dans les phrases suivantes :

(8) Bill didn't drink but Tom did. Quite a lot. He staggered when he was out.

(9) Tom drank a lot and Bill just a little bit. But he staggered when he walked !

où, l'usage fréquent de « stagger » pour *la démarche d'un ivrogne* assure que « he » désigne Tom dans (8) et Bill dans (9) (à cause du connecteur « but »). En effet, ce type de difficulté n'est pas réservé à la machine. Même les humains peuvent rencontrer une difficulté pour répondre à la question « Quel est le verbe anglais/français typique qui décrit la démarche d'un ivrogne ? » si la langue concernée n'est pas la langue maternelle du locuteur. De la même manière, la question « Quelle est la meilleure traduction de « He wrote an article yesterday » entre « Il a rédigé un article hier » et « Il a écrit un article hier » est difficile pour des locuteurs non francophones.

Cette série d'exemples qui touchent à l'association entre mots met en évidence la nécessité d'au moins quatre critères auxquels doivent répondre un mo-

²La traduction de *FreeTranslation*.

³La traduction de *Systrans / Babel Fish*.

dèle ou une méthode en linguistique computationnelle et en psycholinguistique. D'abord, pour la machine ainsi que pour un apprenant d'une langue donnée, une référence lexicale adéquate à consulter est indispensable pour répondre à ce type de questions. En deuxième lieu, un modèle psycholinguistique ou linguistico-informatique qui peut expliciter (1) l'organisation du lexique mental ou du sens des mots (2) le processus concerné est nécessaire. En troisième lieu, une méthode ou un modèle opérationnel est nécessaire pour répondre à ces questions. Dernièrement, un modèle capable d'apprendre automatiquement ce type de relations entre les mots est souhaitable.

Dans ce chapitre, nous examinerons pas à pas ces quatre nécessités en explorant les modèles et les études de la littérature.

4.1 Référence lexicale

Les questions posées ci-dessus à propos de la relation entre *stagger* et *drunk* ou entre *rédigier* et *article* peuvent être résolues facilement s'il existe des références lexicales convenables à consulter. En revanche, l'absence de telles références posera un problème sévère, non seulement à la machine mais aussi aux apprenants d'une langue étrangère. En effet, pour les non-francophones ou pour les machines, trouver le verbe typique qui s'emploie pour décrire « la démarche d'un ivrogne », par exemple, n'est pas un problème tout à fait simple. Les références disponibles comme les dictionnaires usuels ne sont pas très adaptées à cet objectif : un dictionnaire traditionnel est conçu pour trouver les définitions d'un mot à partir de l'entrée et non pour trouver un mot en partant des définitions.

Les dictionnaires utilisables par une machine⁴ permettent de rechercher les

⁴L'existence historique de la machine mécanique utilisant les bandes justifie l'emploi de ce terme au lieu de « dictionnaires électroniques ». Cependant, nous appliquerons le terme dictionnaire électronique avec un sens élargi pour s'opposer à un dictionnaire non-utilisable par une

entrées à partir des définitions. Les dictionnaires utilisables par la machine apparaissent dans les années 60 (Olney 1968, Ide et Véronis 1998) et deviennent les sources de connaissance pour le TAL dans les années 80 (Amsler 1980, Michiels, Mullenders et Noël 1980). Dès lors, les études utilisant ces dictionnaires ont pour but d'en extraire automatiquement les connaissances lexicales et sémantiques. Certains auteurs ont construit un ensemble de connaissances associées aux mots, en utilisant les mots inclus dans les définitions (Wilks, Fass, Guo, McDonald, Plate et Slator 1988, Inkpen et Hirst 2003). Cependant, malgré ce type d'améliorations qui restent limitées, l'ambition initiale n'a pas atteint son objectif. Cet échec pouvait en fait être prévu, car si les dictionnaires utilisables par machine sont une innovation dans l'utilisation des ressources, ils ne le sont pas dans leurs contenus. En effet leurs contenus sont identiques aux dictionnaires classiques qui eux-mêmes rencontrent des problèmes d'incohérence et des problèmes dus à l'absence de connaissances qui dépasseraient le champ de la sémantique lexicale canonique comme le sont les connaissances du monde, les connaissances pragmatiques, etc. Ce type de limite d'un dictionnaire s'observe par exemple dans la question mentionnée ci-dessus : le dictionnaire *Le Robert* propose les verbes qui décrivent une démarche instable comme *chanceler*, *flageoler*, *tituber*, *trébucher*, *vaciller*, etc., mais il n'y a pas de définition ou d'information à propos de « la démarche d'un ivrogne ». D'ailleurs, dans ce dictionnaire les exemples qui pourraient fournir des connaissances indirectes ne sont pas très cohérents : *Le Robert* propose deux phrases « Il chancelle comme un homme ivre » et « Un ivrogne, un malade qui titube » mettant ainsi en évidence que *chanceler* et *tituber* sont les candidats pour le verbe cherché ; cependant les francophones répondront majoritairement *tituber* et non *chanceler*.

Une autre possibilité face à cette limite des dictionnaires classiques et à l'échec

de l'utilisation des dictionnaires électroniques pour étendre la connaissance consiste à développer une ressource indépendante des dictionnaires classiques. Le thésaurus et WordNet (Fellbaum 1998) en sont des exemples. Le thésaurus fournit les informations plus riches et mieux organisées qu'un dictionnaire classique. Par exemple, le *Thésaurus de Larousse* présente les mots dans des catégories classées par thèmes distinctifs comme CONSTRUCTION, ARBRE, PEUR, etc ; dans chaque catégorie, les différents sens/concepts sont organisés par exemple PANIQUE, ANGOISSE, PHOBIE, TIMIDITÉ, etc., pour *peur*.

WordNet est construit pour la consultation mais, en se dégageant de la structure des ressources existantes telles que les dictionnaires, les thésaurus, etc., il a pour but explicite d'organiser le sens des mots. Ce modèle propose des « synsets » qui fournissent l'ensemble des synonymes interchangeable dans un contexte donné. En ce sens, WordNet classe les sens d'un mot de manière traditionnelle. Mais le « synset » est différent d'une simple énumération traditionnelle des synonymes, car d'une part un « synset » est défini de manière globale et non locale : un synset S n'est pas spécifique d'un mot M ; et d'autre part il représente une unité de sens définie de façon plus rigoureuse. Outre cette différence, WordNet suggère une organisation structurale hiérarchique des mots de manière assez complète ainsi que des relations d'antonymie et de méronymie. Par exemple, le mot *sparrow* a les hyperonymes suivants : *passerine*, *passeriform bird*, *bird*, *vertebrate (craniate)*, *chordate*, *animal (animate being, beast, brute, creature, fauna)*, *organism (being)*, *living thing (animate thing)*, *object (physical object)*, *entity*. Ainsi deux mots qui sont dans la même catégorie peuvent être considérés comme proches sémantiquement alors que ce type de mesure n'est pas possible avec un dictionnaire ordinaire.

Bien que WordNet propose ce type de connaissance utile, une connaissance plus étendue comme celle qui lie *stagger et drunk* n'est pas incluse dans ce modèle

pas plus qu'il ne l'était dans les thésaurus.

Le fait que les informations ne soient pas satisfaisantes, notamment pour notre exemple ci-dessus, pour ce qui est des dictionnaires classiques, des dictionnaires utilisables par la machine, des thésaurus et de WordNet est dû au problème fondamental de ces ressources : les informations qu'elles contiennent sont limitées en quantité. Ceci est inévitable par nature, car le recueil des informations est effectué de manière non automatique et cette façon manuelle ne peut pas traiter les informations de grande taille qui ne cessent de s'ajouter chaque jour.

Ainsi, l'extraction automatique de la connaissance ne peut s'améliorer que par l'utilisation de ressources plus étendues comme le sont les corpus. Cette tentative a été initiée par Weiss (1973) ; Kelly et Stone (1975) ont obtenu « des mots clefs du contexte » pour 1800 mots ambigus en utilisant un corpus de 500 000 mots. Les différentes méthodes qui utilisent des corpus seront discutées plus loin en détail.

Le modèle *Script* de Schank (1977) propose onze prédicats primitifs afin de représenter une expression linguistique. Par exemple, il distingue le transfert abstrait, physique et mental qui s'appliquent aux verbes pour les regrouper. Ainsi ce modèle organise les mots ou les concepts de manière très succincte. Mais l'efficacité des règles d'inférence proposées, indispensable pour une tâche de TAL, reste discutable.

4.2 Les modèles d'organisation du sens des mots

Les questions posées au début de ce chapitre demandent deux précisions en ce qui concerne la modélisation linguistique ou psycholinguistique. D'une part, le fait que les locuteurs d'une langue puissent répondre correctement aux questions posées met en évidence le fait qu'il existe soit des liens entre les mots/concepts, par exemple STAGGER et DRUNK dans le lexique mental, soit un processus qui

assure ce lien. D'autre part, un modèle linguistique devrait ou bien avoir le moyen d'intégrer ce type d'information ou bien avoir une méthode pour simuler cette capacité linguistique humaine.

Etant donné qu'il n'existe pas suffisamment de modèles qui rendent compte directement de ce type de relation en-dehors de la sémantique au sens étroit, nous examinerons les modèles qui tentent d'explicitier de manière générale l'organisation du sens des mots, du lexique mental ou des processus langagiers.

Les modèles d'activation propagée supposent un réseau sémantique lié par des nœuds qui eux-mêmes représentent des concepts. L'activation reçue par certains nœuds est soit renforcée soit inhibée selon les architectures des modèles (McClelland et Elman 1986, Collins et Loftus 1975). Par exemple, si le terme A est relié à plusieurs traits T_1 , T_3 et T_5 et le terme B à T_1 , T_4 et T_6 , l'activation de traits T_1 et T_3 pourra activer le terme A plutôt que le terme B. Ce type de modèles a été mis en œuvre avec des réseaux connexionnistes et s'oppose ainsi aux modèles de type symbolique. Les avantages importants de ces modèles sont d'une part la résistance aux éventuels bruits (par exemple : l'ensemble des traits T_1 , T_3 , T_4 et T_5 donnera la réponse la plus probable A malgré l'information incohérente T_4) et d'autre part la capacité à rétablir des informations après une destruction partielle des informations. Une des limites majeures de ce type de modèles est que le nombre de termes utilisés dans le modèle proposé est très petit et il n'est pas évident, pas plus en théorie qu'en pratique, de savoir si le modèle est capable d'extraire les caractéristiques d'un mot parmi les relations complexes établies sur un nombre de mots assez important (pouvant par exemple atteindre quelques centaines de millions).

Quant aux modèles symboliques, l'intégration des relations complexes n'est pas un souci théorique. Car, contrairement aux modèles connexionnistes, les modèles symboliques représentent explicitement les sens ou leurs liens, soit par des

informations sémantiques ou syntaxiques, soit par des règles (Richens 1958, Quillian 1968, Wilks 1968, Small 1987, Hirst 1987, Habert, Naulleau et Nazarenko 1996, Grefenstette et Teufel 1995). A titre d'exemple, le modèle d'Edmonds et Hirst qui rend compte des quasi-synonymes explique la différence subtile entre les quasi-synonymes par leurs caractéristiques distinctives : le CARACTÈRE PÉJORATIF pour *blunder*, l'ERREUR LIÉE À LA MÉMOIRE pour *lapse*, par exemple. Ainsi, intégrer la relation « *stagger – drunk* » ou « *rédiger – article* » dans leur modèle est possible, au moins en théorie. En revanche, l'application effective de ce modèle demande un travail manuel important, ce qui met en doute son emploi pratique.

4.2.1 Le modèle de Ploux

Le modèle de Ploux pour les synonymes représente les liens sémantiques des synonymes dans un espace multidimensionnel (Ploux 1997, Ploux et Victorri 1998). Contrairement aux modèles non continus, ce modèle est capable de représenter les sens d'un mot de manière continue et graduelle. Par exemple, dans une projection des sens du mot *insensible*, les changements progressifs des sens de PHYSIQUE vers ÉMOTIONNEL sont respectés et représentés de manière claire. Ce modèle utilise la notion de *clique* qui est choisie comme unité minimale du sens du mot. L'avantage de ce modèle est qu'il est continu et que les liens complexes entre les mots sont toujours représentables. D'ailleurs le modèle est capable d'organiser les liens à partir de données simples. Par exemple, à partir d'une simple liste de synonymes non organisée telle que { *insensible* : anesthésié, calme, cruel, de marbre, dur, détaché, endormi, endurci, engourdi, froid, glacial, immobile, imperceptible, impitoyable, impénétrable, inanimé, indifférent, inerte, inhumain, insaisissable, mort, neutre, paralysé, sans cœur, sévère, timide, égoïste } et { *égoïste* : avare, bourgeois, exclusif, indifférent, individualiste, insensible, muflé, personnel, plein

de soi, sans-cœur, sans-gêne, égocentrique }, etc.⁵, ce modèle organise et propose une représentation bien classée des sens comme { sévère, dur, sans cœur, inhumain, cruel }, { insaisissable, neutre, imperceptible, timide } et { calme, immobile, endormi, paralysé, inerte }, etc.

Ainsi, l'intérêt de ce modèle consiste en le fait qu'il est capable d'organiser les relations des sens complexes sans système de règles ou consignes préalables. D'ailleurs, étant donné que la plateforme de ce modèle est mathématique, informatique et indépendante d'une langue spécifique, un modèle évolué basé sur ce modèle peut être développé sans changer radicalement l'architecture principale. En effet, le modèle qui représente simultanément deux langues différentes a été développé rendant possible la comparaison entre la différence et la similarité de deux langues (Ploux et Ji 2003). Le détail du principe de ce modèle sera décrit plus loin, lors de la présentation du modèle actuel. A propos des quatre problèmes posés précédemment, comme ce modèle représente les relations synonymiques et non des associations des mots, nous ne le décrirons pas de façon plus étendue.

4.3 Les modèles utilisables dans une tâche de TAL

Si les modèles que nous venons d'examiner concernent directement l'organisation des sens d'un mot ou d'un processus / d'une organisation du lexique mental, il existe d'autres types de modèles et de méthodes linguistico-informatiques qui n'abordent pas directement ce genre de sujet mais qui sont plutôt destinées à s'appliquer dans une tâche réelle de TAL comme la désambiguïsation du sens des mots.

D'abord, il existe des méthodes qui utilisent comme ressource principale les dictionnaires utilisables par la machine. Ces méthodes consistent à utiliser comme

⁵Ces exemples montrent une partie des synonymes et non leur totalité.

mots clefs les mots apparaissant dans la définition d'un sens du mot, soit directement (Lesk 1986), soit par un filtrage approprié (Fass et Wilks 1983, Wilks et Fass 1990) afin de désambiguïser un mot ambigu. L'amélioration de ce type de méthodes consiste à intégrer d'autres informations que la définition sémantique, telle que l'information sur le thème : MÉDICINE-ET-BIOLOGIE ÉCONOMIE, etc. et les sous-thèmes comme ANATOMIE et BIOCHIMIE (Guthrie, Guthrie, Wilks et Aidinejad 1991). Guthrie classe les mots liés à un mot ambigu grâce aux thèmes prédéfinis par le dictionnaire électronique. Pour *bank* les listes *account, cheque, money, keep, draw, safe* et *river, wall, flood, overflow* se rapporteront au couple ÉCONOMIQUE / CODE NULL⁶.

Les méthodes qui utilisent les dictionnaires utilisables par la machine ont une limitation fondamentale et générale concernant la quantité d'informations, car leurs ressources sont dépendantes du contenu des dictionnaires utilisés, dictionnaires dont la quantité d'information est elle-même limitée. Afin de surmonter cette limite, il est indispensable d'utiliser des ressources plus grandes que les dictionnaires. De là vient la nécessité d'utiliser des corpus.

L'utilisation de corpus en tant que ressource est par nature une approche ascendante. Weaver avait déjà insisté sur la nécessité du recours aux méthodes statistiques en 1949 (Weaver 1949/1955). Cette approche ascendante — autrement dit empirique / quantitative / statistique — dans les études linguistiques a attiré l'intérêt dans les années 50 mais elle a été longtemps négligée après certaines critiques — notamment de Chomsky qui a exposé les limites de cette approche non seulement d'un point de vue pratique mais aussi d'un point de vue théorique (Chomsky 1956, Chomsky 1957, Miller et Chomsky 1963). Cependant, grâce à la disponibilité de corpus de très grande taille et à la facilité de leur utilisation, cette approche

⁶Contrairement au sens BANQUE, le sens RIVE n'a pas été classifié et se rapporte au domaine NULL dans la référence utilisée (LDOCE) pour cette étude.

reconquiert l'attention de la linguistique notamment de la linguistique computationnelle et de la psycholinguistique ainsi que de la lexicographie (Sinclair 1985).

Au sein de la linguistique computationnelle on peut distinguer deux types de recherche parmi les applications qui utilisent les corpus comme ressource importante. L'un est utilisé pour l'amélioration de la prédiction d'un mot, notamment pour la reconnaissance de la parole, et l'autre sert à la désambiguïsation du sens des mots. Le principe général du premier consiste à choisir et à trier les mots les plus probables susceptibles de suivre un mot à l'aide de l'entraînement qui prend en compte n mots précédents (sur ce sujet voir Shannon 1948, Baker 1975, Jelinek, Mercer, Bahl et Baker 1977, Furui 1986, Baum 1972, Church et Gale 1991, Bahl, Jelinek et Mercer 1983). La désambiguïsation du sens des mots est essentielle dans plusieurs tâches du TAL comme la traduction automatique, la génération de texte, la recherche documentaire, le traitement de la parole, la correction automatique de l'orthographe, etc.

Un exemple pour l'utilisation de l'approche statistique pour la désambiguïsation du sens des mots se trouve dans le choix entre *make the decision* / *take the decision* lors de la traduction de la locution française *prendre la décision* : la co-occurrence forte entre *make* et *decision* suggère que *make the decision* est le choix correct. Afin d'améliorer la performance d'un système qui ne rend pas compte du changement grammatical de la collocation comme « Je vais prendre ma propre décision » ou « Je me demande par quelle mouche Paul a été piqué »,⁷ on pourrait utiliser une analyse syntaxique (Wehrli 1998). Cependant, une méthode statistique peut s'employer également pour résoudre ce type de problème en considérant le contexte pertinent au lieu de se concentrer uniquement sur le phénomène local à savoir une fenêtre très réduite.

Ainsi, l'approche qui utilise des corpus afin d'en extraire des relations entre

⁷Elles sont les exemples dans Brown, Pietra, Pietra et Mercer 1991 et Wehrli 1998.

mots est basée sur l'hypothèse que les différents sens d'un mot ne sont guère utilisés simultanément dans le même contexte, discours, document ou domaine et que ces sens correspondent aux différents types de documents (Weaver 1949/1955, Gale, Church et Yarowsky 1992, Resnik 1997). Par exemple, les deux sens du mot anglais *suit* (procès, costume) se distinguent selon qu'il s'agit de documents judiciaires ou de documents se rapportant à des vêtements ; *plant* (plante, usine) pour les documents traitant de biologie contre les documents se rapportant à l'industrie. Dès lors que deux types de documents sont suffisamment distincts, certains mots clefs dans ces documents peuvent être utilisés comme indices pour une tâche de désambiguïsation du sens des mots. Par exemple, dans une étude de Yarowsky, les mots *life* et *manufacturing* sont utilisés en tant qu'indice initial pouvant distinguer les deux sens du mot *plant* qui permettront par la suite de faire apprendre d'autres mots clefs comme *growth*, *height*, *flower*, *fruit*, *species*, *car*, *union*, *equipment*, *assembly*, *nuclear*, *job* et qui distingueront ces sens.

Il est vrai que cette méthode produit des mots clefs importants pour les mots ambigus mais les méthodes concernées ne fournissent pas les informations ou les connaissances demandées afin de résoudre les problèmes présentés au début du chapitre. Ceci est dû au fait que ce type d'approche a pour but de désambiguïser les sens des mots et qu'il n'y a pas de moyen de traiter tous les mots de manière automatique. Afin d'avoir des connaissances complètes sur les mots, il faudra nécessairement rendre compte des relations entre les mots clefs et les autres mots en plus de celles qui lient le mot en question et les mots clefs. Par exemple, si un mot cible M_1 possède les mots clefs M_2, M_3, \dots, M_n , il est nécessaire d'organiser de manière structurale les relations entre $\{ M_2, \dots, M_n \}$.

D'ailleurs, ce type d'approche comporte un autre problème qui l'empêche d'être utilisé dans une tâche réelle de TAL : elle nécessite une intervention humaine (car l'annotation des différents sens des mots ainsi qu'une décision sur le

sens sont préalablement indispensables). Afin de surmonter ce problème et d'automatiser le processus, des corpus parallèles ont été utilisés comme ressource (Yarowsky 1995). Cette méthode permet une meilleure automatisation car le système peut distinguer les différents sens d'un mot dans une langue en les comparant avec les différents mots éventuels dans une autre langue. Par exemple, les documents contenant différents sens du mot ambigu *plant* peuvent être classés de manière automatique en comparant les documents français équivalents dans lesquels il existe deux mots *plante* et *usine*. Bien que cette approche propose une solution possible, il reste cependant quelques obstacles. D'une part, les deux types de documents parallèles peuvent comporter des mots non-désambiguïsés, car certains sens du mot peuvent être ambigus dans les deux langues comme *hot* et *chaud*. D'autre part, le fait qu'il n'existe pas suffisamment de corpus parallèles pose un problème non négligeable.

Une autre méthode qui utilise des dictionnaires électroniques a été proposée pour un traitement plus automatique (Karov et Edelman 1996). Karov et Edelman ont d'abord choisi les mots pertinents pour chaque sens d'un mot en utilisant un dictionnaire électronique. Les mots appartenant à la fois aux deux sens dans les exemples du dictionnaire par exemple ont été rejetés. L'avantage de cette méthode par rapport à la méthode de Yarowsky vient du fait qu'elle dépend moins d'un travail manuel, car cette première méthode utilise le dictionnaire électronique pour obtenir les mots clefs initiaux. Cependant, cette approche dépend d'une manière ou d'une autre d'un travail « à la main ». Le dictionnaire lui-même est construit par l'homme et comme nous venons de le voir plus haut, les dictionnaires présentent d'autres problèmes à l'égard de la représentation sémantique du mot.

4.4 Les modèles capables d'apprendre

A proprement parler, les méthodes ou les modèles présentés jusqu'ici n'apprennent pas automatiquement les relations sémantiques ou pragmatiques des mots. Bien que certaines méthodes comme celle de Yarowsky puissent prétendre être des méthodes d'apprentissage automatique puisque sa méthode permet d'apprendre de nouveaux mots clefs en partant des mots clefs initiaux restreints, il ne s'agit pas à proprement parler d'apprentissage. Lorsque nous parlons d'apprentissage automatique nous entendons par cela l'apprentissage par la machine sans aucune intervention ou aide humaine.⁸ Afin de cerner la différence entre ce critère et le critère conventionnel, prenons un cas un peu extrême. Imaginons par exemple que des extra-terrestres arrivent sur une terre ruinée et qu'ils veuillent établir des relations entre les sens des mots à partir de documents textuels uniquement. Dans cette situation, une méthode sémi-automatique telle que celle de Yarowsky ne marchera pas car une étape initiale, à savoir le choix d'un mot d'indice initial, ne peut pas être effectuée automatiquement. En ce sens, la méthode de Yarowsky n'est pas une méthode d'apprentissage automatique, car l'intervention humaine est incontournable.

Schütze propose une approche entièrement automatique pour distinguer les sens d'un mot. Son approche relève de la « discrimination des sens » et non de la « désambiguïsation des sens », car son modèle propose différentes régions dans un espace sémantique au lieu de proposer un sens explicite du mot. Son modèle consiste à créer un espace multidimensionnel, dont les dimensions sont aussi nombreuses que les mots considérés. Ainsi les différents sens peuvent être représentés différemment dans un espace des sens multidimensionnel. Dans ce modèle les

⁸Nous admettons que ce critère est bien plus rigoureux que le critère conventionnel qui emploie le terme « apprentissage » de manière moins rigoureuse. Il sera utile si l'on cherche à développer un système ou un agent intelligent rivalisant avec les êtres humains.

contextes d'un mot ambigu sont représentés par des vecteurs de contexte dans « l'Espace du Mot » qui sont les sommes des vecteurs des mots dans le contexte. La classification de ces vecteurs par la méthode de regroupement est par la suite effectuée avant de comparer leur proximité avec le mot en question dans l'Espace du Mot. Par cette méthode, les mots *judge* et *robe* sont positionnés en fonction de leurs proximités par rapport aux sens LÉGAL et VÊTEMENT respectivement. De la même manière, les mots ayant des sens différents avec des variations graduelles sont représentés par un ordre de mots qui va du sens « légal » vers le sens « vêtement » et *law, judge, statute, suit* du sens LÉGAL vers VÊTEMENT (Schütze 1998).

Une approche similaire à ce modèle a été développée dans le domaine de la psycholinguistique. Ce modèle, nommé l'Indexation de la sémantique latente (LSI) en linguistique computationnelle et l'Analyse de la sémantique latente (LSA) en psycholinguistique est un modèle qui a pour but d'extraire et de représenter le sens des mots dans l'usage contextuel (Landauer, Laham, Rehder et Schreiner 1997, Landauer, Foltz et Laham 1998). Les auteurs défendent la capacité de simulation de LSA pour l'acquisition du vocabulaire, la reconnaissance et la catégorisation des mots, l'amorçage sémantique de la phrase-mot, la compréhension du discours et les jugements sur la qualité de la composition.

Le principe de ce modèle consiste à construire une matrice réduite qui reflète l'appartenance d'un mot à une portion du texte (un paragraphe ou le document entier). La méthode utilisée dans la décomposition de la matrice est l'analyse factorielle (ou la décomposition en valeurs singulières). Ainsi ce modèle utilise à peu près 300 dimensions ce qui représente une réduction par rapport au nombre total de mots. Etant donné que ce modèle a pour but de représenter les sens d'un mot lui-même, désambiguïser ou discriminer les sens du mot n'est pas central dans cette approche. En appliquant ce modèle à une étude sur la prédiction,

Kintsch (2001) propose de rendre compte du changement de sens du mot en fonction du changement du contexte. Par exemple, les différents sens du mot *collapsed* dans « The bridge collapsed » « The plans collapsed » et « The runner collapsed » sont expliqués par les différentes valeurs des cosinus des vecteurs associés aux termes *collapsed* et *bridge*, *plans*, *runner*. Ce modèle ne distingue cependant pas les différents sens d'un mot en tant que tels, contrairement au modèle de Schütze.

Bien que le modèle de Schütze et le modèle LSI soient plus adaptés que les autres modèles pour résoudre les problèmes posés au début de ce chapitre, cependant, quelques problèmes subsistent. Bien que LSI puisse produire des mots sémantiquement proches d'un mot donné, il ne propose pas une organisation comme, par exemple, celle qui est construite dans le modèle de Ploux. D'ailleurs, LSI permet de choisir un mot très proche d'un mot cible même si tous les deux n'apparaissent pas en même temps dans les textes.

Bien que le modèle de Schütze soit capable de discriminer les sens d'un mot, son approche est indirecte. C'est-à-dire, au lieu de produire des mots liés à un mot cible, ce modèle vise à comparer deux possibilités afin de décider qu'elle est la meilleure solution. Ce modèle ne propose pas plus de deux sens pour mot. Mais comme nous allons le voir pour le mot *match*, nous avons besoin de discerner plus de deux sens pour un mot.

4.5 Proposition

Comme nous venons de le voir, il existe des modèles qui abordent les relations d'association entre mots et qui satisfont partiellement les quatre nécessités que nous avons dégagées, à savoir la constitution de ressources, la modélisation de l'organisation du sens des mots, le TAL et l'apprentissage automatique ; il n'existe cependant pas de modèle qui satisfasse à la fois ces quatre demandes. Nous pen-

sons qu'un modèle qui peut répondre à ces quatre nécessités serait souhaitable d'un point de vue à la fois théorique et pratique.

S'il existait par exemple une référence lexicale dynamique qui reflèterait les relations d'association entre mots (ou la relation contexonymique), elle pourrait servir comme base de données et être utilisée d'une part dans des expériences psycholinguistiques ou bien pour tester un modèle informatique qui fait de la simulation et d'autre part comme baromètre fiable dont on peut comparer les résultats à ceux d'expériences psycholinguistiques ou ceux de simulations informatiques. Pour ce faire, il est nécessaire que la référence lexicale soit capable de refléter la complexité des relations contexonymiques au-delà d'une simple énumération des contexonymes. D'ailleurs, une référence fiable sera très utile pour certaines tâches de TAL comme la désambiguïsation du sens des mots. Par exemple, si une référence intelligente est capable de classer les contexonymes vis-à-vis de leurs différents sens, ces contexonymes classés dans différents groupes pourront être utilisés comme indices initiaux dans une méthode de désambiguïsation des sens comme celle de Yarowsky rendant cette méthode plus automatique. L'appariement entre deux langues ou plus sera alors possible pour effectuer des tâches de désambiguïsation sémantique.

Par ailleurs, si une telle référence lexicale est créée de manière automatique, ceci est directement lié à un modèle d'apprentissage automatique. Réciproquement, l'existence d'un modèle capable d'apprendre et d'organiser automatiquement les contexonymes rend possible la création d'une telle référence lexicale. Il existe ainsi un lien constructif entre modèle d'apprentissage automatique et modèle applicable au TAL. Il reste néanmoins à discuter le lien entre un modèle psycholinguistique et un modèle de TAL ou d'apprentissage automatique. Il faudra pour cela répondre aux questions suivantes : est-il possible que ces deux types de modèles aient des principes et des mécanismes tout à fait indépendants ? Quel

est l'intérêt de partager la même architecture entre ces deux types de modèles ?

Avant de répondre à ces questions, examinons brièvement la nature de la simulation pour une modélisation, au regard des capacités humaines ou plus généralement de celles des êtres vivants. Dans le domaine de la bioinformatique, il existe des modèles informatiques qui simulent les fonctionnements des neurones au niveau moléculaire. De plus il existe d'autres types de modèles qui simulent des fonctionnements de plus haut niveau comme les fonctionnements globaux des neurones ou bien les fonctions associées à une région spécifique du cerveau. Dans ce deuxième cas il n'existe plus au sein de ces modèles, de notion / d'élément équivalent aux molécules tels que les neurotransmetteurs, les protéines gamma, etc. Ainsi, certains modèles peuvent atteindre leurs objectifs sans simuler parfaitement les êtres vivants. Par ailleurs, dans certains cas, cette stratégie est incontournable. L'avion en est un exemple. Tandis que le robot volant devrait simuler dans les moindres détails des principes et des mécanismes du vol des insectes ou des oiseaux, l'avion n'a d'autre intérêt que de voler et une simulation parfaite à la manière d'un robot volant n'est ni nécessaire ni conseillée (par exemple, l'avion ne bat pas des ailes).

Dans le domaine de la synthèse de la parole, il existe deux voies différentes de cette nature : tandis que le modèle *clone parlant et la synthèse audiovisuelle* de l'Institut de la Communication Parlée (ICP) tente de simuler non seulement la sonorité humaine mais aussi le mouvement de la bouche, des cordes vocales, etc., la plupart des modèles de synthèse de la parole essaient de simuler uniquement la partie acoustique. Pour ce qui est des modèles de simulation linguistique, il existe d'une part des modèles qui simulent le fonctionnement des régions cérébrales concernées, et d'autre part des modèles qui simulent globalement le processus langagier ou bien la représentation du lexique mental. Si le premier type de modèle a un intérêt croisé avec les neurosciences ou la médecine, le dernier partage plutôt

un intérêt avec la linguistique et la psycholinguistique pour ce qui concerne les relations sémantiques et celles de contexonymie des mots. Par « modèle pour la simulation linguistique » nous désignerons dans cette étude ce dernier type de modèle.

A première vue, les modèles pour le TAL et les modèles psycholinguistiques semblent être indépendants les uns des autres puisque leurs objectifs sont radicalement différents. Cependant, il existe des liens non négligeables entre eux. L'effet d'amorçage, par exemple, examiné dans les expériences psycholinguistiques pour les relations d'association telles que *médecin* — *infirmière doctor* — *nurse* sera utile dans une tâche de TAL de désambiguïsation du mot *doctor* en présence du mot *nurse*. En d'autres termes, si un modèle peut apprendre le lien entre *doctor* et *nurse* et ainsi réduire la distance sémantique entre ces deux termes/concepts ou bien privilégier le sens médical de *doctor* en présence de *nurse*, ceci pourra avoir un effet direct d'une part sur l'amélioration de la performance d'une tâche de TAL et d'autre part pour la simulation d'expériences psycholinguistiques.

En effet, comme le dit Bar-Hillel (1953), la traduction complète nécessite non seulement une sémantique des mots mais aussi une connaissance du monde. Comme nous l'avons proposé dans le chapitre précédent, les contexonymes reflètent ce type de connaissance — au moins partiellement — et si un modèle est capable de produire des contexonymes pour un mot donné, nous pensons qu'il peut servir pour améliorer des tâches de TAL. Par exemple, les contexonymes tels que *stupid* pour *blunder* assurent de manière indirecte la caractéristique PÉJORATIVE du mot *blunder*; s'il existe dans la langue cible des contexonymes équivalents à *stupid* parmi les synonymes du mot cible cherché pour traduire *blunder*, cela permettra de choisir de façon fiable le mot cible.

Il est vrai que l'association du sens des mots/concepts n'est pas toujours due uniquement au renforcement des associations fréquentes entre mots. Par exem-

ple, dans le langage parlé, les caractéristiques prosodiques comme l'intonation, l'accentuation, le rythme, l'intensité, etc. et les traits paralinguistiques comme les expressions faciales, les gestes, etc. jouent un rôle important, ainsi que les traits extra-linguistiques, pour le renforcement des liens sémantiques/pragmatiques.

De ce fait, les statistiques sur la co-occurrence des mots ne peuvent pas expliquer parfaitement tous les liens sémantiques et pragmatiques. Pour cette raison, il existe des modèles qui tentent d'intégrer des informations sensorielles liées à la vision et à l'audition en plus des informations textuelles. Nous reconnaissons que les informations textuelles ne sont pas des ressources complètes. Par exemple, tandis que le lien entre *cheval* et *monter* est observé dans les textes aussi fréquemment que dans les ressources visuelles comme les dessins, les peintures, les photographies, les films, etc., le lien entre *éléphant* et *grand* dans les textes sera moins fréquent que dans les ressources visuelles, car le mot *éléphant* seul suffit à transmettre le sens GRAND et le mot *grand* sera accompagné d'*éléphant* moins systématiquement que *monter* pour *cheval*. Nous pensons néanmoins qu'une vaste quantité d'informations textuelles compensera partiellement cette limite. C'est-à-dire, si le lien entre deux mots/sens A et B est pertinent, ce lien sera présenté dans les textes plus fréquemment que le lien non pertinent, ce qui donne la possibilité de différencier ces deux cas. Par exemple, des phrases telle que « Cet éléphant est vraiment grand », « L'éléphant est plus grand que j'avais imaginé », « L'éléphant est un grand animal » seront plus fréquentes que des phrases telle que « Cet éléphant est plus petit que le soleil », « L'éléphant est plus moche que j'ai imaginé », « L'éléphant ne peut pas voler ». En effet, la quantité de corpus utilisés actuellement dans les recherches linguistiques computationnelles dépasse très largement la quantité d'information moyenne qu'un individu traite pendant toute sa vie — une taille moyenne équivalente à 200 méga octets (Iwańska 2000), ce qui rendrait possible de tirer les liens sémantico-contextuels pertinents même si ce type d'in-

formation est présentée moins systématiquement que l'observation sensorielle.

De ce fait, nous proposons un modèle qui produit explicitement les contextonymes ; ce modèle propose aussi une organisation convenable qui peut être utilisée soit dans une tâche de TAL, soit dans la simulation du lexique mental ainsi que comme référence lexicale.

Chapitre 5

Modèle automatique d'organisation des contexonymes (ACOM)

Le modèle automatique d'organisation des contexonymes ACOM¹ (Ji et Ploux 2003) qui a été développé dans cette étude utilise la même notion de *clique* et la représentation géométrique que Ploux et al. ont développé (Ploux 1997, Ploux et Victorri 1998, Ploux et Ji 2003). Nous désignerons ce dernier modèle par son nom actuel Atlas Sémantique.

Contrairement à la plupart des modèles sémantiques, l'Atlas Sémantique organise automatiquement du sens des mots sans la consultation de la classification du sens des dictionnaires, des thésaurus ou du WordNet. Par exemple, l'Atlas Sémantique n'utilise pas de classification du sens telle que PEU DE GRAISSE, PEU ÉPAIS, PEU ABONDANT, PEU D'IMPORTANCE, etc. pour le mot *maigre* des références lexicales ; il n'utilise en effet que la liste des synonymes non organisés comme la suivante :

(10) maigre : *hâve, émacié, sec, fin, pauvre, stérile, petit, médiocre*

¹Automatic Contexonym Organizing Model.

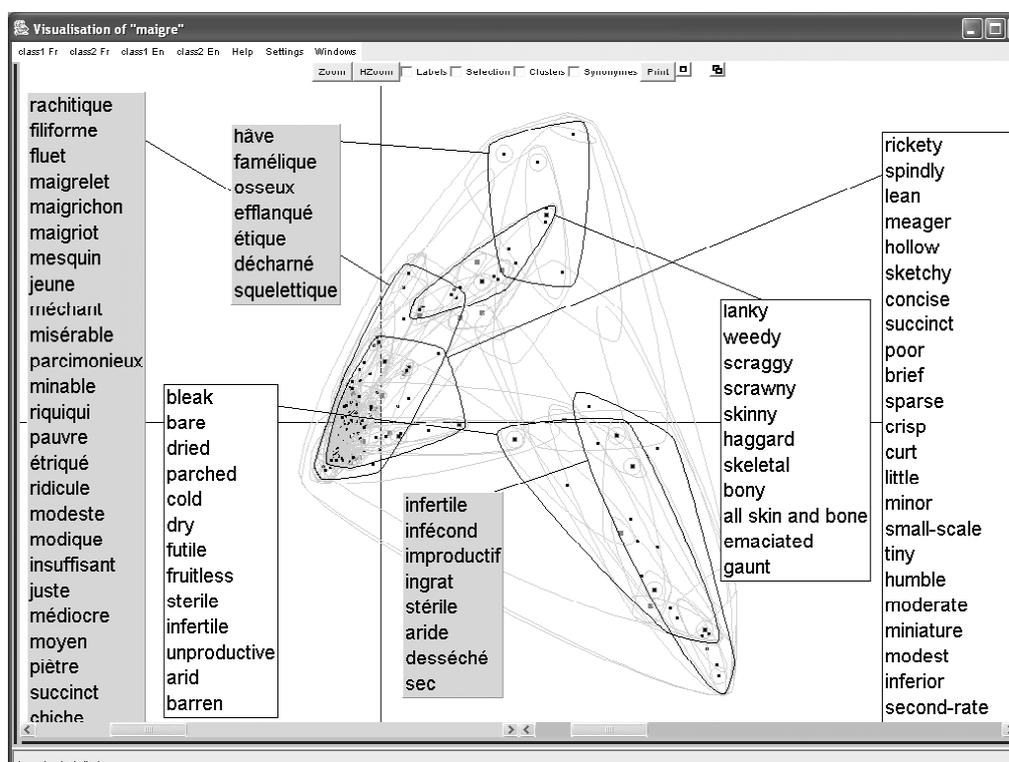


FIG. 5.1 – Rerésentation du mot *maigre* et son équivalent anglais.

Le modèle sémantique pour deux langues, développé par Ploux et Ji (2003), intègre un mécanisme matriciel qui permet de sélectionner les mots de la langue cibles pour des mots donnés de la langue source. Dans ce modèle, les différents sens (S_1) d'un mot sont comparés aux les sens des mots d'une autre langue qui correspondent à l'ensemble des mots ayant les sens S_1 . La figure 5.1 montre l'appariement des sens du mot *maigre* avec les mots correspondants en anglais. On peut constater dans cette figure que les sens STÉRILE, FIN et PAUVRE sont bien distingués.

En consultant les synonymes de chaque fils dans (10), l'Atlas Sémantique classe et organise les synonymes selon leur sens de manière automatique. Pour ce faire, Ploux (1997) a proposé la notion de *clique* définie ici comme « unité »

minimale du sens qui sert à réorganiser le sens des mots. Le calcul des cliques permet en suite de représenter les différents sens dans un espace sémantique multidimensionnel.

Malgré cette capacité de classification distingué du modèle, le modèle Atlas Sémantique dépend des ressources manuellement traitées. C'est-à-dire, bien que ce modèle ne dépende pas de la classification lexicographique des sens d'un mot, la liste des synonymes préparée par l'homme pour chaque terme est indispensable pour le modèle.

En revanche le modèle actuel (ACOM) dans cette étude ne dépend aucunement des ressources organisées par l'homme mais utilise directement les textes bruts eux-mêmes comme par exemple, les articles de journaux, des textes littéraires, etc. En outre, une série de mécanismes spécifiques qui permettent de choisir les contexonymes de manière automatique a été intégrée dans le modèle actuel.

Dans ce chapitre, la série des étapes pour avoir la représentation des contexonymes pour un mot donné sera expliquée par étape par étape.

5.1 Procédure d'organisation du contexonyme

Etape 1

Pour un corpus donné, les co-occurrences de tous les mots dans un passage défini (une phrase ou une fenêtre) sont comptées et enregistrées. Chaque entrée W_i^n ($1 \leq i \leq N$, où N est égal au nombre total de types² dans le corpus) possède des fils c_j qui sont rangés dans l'ordre descendant des co-occurrences avec W_i^n ; les fils dont les co-occurrences sont inférieures au 10 000^e de la fréquence globale de l'entrée W_i^n sont éliminés afin de réduire le bruit :

²Le nombre de mots différents.

$$W_i^n : c_1, c_2, \dots, c_n.$$

Etape 2

Pour le mot cible, le tableau d'association des mots est construit en utilisant quatre paramètres.

Etape 2-1

Afin d'éliminer les fils qui ne sont guère co-occurents avec le W_i^n , une première portion α (où $0 < \alpha \leq 1$) de mots est choisie. Ainsi W_i^n devient :

$$W_i^n : c_1, c_2, \dots, c_k,$$

où $k = n\alpha$ et n est le nombre initial des fils de W_i^n .

Etape 2-2

Le paramètre β ($0 < \beta \leq 1$) sert à enlever les mots qui sont rarement co-occurents avec le fils c_j :

$$c_j^m : g_1, g_2, \dots, g_l \quad (1 \leq j \leq k, \quad l = m\beta).$$

De cette manière, le tableau d'association du mot (Tab. 5.1) est obtenu. Avec cette méthode, les contexonymes liés au sens rarement utilisés du mot W_i^n ne sont pas choisis quand la valeur du paramètre β reste petite. Ceci ne pose pas de problème néanmoins dans la mesure où les contexonymes reflètent des associations de mots attestées. Par exemple, il est difficile de supposer que le mot *fast* seul fait penser chez les anglophones au sens *hungry* dont l'emploi est très rare.³

³Dans le corpus utilisé dans cette étude, *hungry* peut être choisi comme contexonyme pour *fast* avec une valeur du paramètre α supérieure à 0,14 et une valeur du paramètre β supérieure à 0,09.

Entrée	Choisi	Enlevé
W_i^n	c_1, c_2, \dots, c_k	c_{k+1}, \dots, c_n
c_1^m	g_1, g_2, \dots, g_l	g_{l+1}, \dots, g_m
...		
c_k^p	h_1, h_2, \dots, h_q	h_{q+1}, \dots, h_p

TAB. 5.1 – Tableau d'un candidat du contexonyme.

Contrairement aux autres modèle comme LSI, s'il n'existe pas de co-occurrence entre le W_i^n et le mot M , M ne peut même pas être un contexonyme de W_i^n . De fait, ceci est conforme à notre objectif qui est de produire les contexonymes. Car, selon la caractéristique du contexonyme introduite dans le chapitre trois, deux mots qui ne s'emploient jamais ensemble sont dépourvus de liens contextuels. A notre avis, un nombre très important de textes permet de surmonter le problème de la rareté des données.

Étape 2-3

Le paramètre γ ($0 < \gamma \leq \beta$) a le même rôle que β sauf que γ est égal ou inférieur à β ce qui rend possible le fait d'avoir différents ensembles de cliques sans changer les contexonymes globaux obtenus dans l'étape précédente. Ceci donne un autre tableau d'association des mots (Tab. 5.2), qui sera utilisé ultérieurement pour obtenir des cliques ($l' = m\gamma$ et $q' = p\gamma$). En principe, une valeur plus petite du paramètre β permet d'obtenir les candidats des contexonymes plus rigoureux. De la même manière, une valeur du paramètre γ inférieure à celle du paramètre β impose une condition plus restrictive. Puisque c'est le paramètre β et non le paramètre γ qui décide les premiers candidats des contexonymes, le paramètre γ ne contribue pas à la sélection des contexonymes. Le paramètre γ peut jouer un rôle

Entrée	Choisi	Enlevé
W_i^n	c_1, c_2, \dots, c_k	c_{k+1}, \dots, c_n
C_1^m	$g_1, g_2, \dots, g_{l'}$	\dots, g_l, \dots, g_m
...		
C_k^p	$h_1, h_2, \dots, h_{q'}$	\dots, h_q, \dots, h_p

TAB. 5.2 – Deuxième tableau du contexonyme.

cependant pour l'établissement des relations entre les contexonymes puisqu'il sert à enlever les liens faibles entre contexonymes quand il est inférieur à β . Pour cette raison, la granularité des cliques (le nombre de contexonymes contenus dans une clique) est modifiée par ajustement de ce paramètre. Par exemple, les fils entre $h_{q'+1}$ et h_{q-1} dans le tableau 5.2 perdent le lien avec C_k^p contrairement aux autres fils $h_1, h_2, \dots, h_{q'}$. En d'autres termes, ces premiers ne forment pas de *clique* avec C_k^p .⁴

Étape 2-4

Le paramètre δ est un booléen. Si l'entrée W_i^n n'est pas trouvée dans les fils c_j (g_1, \dots, g_l) dans le tableau 5.1, c_j lui-même dans W_i^n et le rang c_j (qui contient les fils de c_j) sont enlevés des deux tableaux lorsque δ est actif (dans cette étude, δ a été rendu actif). Cette étape de filtrage produit l'ensemble des contexonymes suivants (C_i^m) pour W_i^n :

$$C_i^m = \{c_i : 1 \leq i \leq k, c_i \notin D\} \quad (k = n\alpha),$$

où D est l'ensemble des c_j mots supprimés par le filtrage.

⁴La notion de *clique* est expliquée de façon plus détaillée dans l'étape trois.

Etape 3

Les *cliques* sont calculées à partir de ces deux tableaux. Une clique est un terme mathématique dans la théorie des graphes qui désigne un sous-graphe maximal et complet. Si w_1 possède w_2 et w_3 comme membres et vice versa pour w_2 et w_3 , alors w_1 , w_2 et w_3 forment une clique. Autrement dit, si par exemple w_3 n'a que w_1 comme membre, ces deux éléments ne peuvent pas constituer une clique. Si w_1 , w_2 , w_3 , et w_4 constituent une autre clique, cette dernière *absorbe* la clique w_1, w_2, w_3 , en ne donnant qu'une seule clique.

Le tableau 5.2 sert à calculer ces cliques. Composées de plusieurs ensembles des mots, les cliques sont considérées dans notre modèle comme unités minimales d'un contexonyme et elles représentent des sens plus subtils que le mot lui-même.

Etape 4

L'analyse factorielle de correspondance proposée par Benzécri (1981, 1992) a été utilisée afin de représenter les corrélations entre les cliques et les termes. Le résultat est représenté par une forme géométrique dans un espace sémantique qui a autant d'axes que le nombre total de contexonymes choisis, de telle façon que chaque axe puisse représenter le mot correspondant. La distance du χ^2 entre deux cliques, y_i et y_j , est calculée afin de représenter les cliques dans un espace multidimensionnel :

$$\chi^2(y_i, y_j) = \sum_{k=1}^n \frac{x_{..}}{x_{.k}} \left(\frac{x_{ik}}{x_{i.}} - \frac{x_{jk}}{x_{j.}} \right)^2,$$

où $x_{..} = \sum_{i=1}^n \sum_{j=1}^p x_{ji}$, $x_{i.} = \sum_{k=1}^p x_{ki}$ et $x_{.i} = \sum_{k=1}^n x_{ik}$; n est le nombre total de contexonymes et p le nombre total de cliques; x_{ji} est égal à 1 si le i^e contexonyme appartient à la j^e clique, et égal à 0 sinon. Les distances entre cliques, définies à partir des coordonnées calculées, reflètent leurs similarités.

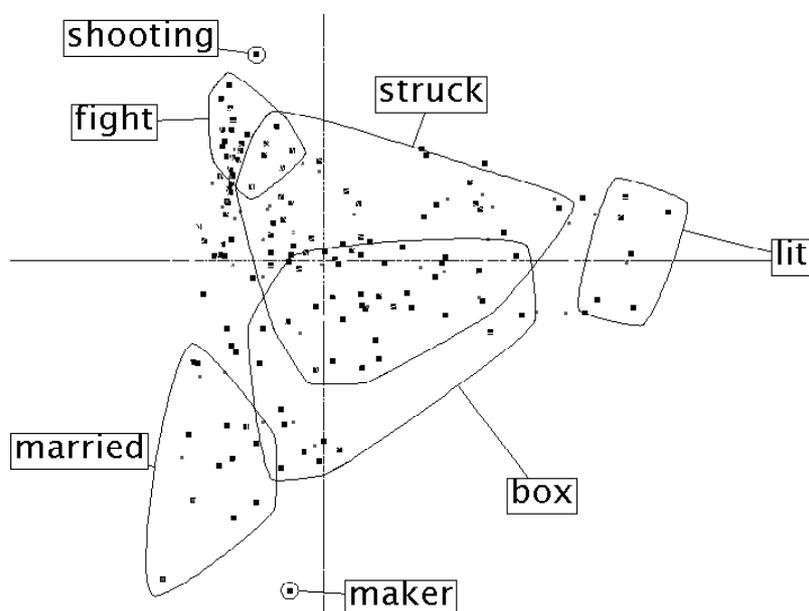


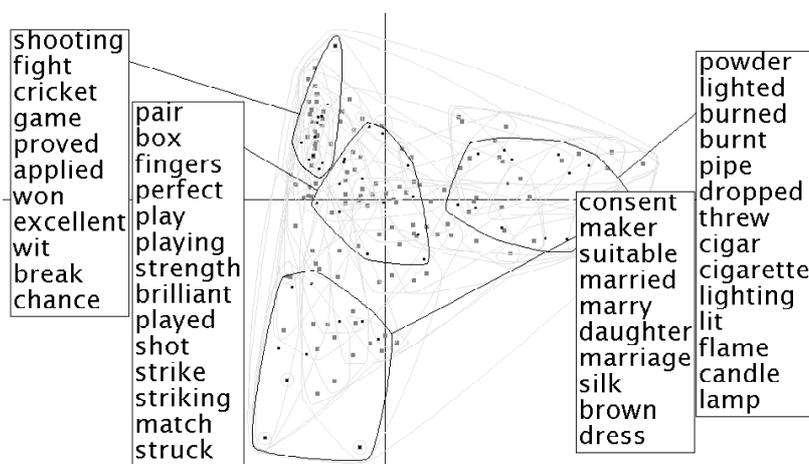
FIG. 5.2 – Certains contexonymes de *match*.

Lorsqu'un contexonyme i appartient à un nombre important de cliques alors il n'est pas discriminant et ceci est pris en compte dans l'équation car le coefficient $\frac{1}{x_{.i}}$ est faible.

On peut encore distinguer deux sous cas : (1) quand les cliques y_i et y_j partagent beaucoup de contexonymes et alors ceci est pris en compte dans l'équation et leur distance mutuelle est réduite (2) lorsque de nombreux contexonymes appartiennent à une même clique alors elle est peu discriminante pour le sens et l'équation permet de situer la clique près de l'origine des axes.

Etape 5

Les cliques sont projetées sur l'espace multidimensionnel et le plan principal est choisi pour représenter au mieux les spécificités des contexonymes. Dans cet espace, les cliques sont représentées par des points. Ces points sont les points de

FIG. 5.3 – Classification des contexonymes de *match*.

petite dimension dans la figure 5.3. Quant aux *contexonymes*, ils sont représentés par les points de plus grande taille et chacun de ces points dans cet espace multidimensionnel correspond à un contexonyme. Un point associé à un contexonyme est en effet le centre de gravité des cliques auxquelles appartient ce contexonyme. C'est-à-dire, si un contexonyme M appartient aux cliques $c_1(x_1, y_1)$, $c_2(x_2, y_2)$, \dots , $c_n(x_n, y_n)$,⁵ le point correspondant à ce contexonyme M a pour coordonnées :

$$M(x, y) = \left(\frac{1}{n} \sum_{k=1}^n x_k, \frac{1}{n} \sum_{k=1}^n y_k \right).$$

La classification des contexonymes correspond en effet à la classification de ces points. Puisqu'il existe deux types de points dans l'espace sémantico-contextuel, il existe aussi deux façons de procéder à une classification. D'abord, on peut envisager la classification par les points qui correspondent aux contexonymes et non aux cliques. Comme chaque clique correspond à un point dans la représentation géométrique, ces points peuvent être classés aussi en suivant le même principe que

⁵Puisque les points sont maintenant projetés sur un plan principal, les points sont identifiés par deux coordonnées.

celui utilisé pour les contexonymes. La différence entre les deux méthodes vient des différences représentationnelles entre *clique* et *contexonyme* et non d'une différence méthodologique.

Comme nous l'avons examiné plus haut, le point correspondant à un contexonyme est lui-même associé à un mot. De ce fait, le résultat de la classification des points de contexonymes produit directement la classification des contexonymes comme on peut le voir dans le tableau A.1. En revanche, le point correspondant à une clique ne représente pas un mot mais plutôt « sens minimal du mot ». Ainsi, le résultat de la classification des points correspondant aux cliques ne représente pas directement la classification des contexonymes. Pour avoir la classification des contexonymes, il faudra donc une étape supplémentaire. Cette étape consiste à choisir les contexonymes qui n'appartiennent qu'à un des groupes déjà classés par les points cliques. Autrement dit, les contexonymes qui appartiennent à plus d'un groupe sont enlevés car, ils ne sont pas discriminés. Ainsi, cette méthode, en excluant les contexonymes moins spécifiques ou bien trop génériques, permet d'obtenir les contexonymes qui se distinguent le mieux.

De manière générale, il existe deux types de méthodes pour classer des données. L'un regroupe les méthodes non hiérarchiques comme K-mean, Fuzzy C-means et l'autre regroupe les méthodes hiérarchiques. L'avantage de la méthode non hiérarchique vient du fait qu'elle permet d'échanger les membres d'un groupe à l'autre au cours de la classification. Le désavantage de ce type de méthode consiste en fait qu'il faut choisir le nombre de groupes à classer au préalable et qu'il est obligatoire de recalculer si le nombre de groupes à classer change après la fin du calcul. En revanche, une méthode hiérarchique permet de faire varier le nombre de groupes à classer une fois le calcul terminée. D'ailleurs, la méthode hiérarchique convient à notre besoin par son principe de la construction qui consiste à former un groupe pour les points les plus proches parmi l'ensemble des points.

Ainsi, les points très proches entre eux sont considérés comme un seul point à l'étape suivante. Cette caractéristique est conforme tout à fait à notre objectif, car les cliques qui se ressemblent entre elles doivent être considérées comme un seul groupe vis à vis des cliques très hétérogènes.

Les caractéristiques détaillées du modèle sont expliquées ci-dessous à l'aide d'exemples.

5.2 Caractéristique du modèle

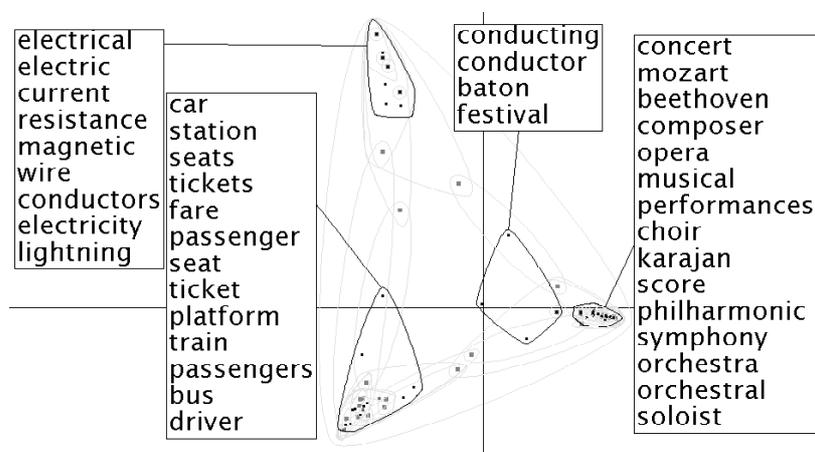
Pour notre modèle, nous avons utilisé comme corpus les textes du Projet Gutenberg (PG) qui comprend des textes, les littératures, des essais, etc. Aucun dictionnaire électronique ou encyclopédique n'a été inclus dans le corpus à traiter. La base de données ainsi construite a été fusionnée avec une autre base de données séparée, construite à partir du *British National Corpus* (BNC). Le nombre total des tokens⁶ dans le corpus est supérieur à 300 millions. Pour les mots français, cinq années de journaux *Le Monde* et *L'Humanité* ont été utilisées.

Avec $\alpha = \beta = \gamma = 0,05$, 50 contexonymes et 133 cliques ont été obtenus pour le mot cible *match*. Une partie de ces cliques est constituée de :

- 1 : applied, mariage, match, proved
- 6 : box, candle, dropped, lighted, match, struck, threw
- 68 : burned, candle, flame, lamp, lighted, lit, match
- 93 : cigar, cigarette, lighted, lit, match, pipe
- 109 : fight, game, match, proved, shot, won

Tandis que chacun des contexonymes *shooting* et *maker* n'appartient qu'à une seule clique, *struck* appartient à 49 de ces cliques. En d'autres termes, *maker* n'est associé qu'à une seule valeur sémantique minimale et *struck* à 49 « valeurs sémant-

⁶Le nombre total de formes d'un texte.

FIG. 5.4 – Représentation de *conductor*.

tiques minimales différentes ». Cette différence est représentée sur la figure 5.2 par la région que chacun des contexonymes couvre, c'est-à-dire, par ses points qui correspondent aux cliques.

La classification peut être faite à partir des cliques ou bien à partir des contexonymes. Dans cette étude, le contexonyme est utilisé pour tester le modèle. La figure 5.3 montre le résultat de cette classification. La figure 5.4 est un autre exemple d'une telle représentation⁷.

En général, des contraintes plus rigoureuses, à savoir des valeurs moins importantes pour α , β et γ produisent des contexonymes moins nombreux que des contraintes moins strictes. Voici certains exemples :

- **drunken** ($\alpha = 0,05$ $\beta = 0,05$ $\gamma = 0,05$) : [brawl] [brute] [drink, drunk, sober, wine] [reeled, staggered] [reeling, staggering] [sailor] [stupor]
- **drunken** ($\alpha = 0,05$ $\beta = 0,10$ $\gamma = 0,05$) : [(coarse, shouts, street, streets, dirty, songs, mad, fellow, driver, driving, dancing, fool, singing,

⁷A partir d'ici, au plus 30 contexonymes les plus étroitement liés sont présentés pour chacun des exemples.

drunken, laughter) (drunk, brute, asleep, wine, drank, sober, drink, song)
 (stupor, sailor, sailors, brawl, crying) (killed) (fury)} {eaten, feast} {brutal,
 mob} {reeled, staggered} {reeling, staggering}} [orgies] [stagger] [swea-
 ring]

- **bucket** ($\alpha = 0,05$ $\beta = 0,05$ $\gamma = 0,05$) : [buckets] [contents, emptied,
 tin] [mop] [rope]
- **bucket** ($\alpha = 0,05$ $\beta = 0,10$ $\gamma = 0,05$) : [{carrying, empty, filled,
 tin, wash, buckets, fill, contents, emptied, poured} {bottom, mop, bucket,
 ice, wooden} {chain, rope} {file, record}] [dipped] [kick, kicked] [packing]
 [pump] [spade]
- **rédigé** ($\alpha = 0,05$ $\beta = 0,10$ $\gamma = 0,05$) : {lui-même, rendu, rédigé,
 signé, acte, article, présenté, publication, daté, document, base, instruction,
 chambre, avis, bureau, communiqué} {intitulé, lire, professeur, commun,
 essentiel, pages, publié, guide, manifeste, ouvrage, langue, mémoire, rap-
 ports} {code} {tiberi} {alinéa}
- **stagger** ($\alpha = 0,05$ $\beta = 0,05$ $\gamma = 0,05$) : [drunken] [reel]
- **staggered** ($\alpha = 0,05$ $\beta = 0,05$ $\gamma = 0,05$) : {clutching, staggered,
 forward, knees} {blow, sank, backward, cry} {stumbled} {drunken, reeled}

où, les crochets indiquent les relations disjointes entre les contexonymes, les accolades désignent les classifications sur le plan principal, et les parenthèses, les classifications sur un plan non-principal. Les contexonymes *kick* et *kicked* pour le mot *bucket* suggèrent l'idiotisme *kick the bucket*, et le contexonyme *article* pour le français *rédigé* reflète leur proximité.

Chapitre 6

Test sur des exemples

Dans ce chapitre nous présenterons les résultats du test du modèle sur des exemples variés. Ce sont des exemples choisis de façon aléatoire, des exemples parus dans les littératures en linguistique computationnelle et en psychologie cognitive.

6.1 Test sur des mots choisis de manière aléatoire

Vingt mots à tester dans le modèle ont été choisis de manière aléatoire. Une méthode d'écart uniforme simple (*uniform deviate method*) s'est montrée inadéquate car, par cette méthode, trop de mots rarement utilisés sont choisis. La méthode consiste à choisir des mots aléatoires parmi tous les types de mots dans le corpus (220 mille mots). A l'opposé de cette méthode on trouve la méthode d'écart uniforme pondéré selon laquelle les mots sont choisis parmi les tokens dans le corpus (comprenant ici 240 millions de mots). Ainsi, si la première méthode consiste à choisir au hasard un mot vedette dans un dictionnaire, la deuxième méthode consiste à choisir un mot en ouvrant un livre dans une bibliothèque. Contrairement à la première méthode, la deuxième méthode donne trop de poids aux mots globa-

lement fréquents tels que *prendre, mettre, devoir* qui ne sont pas significatifs pour tester la performance du modèle à choisir des contexonymes pertinents. Enfin, afin de choisir l'ensemble de mots le plus significatif, les mots d'entrée ont été divisés en trois groupes : le premier est constitué de deux mille mots ayant les fréquences les plus élevées, le second contient les dix mille mots qui les suivent et le troisième est le reste des mots (environ onze mille mots) ; 2, 15, 3 mots ont été ainsi choisis pour chaque groupe respectivement. L'Ecuyer utilise l'algorithme Bays-Durham shuffle afin d'obtenir des mots aléatoires d'écart uniforme (L'Ecuyer 1998).

Les valeurs respectives $\alpha = 0,1$, $\beta = 0,25$, $\gamma = 0,1$ pour les candidats des contexonymes ont été choisies. Le tableau A.1 en Annexe A montre l'ensemble des contexonymes ainsi obtenus pour les 20 mots. Le décompte a été fait avec un découpage automatique des phrases. Les 500 premiers mots les plus fréquents ont été enlevés avant cette sélection. Tous les mots sont présentés en minuscules même si ce sont des noms propres. Les différents groupes classifiés par le calcul sur le plan principal, et positionnés par rapport au centre de masse des cliques, sont représentés par des accolades. C'est-à-dire que, chaque ensemble dans le tableau correspond à un groupe classé dans l'espace sémantico-contextuel. L'ensemble des résultats se trouvent dans le tableau A.1. Les résultats mettent en évidence le fait que le modèle produit des contexonymes pertinents. Par exemple, les contexonymes du mot *blasted* ([health] desséché, [tree] frappé par la foudre, [hopes] anéanti, [annoying] fichu¹) montrent bien les usages du mot : *ruined, tree* et *health* dans le groupe 1 reflètent à la fois le sens DESSÉCHÉ et FOUDROYÉ ; *dammed, hell, hopes* et *curse* reflètent ANÉANTI et FICHU ; et *lightning, oak* pour FOUDROYÉ ; les contexonymes du nom de la ville *Canterbury* reflètent les notions liées à son histoire et à son rôle : *Henry, York, England, Pope, London, cathedral, bishops, cardinal, pilgrims, etc.* Les contexonymes de *eaves* (avant-toit) comme

¹Les définitions du *Collins English-French Dictionary*.

steep, porch, thatch, roofs, straw, sparrows, birds, nest, drip reflètent les images visuelles associées à un avant-toit. Les contexonymes comme *period, chance, rarely* du verbe très général *occur* indiquent son usage, ainsi que ceux de *end* comme *length, middle, opposite, lower, begin, week, etc.*, bien qu'ils soient moins remarquables que les autres exemples.

Outre cette caractéristique, le tableau A. 1. montre également que la classification des contexonymes est correctement atteinte comme on peut le constater sur les exemples *blasted*.

La classification des contexonymes ainsi choisis montre une performance meilleure que la classification non hiérarchique. On peut remarquer que, puisque la classification est faite sur les coordonnées de cliques plutôt denses, cette classification ne doit pas être considérée comme une classification absolument correcte de la géométrie : comme on voit sur la figure 5.4, deux cliques appartenant aux différents groupes pourront avoir une distance sémantique plus courte que celle entre des cliques situées à l'intérieur d'un même groupe.

6.2 Test sur les exemples d'Edmonds et Hirst

Lors de la discussion sur leur modèle, qui tient compte de la quasi-synonymie, Edmonds et Hirst ont examiné de près les différences subtiles entre les mots *blunder, error, lapse* et *slip*, et les paires de mots *order/enjoin, forest/woods* (Edmonds et Hirst 2002).

Nous avons testé ces mots sur notre modèle. Le tableau 6.1 a été obtenu par le test sur le corpus fusionné du GP et du BNC. Le résultat obtenu par le test sur le corpus GP est présenté dans le tableau B.1. en Annexe B. Comme le montre le tableau 6.1, tandis que *blunder* admet les contexonymes *stupid* et *stupidity*, il n'existe pas de tels contexonymes pour *error*, suggérant en effet que ce pre-

Mot	α	β	γ	Contexonymes
blunder	0,05	0,10	0,05	{blunder, mistake} {commit, committed} {stupid}
	0,30	0,30	0,30	[[corrected, political, reckon, blunder, mistake, minister, serious, fatal, pardon, terrible, awful, joke] {unpardonable, gross, committing, stupidity, grievous, stupid, ignorance, guilty, commit, committed, excuse, mistakes} {survived, tragic} {egregious}] [speelman] [tactical]
lapse	0,10	0,20	0,10	{considerable, mere, ten, twenty, lapse, slow, absence, rate, sudden, fifty, forgotten, memory, allowed, minute, months} {evidence, species, progress, century, original, changes, vast, ages, centuries} {recall, moments, interval, minutes} {geological, organic} {strata} {momentary}
slip	0,05	0,10	0,05	{past, slip, tried, run, fall, hold, managed, opportunity, try, allowed, easy, chance, easily} {tree, watch, rope, letting, quietly} {foot, caught, front, drew, fingers, neck, handed, pocket} {tongue, book, paper, written, wrote}
enjoined	0,10	0,20	0,10	{commands, strictly, obedience, instructions, obey, strict, commanded, enjoined, abstain, duty, expressly, orders} {multitude, earnestly, silence} {penance, perform, priests} {secrecy}
ordered	0,05	0,05	0,05	{captain, horse, immediately, company, send, placed, ready, six, service, ordered, party, court, pay} {troops, attack, enemy, city, command, line, horses, officers, war, army, soldiers} {carriage, dinner, master, table} {march} {costs}
error	0,05	0,05	0,05	{trial, evil, human, errors, lead, false, wrong, error, truth, correct, ignorance, fatal, committed, opinion, due, judgment, causes, serious, avoid, fault} {fallen, source, lies, ways, discovered, common, fall} {mistake} {mistaken, supposing}

TAB. 6.1 – Résultats du test sur les exemples d'Edmonds et Hirst.

mier mot a STUPIDITÉ comme connotation contrairement à ce dernier. Des contexonymes comme *unpardonable*, *fatal*, *grievous*, *awful*, *indiscretion* et *egregious* caractérisent le mot cible *blunder* par les traits FORCE(STRENGTH), RESPONSABILITÉ(BLAMEWORTHINESS), et LE CARACTÈRE PÉJORATIF, contrairement au mot *error*. Les contexonymes de *lapse* comme *forgotten*, *memory*, et *minutes* reflètent également l'usage des mots ; les contexonymes *written*, *wrote*, *lines* et *tongue*, parmi les sens différents du mot *slip*, suggèrent qu'il s'emploie pour les erreurs de la parole ou de l'écrit.

Le test sur *woods* a produit les contexonymes *houses*, *path*, *walk*, et *walking*, qui n'ont pas été classés comme contexonymes pour *forest*, tandis que *deer*, *beasts*, *hunting*, *castle* et *knight* ont été considérés comme étant les contexonymes exclusivement réservé à *forest*. Ce résultat est cohérent avec l'observation de Room (1985, comme cité dans (Edmonds et Hirst 2002)).

En général, des différences très subtiles entre des mots similaires, comme discuté dans le modèle FLK (Edmonds et Hirst 2002), ont été observées et raisonnablement représentées dans les résultats de test de notre modèle. En outre, nous avons trouvé d'autres caractéristiques qui n'ont pas été discutées dans leurs études originales : les contexonymes de *error* comme *science*, *fundamental*, *discover* et *method* reflètent l'usage scientifique de ce mot ($\alpha = 0,07$ $\beta = 0,15$ $\gamma = 0,05$ pour le corpus GP) ; les contexonymes *coffee*, *wine*, *supper* et *tea* pour *order* suggèrent que le verbe s'emploie pour demander des boissons ($\alpha = \beta = \gamma = 0,05$ pour le corpus GP). Cette information n'est pas négligeable, car l'anglais *order* dans cette situation devrait être traduit en français par *commander* et non par *ordonner*.

6.3 Test sur les exemples de Dagan et Itai

En discutant le problème du choix d'un mot cible correct dans la traduction automatique des langues, Dagan et Itai ont présenté quelques exemples : *sign* (plutôt que *seal*, *finish* ou *close*) est le verbe correct à utiliser avec *treaty*, et *treaty* (plutôt que *contract*) est le mot correct lorsqu'il est employé avec *peace*. Dans la phrase suivante, le premier mot entre accolades est le choix correct pour chacune des paires (Dagan et Itai 1994) :

Diplomats believe that the joining of Hon Son {**increases** | enlarges | magnifies} the chances for achieving {**progress** | advance | advancement} in the {**talks** | conversations | calls}.

Les cinq mots *peace*, *treaty*, *chances*, *achieved* et *diplomats* ont été testés par le modèle constitué à partir des corpus GP et BNC. Dans le tableau 6.2, on constate que les mots en gras dans la liste des contexonymes sont plus étroitement reliés au mot d'entrée que des mots similaires (par exemple, *sealed*, *closed*, *enlarge*, *conversations*, etc.), qui n'ont pas été choisis comme contexonymes.

6.4 Test avec la méthode de fusion

Puisque les informations pouvant être recueillies à partir des contexonymes du mot cible sont cohérentes, cette capacité peut être utilisée de façon directe pour certaines tâches de TAL. Reprenons la phrase introduite dans le chapitre trois.

The final was Hewitt's first and Sampras' 17th, but the less experienced 20-year-old Australian was much more energetic. After consecutive wins against former champions Pat Rafter, Andre Agassi and Marat Safin, Sampras appeared to have nothing left for his second match in barely 24 hours.

peace	{army, france, peace, war, law, nation, city, sense, free, cause, desire} {earth, happiness, live, spirit, hold, land} {foreign, nations, terms, justice, states, united} {heaven, happy, soul, joy, quiet} {security, treaty }
treaty	{terms, rights, united, states, treaty, france, french, british, foreign, spain, peace , powers, agreed, subject, war, article, britain} {alliance, majesty, concluded, emperor} {territory, commerce, american, citizens, congress, mexico} { signed , ratified, senate}
chances	{calculated, finding, risk, probability, ten, election, favour, nine, success, chances, favor} {game, missed, desperate, plenty, calculate, chance, getting, happen, escape, survival} { increase , increased, reduce, diminished, improve} {promotion, victory, winning}
achieved	{achieved, real, period, process, history, result, increase, development, greater, actually, considerable, progress , growth, results, purpose, per, objectives, using} {economic, aim, position, areas, ways} {success, record, sales, task} {independence, status} {victory}
diplomats	{libya, embassy, p., talks , sanctions, saudi, jan., libyan, reported, iraqi, kuwait, diplomatic, embassies, diplomats, ambassador, foreign} {politicians, journalists, consuls, countries, officials, intelligence, claimed, ministry} {expulsion, expelled, yugoslav} {diplomacy, eighteenth} {statesmen}

TAB. 6.2 – Test sur les exemples de Dagan et Itai.

Entraîné sur les cinq ans de textes des journaux *Le Monde* et *L'Humanité*, notre modèle produit les contexonymes *finale* et *match* pour les mots *Agassi*, *champions* et *victoires* ; et *finale* pour *Rafter* et *Sampras*. Ce résultat met clairement en évidence les mots cibles adéquats pour la traduction.

Toutefois, deux problèmes restent non résolus : en premier lieu, contrairement au stade de la sélection du mot dans la langue cible, aucune désambiguïsation n'est effectuée pour la langue source ; en deuxième lieu, le problème de la rareté des données (Lee et Pereira 1999) n'est pas couvert par cette approche directe. Le premier problème implique d'assigner le sens du mot *match*, dans le paragraphe concerné, vers un cluster correct. Cependant, comme le montre la figure 5.3, il n'y a aucun contexonyme partagé en commun avec le texte cité ci-dessus.

Nous présentons la méthode de fusion des contexonymes comme une solution possible. Cette méthode consiste à consulter plus d'un mot afin d'obtenir les contexonymes. Considérons par exemple le mot W_i^m qui apparaît p fois dans le corpus et W_j^n , q fois comme ci-dessous :

$$W_i^m[p] : c_1[p_1], c_2[p_2], \dots, c_k[p_k], \dots, c_m[p_m]$$

$$W_j^n[q] : d_1[q_1], d_2[q_2], \dots, d_l[q_l], \dots, d_n[q_n].$$

Le tableau de fusion devient :

$$W_i^m + W_j^n[p + q] : c_1[p_1], c_2[p_2], \dots, c_m[p_m], d_1[q_1], d_2[q_2], \dots, d_n[q_n].$$

Si par exemple c_1 et d_2 sont le même mot, ce tableau devient :

$$W_i^m + W_j^n[p + q] : c_1[p_1 + q_2], c_2[p_2], \dots, c_m[p_m], d_1[q_1], \dots, d_n[q_n].$$

Or, si la fréquence globale du mot W_i^m est trop importante par rapport à W_j^n , la contribution de ce dernier pourrait ne pas être prise en compte dans ce type de fusion. Afin de compenser cet effet global de fréquence, on a utilisé la méthode

match + wins	matches, strength, play, chance, won, league, cricket, games, fight, club, final, game, points, united, record, victory, team, season, marriage, daughter, prix, match, box, draws, wins, loses, lighted, cigarette, lit, whoever
match + wins + champions	won, game, division, champions, defeat, final, united, games, champion, club, yesterday, season, draw, points, victory, strength, fight, play, team, players, australia, chance, defending, match, record, matches, wins, daughter, struck, lighted
match + agassi + sampras	tennis, champion, minutes, struck, play, final, davis, won, doubles, yesterday, players, michael, game, player, jim, tournament, today, defeat, victory, weeks, opening, monday, sets, agassi, match, wimbledon, edberg, goran, ivanisevic, sampras
match + champions + wins + agassi + sampras	final, champion, won, defeat, play, game, champions, doubles, player, league, players, tennis, davis, sets, minutes, michael, tournament, team, victory, season, yesterday, today, wins, weeks, jim, agassi, match, edberg, ivanisevic, sampras

TAB. 6.3 – L'effet de fusion pour *match* et ses voisins ($\alpha = \beta = \gamma = 0,05$). Les contextonymes sont ordonnés par la proximité à l'origine, en d'autres termes par la représentativité.

de fusion normalisée. Cette méthode donne un nouveau tableau :

$$W_i^m + W_j^n [2N] : c_1 \left[\frac{Np_1}{p} + \frac{Nq_2}{q} \right], c_2 \left[\frac{Np_2}{p} \right], \dots, c_m \left[\frac{Np_m}{p} \right], d_1 \left[\frac{Nq_1}{q} \right], \dots, d_n \left[\frac{Nq_n}{q} \right],$$

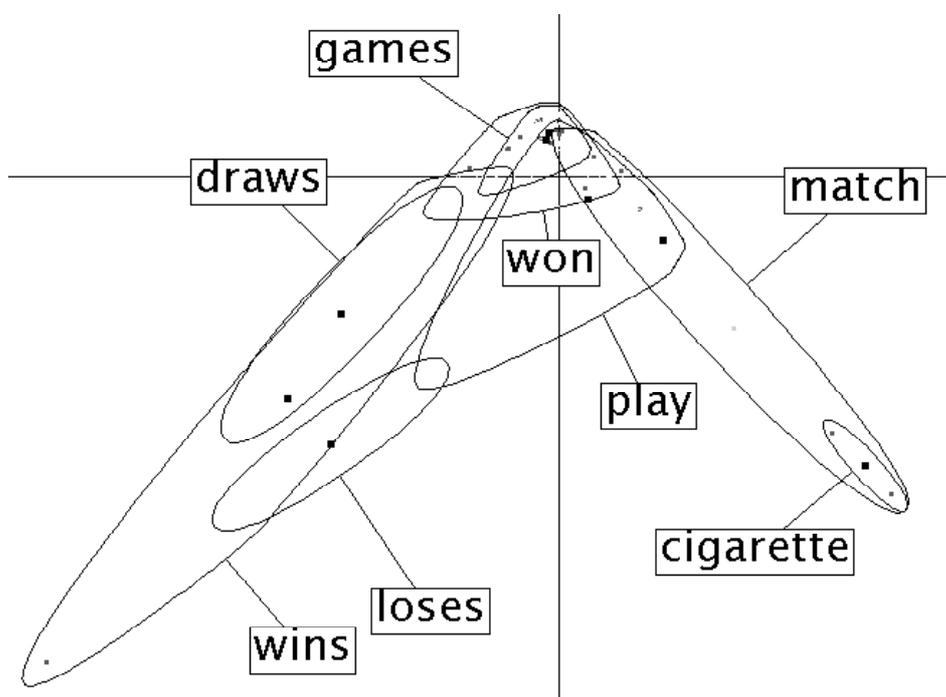
où, N est un nombre suffisamment grand.

Le tableau 6.3 met en évidence l'exclusion graduelle des contexonymes les moins reliés comme *cigarette* et *marriage*, et l'inclusion graduelle des contexonymes les plus pertinents comme *player* et *tennis*. De la même manière, dans la représentation géométrique de *wins + match* (Fig. 6.1), les régions correspondant aux contexonymes *play* et *won* se situent au centre, marginalisant la région qui correspond à *cigarette*.

Une autre manière de faire la distinction entre les sens du mot cible consiste à observer les *contexonymes de liage*. Par exemple, *match* et *wins* n'ont pas de cliques partagées et les zones qu'ils recouvrent sont disjointes dans la projection principale. Cependant, ils sont connectés par des contexonymes intermédiaires *play* et *games*, qui partagent à leur tour des zones avec les cliques de *match* et celles de *wins*.

Une solution potentielle pour le problème de la rareté des données est de construire une liste de décision, comme proposé par Yarowsky (1995), en utilisant les contexonymes comme mots clefs initiaux. Pour une tâche de désambiguïsation, la classification par les points correspondant aux cliques, et non par les points correspondant aux contexonymes, est le meilleur choix.

Bien que le modèle actuel ne soit pas conçu pour résoudre un tel problème, la méthode peut être considérée comme une solution possible. Par exemple, *stars* et *mathematician* ne partagent aucun contexonyme mais ils sont reliés par *astronomer* dans la méthode de fusion, suggérant que *stars* devrait être interprété comme ÉTOILE et non VEDETTE.

FIG. 6.1 – Fusion de *wins* et *match*.

Proportion des sujets qui ont produit chaque réponse	Nombre de contexonymes	Nombre de contextes identiques	Nombre de mots communs entre les sujets	Nombre de sujets / Nombre de contexonymes
Pour tous les mots	588		2356	25 %
Plus de 10 %	169		315	54 %
Plus de 20 %	85		129	66 %
Plus de 30 %	49		70	70 %
Plus de 40 %	31		40	78 %
Plus de 50 %	19		24	79 %

TAB. 6.4 – Résultats du test sur les exemples de Hirsh et Tree.

6.5 Test sur les exemples de Hirsh et Tree

Hirsh et Tree (2001) ont testé 90 sujets, adultes, sains, sur les associations de mots. Les sujets sont répartis en deux groupes : 45 sujets âgés (entre 66 et 81 ans) et 45 sujets jeunes (entre 21 et 30 ans). Pour l'ensemble des mots présentés, les sujets écrivent le mot qui leur vient à l'esprit immédiatement. Par exemple, pour le mot *beurre*, les sujets pourront écrire *pain*, *lait* ou *verre*. L'ensemble des 90 mots a été testé et dans l'annexe C on trouvera leurs résultats.

Comme on peut le voir dans le tableau 6.4, le taux d'appariement entre les contexonymes produits par le modèle et les mots associés choisis par les sujets augmente si les mots choisis par les sujets font état d'un consensus global. Par exemple, pour l'ensemble des mots associés choisis par moins de dix pourcents des sujets, le taux d'appariement des contexonymes du modèle est 54%, tandis que pour les mots choisis par 40 pourcents des sujets, le taux d'appariement est 78%.

Les résultats plus détaillés du test sont ajoutés en troisième colonne pour comparer ces deux résultats. Comme on peut le constater, notre modèle produit des contexonymes très pertinents comparables aux résultats des sujets humains.

6.6 Test sur deux corpus différents

Comme nous l'avons dit dans les chapitres précédents, la représentation d'un concept ou du sens des mots différerait selon l'environnement. D'ailleurs, nous avons avancé, dans le chapitre deux, l'idée que malgré la différence environnementale, la notion centrale attachée au concept reste identique.

Afin de voir si ces deux caractéristiques que nous avons soutenues sont respectées dans le modèle, nous avons testé quelques mots pour le modèle qui a appris séparément deux corpus différents. Les corpus utilisés sont cinq ans de textes des journaux *Le Monde* et *L'Humanité*.

Comme on peut le constater dans le tableau 6.5, pour les mots censés être identiques par leur sens dans ces deux journaux, comme par exemple *film*, il n'existe pas de différence significative entre les contexonymes obtenus séparément. En revanche, les mots censés être différents entre les deux journaux, comme par exemple *politique*, sont représentés différemment par le modèle : le résultat du modèle qui a utilisé *L'Humanité* comme corpus présente des contexonymes différents de ceux obtenus en utilisant *Le Monde* comme corpus. Par exemple, les contexonymes *PCF*, *communistes*, *Hue*, *chômage* pour le mot *politique* se distinguent des contexonymes du *Le Monde*.

Par ailleurs, la spécificité commune des corpus, à savoir les textes de journaux est reflétée par les résultats obtenus pour le mot *américain*. Les contexonymes produits par le modèle figurant dans le tableau 6.5 reflètent en effet l'intérêt des médias pour l'aspect politique.

Mot	α	β	γ	Contexonymes
film Le Monde	0,01	0,01	0,01	{ films, festival, réalisateur, réalisé, v, arte, cinéaste, min, scénario, noir, o, tourné, personnages, images, john, roman, personnage, cannes, sortie, documentaire, comédie, amour, meilleur, tournage, tf, e, série, acteurs, vidéo, genre }
film L'Humanité	0,01	0,01	0,01	{ caméra, cinéaste, images, personnages, regard, télé, réalisateur, montre, filmer, films, parle, réalité, enfant, usine, caméras, voit, montage, vérité, amour, tournage, écran, écriture, distance, spectateur, sorte, personnage, type, engagement, minutes, trente }
politique Le Monde	0,01	0,01	0,01	{ monétaire, étrangère, classe, commune, démocratie, discours, militaire, centrale, régime, partis, soutien, bureau, front, responsabilité, véritable, budgétaire, socialiste, relations, unique, élection, dirigeants, économiques, parlement, solution, opinion, égard, ministres, monnaie, élus, processus }
politique L'Humanité	0,01	0,01	0,01	{ pcf, communistes, citoyens, démocratie, partis, changement, monétaire, véritable, commune, intervention, hue, vote, sociaux, plurielle, peuple, chômage, opinion, engagement, solution, construction, sociales, discours, humanité, processus, unique, objectif, changer, socialiste, ouvre, classe }
américain Le Monde	0,01	0,01	0,01	{ clinton, bill, washington, britannique, mondial, new, e, trésor, john, département, york, allemand, george, numéro, filiale, congrès, dollar, géant, japonais, modèle, richard, o, russe, mondiale, bush, site, commerce, michael, capital, rachat }
américain L'Humanité	0,01	0,01	0,01	{ clinton, bill, washington, mondial, e, peuple, bush, congrès, new, champion, otan, britannique, russe, militaire, york, onu, george, visite, mètres, john, irak, numéro, chine, william, noir, département, processus, mondiale, indiqué, kosovo }

TAB. 6.5 – Résultat du test pour deux corpus séparés (*Le Monde* et *L'Humanité*). Une fenêtre de taille 25 a été utilisée pour chaque cas.

Outre ces résultats, plus de 100 000 types de mots peuvent être testés sur le site <http://dico.isc.cnrs.fr/dico/context/search>.

Chapitre 7

Discussion et conclusion

7.1 Une étude pluridisciplinaire

Dans cette étude nous avons examiné des sujets assez variés comme la notion de concept, le sens des mots et les modèles informatiques, psycholinguistiques et linguistiques. Les disciplines liées à ces sujets sont aussi très diverses — la philosophie, la linguistique, la psycholinguistique, la linguistique computationnelle, l'intelligence artificielle, les sciences cognitives, etc. Il est vrai que chacune de ces disciplines garde son propre intérêt distinct et qu'elle adopte une approche spécifique à son objectif. Par exemple, en ce qui concerne le concept, la philosophie s'intéresse à la nature fondamentale du concept, la linguistique au lien entre le concept et les sens des mots, la psycholinguistique au lien entre représentation et processus mentaux et la linguistique computationnelle à la modélisation formelle. De ce fait, traiter dans une étude des sujets censés appartenir à ces différentes disciplines peut comprendre le risque d'étudier ces sujets de manière superficielle ou bien de dévier de la direction principale en examinant chaque sujet de façon trop détaillée.

D'un autre côté, une étude pluridisciplinaire a des avantages par lesquelles on

pourra compenser ce type de risque. En effet, un simple parcours de l'histoire des sciences montre des effets constructifs et interactifs entre différents domaines de recherche.

D'une part, les sciences issues de domaines différents s'influencent entre elles en s'inspirant du mode de pensée de chacune d'elles. La physique newtonienne par exemple a engendré le déterminisme mécaniste comme celui de Laplace tandis que la théorie de la relativité a influencé le relativisme comme le relativisme éthique chez Weber (1917/1949, 1919/1975). L'influence des sciences sociales sur les sciences naturelles existe aussi comme par exemple celle de la théorie de Malthus sur la théorie darwinienne. De manière similaire, la prospérité florissante du behaviorisme avant les années 60 a influencé plusieurs domaines de recherches et réciproquement son déclin est lié aux critiques provenant d'autres disciplines.

D'autre part, outre ces aspects interdisciplinaires, les changements au sein des disciplines elles-mêmes au cours de ces dernières décennies méritent notre attention. Par exemple, le domaine de la recherche linguistique s'est largement étendu aujourd'hui par rapport à ce qu'il était au début du dernier siècle : la pragmatique et la prosodie en sont des exemples. Ce type d'élargissement intradisciplinaire est en effet dû au fait qu'il n'est pas possible de progresser sans rendre compte d'aspects considérés comme initialement non reliés. En effet, en ce qui concerne la linguistique, l'approche restreinte à la sémantique lexicale montre une limite pour l'explication de divers aspects linguistiques. Par exemple, Reboul (à paraître) remarque que la théorie du lexique génératif dont l'intérêt porte plutôt sur le lexique sémantique a des difficultés à expliquer les différents sens du mot *began* entre la phrase « Mary began her book » et la phrase « Virginia Wolf began her book » à moins que l'on ne recoure à une connaissance encyclopédique voire pragmatique.

Outre cette tendance à l'influence mutuelle interdisciplinaire et à l'extension intradisciplinaire, la reprise de méthodes considérées comme obsolètes permet de

faire progresser les études. Par exemple, les méthodes statistiques, qui sont considérablement appréciées aujourd'hui dans les applications du TAL, ont en réalité été rejetées dans les années 60, à la fois théoriquement et pratiquement, face à l'émergence de méthodes non statistiques. De fait, une théorie linguistique qui préconise une approche non statistique, notamment celle de Chomsky, est un progrès évident contre une approche statistique naïve. Ce progrès incontestable n'explique cependant pas tous les aspects linguistiques ; il ne garantit pas non plus la possibilité d'avoir un système capable de traiter de manière parfaite des tâches de TAL voire de comprendre le langage humain. De ce fait, la réapparition de méthodes statistiques dans le domaine du TAL qui montre des résultats intéressants pour les études linguistiques computationnelles ne doit pas être comprise comme un retour en arrière mais devrait être comprise comme la négation de la négation de Hegel, autrement dit comme un progrès. En effet le progrès est souvent observé dans des études par alternance de deux attitudes opposées, ce qui n'est pas néanmoins un simple jeu à somme nulle. Le retour du dualisme cartésien contre le monisme naïf par exemple est un progrès au lieu d'une simple répétition du dualisme de l'antiquité ; inversement, le monisme hégélien est un progrès vis-à-vis du dualisme qui est devenu cette fois-ci naïf et incomplet relativement à ce retour du monisme.

Notre position se situe dans cette perspective. Nous préconisons un progrès possible en adoptant une approche statistique au sein de la linguistique sémantique, de la linguistique computationnelle et de la modélisation psycholinguistique. En ce sens nous indiquons la limite d'une approche non statistique telle que l'approche syntaxique, symbolique, etc. Nous reconnaissons néanmoins que la combinaison de ces deux types d'approches ou bien que les approches non statistiques qui utilisent les résultats d'approches statistiques puissent porter d'autres types de progrès envisageables. Ainsi, l'absence d'une discussion profonde sur

la syntaxe dans cette étude ne présage pas l'avantage et une légitimité absolue de notre approche statistique ; de fait, la voie est ouverte pour une amélioration utilisant des approches différentes de la nôtre.

Compte tenu de ce fait, nous discuterons ce que nous avons apporté dans cette étude et ce qu'il faudra faire en perspective.

7.2 Modélisation du concept

Notre modèle propose une représentation complexe et classificatoire des concepts. Par exemple, les différents sous-ensembles de contexonymes du mot *match* montrent clairement l'existence de différents concepts distincts du mot d'une part et la possibilité d'apprendre ces concepts de manière automatique d'autre part.¹ Le concept ainsi représenté par les contexonymes explique en effet plusieurs aspects relatifs à l'étude des concepts.

D'abord, le modèle propose une représentation des concepts de manière à la fois graduelle et distinctive. La critique fondamentale contre la nature « polysémique » du concept chez les atomistes comme Fodor et Laurence a deux types d'arguments. D'une part, la séparation des différents concepts d'un mot est dépourvue de critère général. Par exemple, d'après Fodor, la distinction du *keep* de Jackendoff² n'est pas exhaustive, ce qui rend cette séparation plus ou moins ad hoc. D'autre part, si nous avons bien interprété l'argument de Fodor, la division des concepts d'un mot conduira à un enchaînement des divisions à l'infini, puisque l'établissement d'un critère orthodoxe de ce type de division n'est pas possible.

¹Dans ce cas, les différents concepts MATCH COMME SPORT, ALLUMETTE et APPARIEMENT correspondent aux différents sens du mot MATCH.

²Jackendoff distingue différentes notions : l'emplacement et la spécificité (Harry kept the bird in the cage), la possession (Susan kept the money), un script de propriétés (Sam kept the crowd happy), un programme d'activité (Let's keep the trip on Saturday).

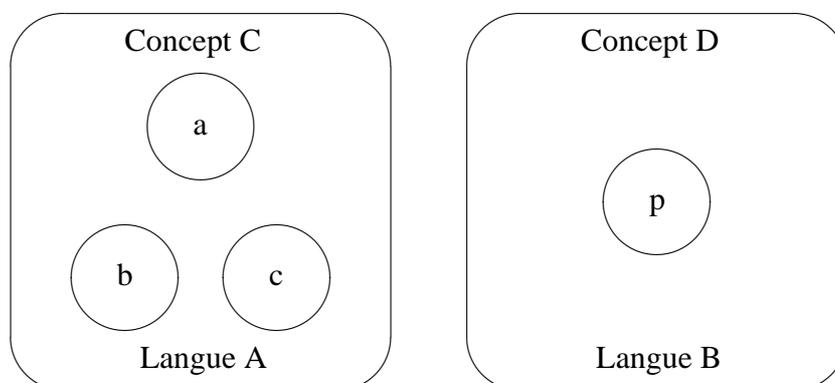


FIG. 7.1 – Représentation schématisée du concept C dans la langue A et du concept D dans la langue B. a, b, c et d sont les concepts de bases.

Le fait que notre modèle représente les concepts avec un critère très formel et général au lieu d'un critère spécifique pour chaque mot répond à la critique de Fodor. La représentation graduelle des concepts nous permet d'éviter d'adopter des définitions afin de séparer les différents concepts d'un mot. L'absence de définition n'entraîne pas cependant l'absence de distinction entre les différents concepts. Les régions localisées dans la représentation du mot reflètent en effet très nettement les différents concepts d'un mot comme MATCH POUR SPORT, ALLUMETTE et APPARIEMENT pour le mot *match*. Ces concepts distincts correspondent au concept de base que nous avons proposé dans le chapitre deux. Cette distinction est très nette pour un mot comme *conductor* et moins claire pour d'autres exemples.

Par ailleurs, tous les mots ne sont pas nécessairement divisés en différents concepts : il existe des mots qui correspondent à un seul concept de base. Par exemple, *astronef*, *croquet*, *devancier* et *ostéomalacie* peuvent être considérés comme ayant un seul concept de base. La figure 7.1 schématise cette notion d'équivalence entre un concept et un concept de base. Le concept C dans la langue A est composé de

trois concepts de base a , b et c , et le concept D dans la langue B ne possède qu'un seul concept de base p . Si le concept a et le concept p sont équivalents, leur représentation peut être similaire. Si le concept ALLUMETTE (a dans la figure 7.1) pour le mot anglais *match* (correspondant au concept C) est considéré comme concept de base, le mot français *allumette* (correspondant à la fois au concept D et au concept p) doit l'être. Ainsi dans la représentation géométrique, la région globale de p ou D (R_1) représente une structure aussi simple que celle de la région localisée du concept de base a (R_2) dans la représentation du concept C .

Cette interprétation de la représentation du modèle est intuitive aussi, car la séparation des concepts est tantôt évidente pour des mots comme CHAUD, FROID, INSENSIBLE, etc. tantôt rare pour des mots comme ASTRONEF, CROQUET, DEVANCIER, OSTÉOMALACIE, etc. Le concept de base du modèle n'empêche néanmoins pas d'analyser de manière plus profonde le concept associé au mot. Comme on peut le constater dans la représentation du mot *conductor*, les liens entre les régions correspondant à un concept de base sont encore analysables. Par exemple, le recouvrement de *tickets* dans la représentation du mot *conductor* entre les concepts de base liés à la MUSIQUE et au TRANSPORT met en évidence le fait que le terme « tickets » est un trait partagé par ces deux concepts de base et ces deux concepts seulement (Fig. 7.2).

Ces doubles caractéristiques à savoir la représentation du concept par des concepts de base et la capacité à analyser plus profondément les concepts de base permettent d'établir une nouvelle théorie du concept. D'une part, l'existence de concepts de base permet de maintenir un « principe partagé ». On sait très bien qu'un mot dans une langue n'a pas fréquemment son équivalent dans une autre langue. Malgré cette discordance, il est difficile d'affirmer qu'il n'y a pas d'équivalent conceptuel entre les gens ayant des langues maternelles différentes,³ car dans ce

³Il existe des recherches qui affirment cet argument, à savoir que le concept chez l'humain

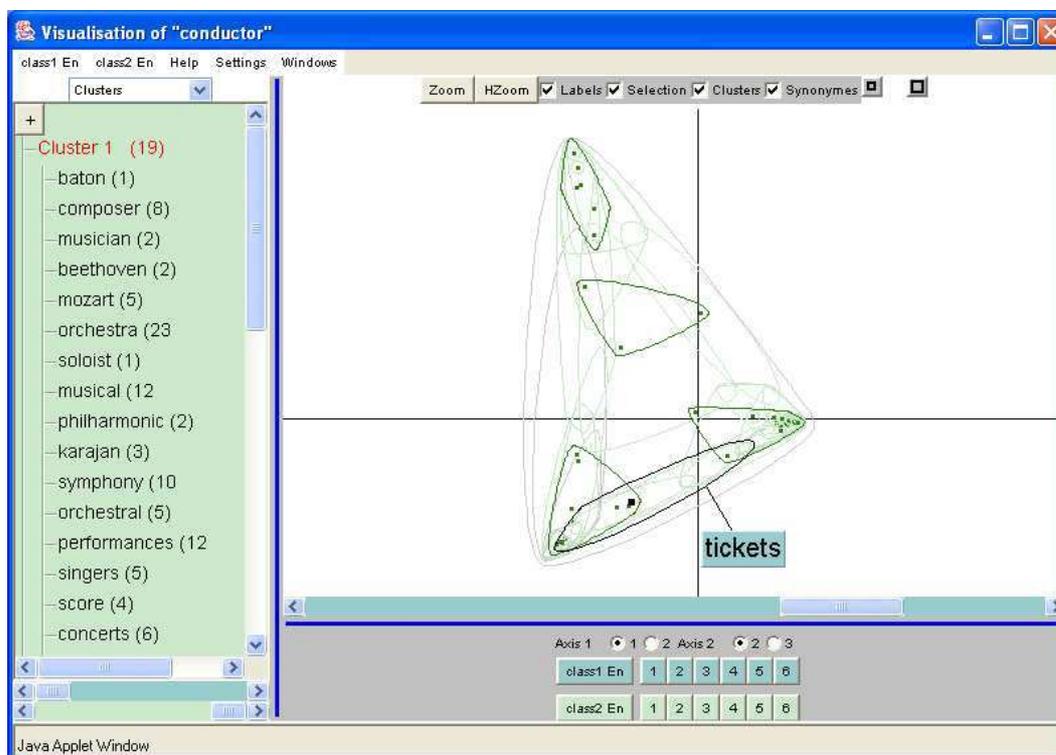


FIG. 7.2 – La région correspondant à *tickets* dans la représentation de *conductor*.

cas les caractéristiques cognitives communes des humains devraient être mises en question. Le concept de base est en général commun sans aucun égard aux langues. Par exemple, le concept ALLUMETTE du mot anglais *match* équivaut au concept ALLUMETTE du mot français *allumette*. Ainsi, notre modèle avec sa notion de concept de base s'oppose à la position des atomistes qui identifie un mot en tant que concept en refusant sa décomposition. Et il nous semble que l'argument atomistique rencontre des difficultés majeures vis-à-vis de ce type d'exemples.

D'autre part, la nature continue de notre modèle permet d'analyser de manière plus profonde un concept de base. En effet, chacune des cliques appartenant à est différent selon la langue utilisée mais cette observation ne s'applique qu'à des cas spécifiques (Bloom et Keil 2001).

un concept de base se positionne différemment, représentant ainsi les différences subtiles linguistiques ou extralinguistiques.

Par ailleurs, la représentation par les contexonymes permet de dépasser les limites de la théorie définitionnelle du concept. Le contexonyme *drunk* pour le mot *stagger* par exemple reflète bien des connaissances difficiles à définir par une approche définitionnelle. Cet aspect sera discuté plus en détail lors de la discussion portant sur l'aspect informatif du modèle.

Nous donnons ici un exemple montrant que le modèle répond bien à l'observation psycholinguistique d'effet prototypique et les observations selon lesquelles le concept chez l'homme n'est pas organisé comme une définition. D'une certaine manière, notre modèle suggère une explication possible pour l'effet prototypique : d'une part le recouvrement central et générique d'un contexonyme dans la représentation de *bird* ou *sparrow+swan+robin+dove+eagle* ; d'autre part, les traits communs que possède le concept prototypique sont représentés par des liens plus importants entre ce concept et les autres concepts.

Les liens du modèle avec l'extension du concept

Dans notre modèle, un concept est représenté par ses contexonymes et par l'organisation de ces contexonymes. Autrement dit, il n'existe pas de définition pour un concept. Malgré l'absence de définition, la capacité à pouvoir séparer les différents concepts de base, comme nous venons de le voir, permet à ce modèle de répondre à certains problèmes relevant de l'étude des concepts. En outre, notre modèle propose une solution authentique à l'extension du concept comme nous allons le voir.

Comme nous l'avons examiné dans le chapitre deux, la définition d'un concept rencontre deux problèmes majeurs concernant l'extension du concept. D'une part, la définition ne correspond pas parfaitement à l'extension. Par exemple, la défini-

tion du poisson comme « animal vertébré inférieur, vivant dans l'eau et muni de nageoires »⁴ ne s'applique pas aux poissons sans nageoire. D'autre part, il est très contestable qu'un concept qui englobe toutes les extensions (contenu large) puisse être « le » concept. De fait, Fodor présuppose ce type de concept lorsqu'il dit que les contenus du concept de l'eau sont différents entre la terre et la Terre Jumelle, tandis que Putnam sépare le concept de son extension.

Dans notre modèle, le concept n'est représenté que par les contexonymes et par leur organisation. Cette déclaration ne présuppose donc pas qu'un concept puisse recouvrir ou non son extension : le concept peut comprendre soit un contenu étroit soit un contenu large. En d'autres termes, dans certains cas, la représentation d'un concept peut englober toutes les extensions et dans d'autres cas, une partie des extensions seulement. Cette idée suppose en effet que le recouvrement parfait incluant toutes les extensions n'est pas la condition nécessaire pour la constitution d'un concept.

Examinons maintenant comment notre modèle répond au vif débat autour de l'extension du concept. Les deux mots *elm* et *beech* qui relèvent d'un concept unique — au moins pour Putnam — révèlent un problème sévère concernant le lien entre représentation mentale et monde. Selon notre hypothèse et l'architecture du modèle, si les deux concepts partagent les mêmes contexonymes et la même organisation, ces deux concepts sont considérés comme identiques. De ce fait, l'identité du concept entre *elm* et *beech* correspond à l'identité des contexonymes et à l'organisation de ces mots.

De fait, il est difficile de trouver de tels exemples, car ce genre de redondance linguistique — à savoir l'utilisation de deux mots distincts pour désigner un même concept — est un cas assez exceptionnel, ce qui est corrélé d'ailleurs à l'observation qu'il n'existe guère de synonymes absolus : même si deux synonymes se

⁴La définition du *Robert*.

ressemblent considérablement, il y a toujours des différences très subtiles (voir *forest/wood* ou *blunder / mistake / lapse* par exemple).

En revanche les contexonymes de *beech* et *elm* montrent une similarité remarquable : *trees, tree, oak, wood, leaves, woods, green, forest, road* pour *beech* et *trees, tree, oak, street, green* pour *elm* avec les paramètres fixés à 0,02. Cette similarité suggère en effet que les deux concepts sont quasi-identiques et en ce sens nous partageons le point de vue de Putnam qui affirme leur identité.

D'après Putnam, l'affirmation de l'identité du concept nuit gravement à la validité de la théorie de la représentation mentale, car d'une part, il est évident pour Putnam (et Fodor) qu'il existe — puisqu'il y a deux mots différents et deux types d'arbres — une différence entre *elm* et *beech*, et d'autre part, la représentation mentale ne peut pas expliquer cette différence « évidente », puisqu'elle ne peut pas distinguer deux choses différentes.

Quant à Fodor, sa position se résume ainsi : en général, si l'extension est différente le concept est différent aussi. Fodor a reconnu que les concepts de l'eau entre les gens de la terre et la Terre Jumelle sont différents. Par cette affirmation, il embrasse un contenu large, autrement dit, il accepte que le concept soit défini par une caractéristique externe. La confusion entre deux concepts différents est considérée soit comme improbable (cas pour la Terre Jumelle) soit comme étant une erreur ou un acte non attentionnel (cas pour *elm et beech*). Fodor compare cette situation à l'usage du papier de tournesol : de la même manière qu'on peut consulter le papier tournesol pour savoir si un liquide est acide ou non, on peut consulter un expert pour savoir si l'arbre en question est un *elm* ou un *beech*.

Au fond, chez Fodor, les notions telles que expert, acte non attentionnel et erreur sont des moyens qui défendraient sa théorie de la représentation mentale qui est, d'après lui, censée expliciter toutes les situations mentionnées.

Notre modèle peut, à notre avis, répondre à cette question. Notre solution con-

siste à générer la représentation du concept de manière dynamique. La représentation du concept dans notre modèle est en principe basée sur le corpus sur lequel le modèle a été « entraîné ». Cet aspect ascendant du modèle prévoit que la représentation des contonymes varie, si le corpus utilisé varie. De fait, la représentation du mot *vision* par exemple basé sur un corpus général comme BNC et sur un corpus spécialisé créé à partir d'articles de *The MIT Encyclopedia of the Cognitive Science* est très différente :

- vision : dream, clear, eye, future, soul, sight, beyond, human, sense, field, dark, beauty, rose, range, form, mental, past, earth, spirit, across, figure, line, strange, bright, truth, reality, sudden, dead, sound, picture ($\alpha = \beta = \gamma = 0,01$ pour le corpus « BNC + GP »).
- vision : visual, color, objects, processing, mid-level, motion, Vision, information, perception, high-level, cortex, image, spatial, Color, object, conscious, brain, field, areas, human, studies, recognition, N, example, surface, system, primates, role, mechanisms, models ($\alpha = \beta = \gamma = 0,03$ pour le corpus d'articles de sciences cognitives).

De même, il est facile d'imaginer que les concepts *beech* et *elm* seront différents si un corpus important basé sur la botanique est utilisé comme ressource d'entraînement. Ainsi s'explique la différence des représentations des concepts *elm* et *beech* chez les experts. Il faudra noter que ce type de résultat est obtenu sans modification du principe du modèle.

Cette approche est très intuitive aussi, car le concept associé aux mots rarement utilisés comme *méronyme*, *hyperonyme*, etc. est, contrairement à ce qui se passe pour les spécialistes, absolument vide pour les gens n'ayant pas de telles connaissances. Il est incontestable que la différence entre ces deux groupes de gens provient du fait qu'un groupe utilise, voit ou entend plus fréquemment ces mots que l'autre.

Cependant, ce genre d'approche qui différencie les concepts par rapport aux groupes (sociologiques par exemple) risque de se heurter à l'individualisme. En ce sens, on peut comprendre la raison pour laquelle Fodor s'est détourné de ce type de solution concernant les problèmes de la Terre Jumelle et *elm/beech*. Cependant, notre approche qui consiste à utiliser différents corpus pour répondre à ce genre de problème est loin d'être une approche ad hoc, car d'une part le principe du modèle ainsi que son architecture ne sont pas modifiés par un changement de corpus et d'autre part, un changement de la représentation du concept selon le corpus utilisé est tout à fait légitime (ceci permet par exemple d'utiliser le modèle pour un corpus varié provenant de différentes langues). Plus précisément, bien que les humains possèdent, grosso modo, les mêmes concepts de base, indifféremment des langues utilisées, il existe toujours des différences subtiles entre les concepts de base, par exemple entre *neck* et *cou*. Ainsi, de même qu'il est possible d'avoir, chez les humains, différentes représentations du concept selon la langue utilisée, il est tout à fait possible également de posséder différentes représentations pour un concept spécifique entre les spécialistes du domaine et les non spécialistes.

En résumé, la clef de notre approche concernant l'extension et le contenu du concept consiste à accepter le concept dynamique qui varie selon les différentes langues ou les différents domaines qui reflètent différentes représentations mentales dépendant de la spécificité — langue ou domaine. Cette approche est ainsi radicalement différente de l'approche qui tend à distinguer le contenu étroit/large ou l'extension/intension du concept.

7.3 Modèle sémantique

Notre modèle n'est pas enfermé sur le mot lui-même. Comme nous venons de le voir, dans le modèle, plusieurs mots peuvent partager un même concept et

un mot peut contenir plusieurs concepts. Pour une langue donnée, le modèle est ainsi construit : un même sens peut correspondre à plusieurs mots et plusieurs sens peuvent correspondre à un mot. En ce sens, l'unité fondamentale dans le modèle est l'unité de base du sens et non pas une unité lexicale comme le mot ou le lexème. L'unité minimale du sens, à savoir la clique, quant à elle, permet d'organiser les contonymes d'un mot pour représenter le sens des mots de manière graduelle et distinctive.

Quant à la distinction entre connaissance encyclopédique et connaissance sémantique, notre modèle ne distingue pas cette dichotomie. Par exemple, pour le concept *tennis*, *wimbledon* ou *Roland-garros* sont proposés *court*, *courts*, *play*, *table*, *game*, *lawn*, *playing*, *club*, *golf*, *players*, *played*, *ball*, etc comme contonymes parmi d'autres. Ainsi ce modèle peut représenter des connaissances encyclopédiques assez générales comme le montrent les contonymes de *Egypt*, *Marx*, etc. :

- Egypt : land, Syria, Israel, ancient, Nile, Moses, sea, egyptian, Joseph, India, Egyptians, pharaoh, war, city, east, army, Arabia, Greece, countries, upper, Asia, gods, Africa, Arab, Rome, desert, history, kings, Cairo, river, ..., pyramids, ..., Cleopatra, Alexander⁵
- Marx : Engels, social, society, class, theory, history, production, Karl, political, capitalist, ideology, capitalism, labour, economic, capital, ideas, property, system, historical, development, societies, human, relations, form, different, concept, analysis, Althusser, Hegel, german, account, Morgan, modern, revolution, writings, argued, division, Lenin, proletariat, Weber, material, particular, private, working.⁶

Cette caractéristique du modèle est liée aussi à la représentation du stéréotype

⁵ $\alpha = \beta = \gamma = 0,01$ et $0,03$ pour le corpus « BNC + GP ».

⁶ $\alpha = \beta = \gamma = 0,03$ pour le corpus « BNC + GP ».

du mot que nous allons voir plus bas. En ce sens, ce modèle a un lien avec la linguistique sociale.

Il est souvent énoncé que l'opinion publique est gérée par les médias (Chomsky 2002). Ceci met en question en effet la représentation mentale que l'on crée à partir des informations passées dans les médias. Par exemple, si un pays est souvent présenté par l'utilisation de certains mots de manière répétitive, ces mots risquent de former le stéréotype du pays. Notre modèle reflète ce type de tendance. Par exemple, tandis que les contexonymes du mot *Colombia* basé sur un corpus général sont *peru, venezuela, states, united, mexico, brazil, countries, panama, ecuador, bolivia, president, argentina, american, chile, republic, drug, treaty, etc.* les contexonymes basés sur des textes de journaux sont plus stéréotypiques : *Farc, armées, révolutionnaires, guérilla, brésil, mexique, venezuela, Amérique, argentine, Pérou, bogota, drogue, Chili, équateur, lutte, paramilitaires, violence, bolivie, colombien, latine, groupes, pastrana, trafic, etc.* Ces résultats mettent en évidence d'une part une caractéristique de la représentation du modèle qui varie selon le corpus utilisé et d'autre part le mécanisme de la création des stéréotypes par simples répétitions.⁷

De plus, notre modèle ne distingue pas au préalable l'homonymie et la polysémie. Cette caractéristique est conforme aux résultats des expériences psycholinguistiques qui n'observent pas de différence systématique entre les mots homonymiques et les mots polysémiques (Klein et Murphy 2001).

⁷Sous un autre angle, cette observation sera très utile. D'une part, on peut observer une représentation moyenne dans une société pour un pays ou un groupe si on dispose d'un corpus suffisamment large et adéquat ; cette représentation sera plus que la connaissance trouvée dans un encyclopédie. D'autre part, si la représentation d'un terme/concept est trop biaisée, comme c'est le cas pour la Colombie, les médias pourraient essayer de prendre en compte les résultats afin d'éviter quand cela est nécessaire l'emploi de certains mots qui sont déjà trop répétés puisqu'ils sont déjà inclus dans l'ensemble des contexonymes.

Le fait que le modèle ne distingue pas l'homonymie et la polysémie n'entraîne cependant pas que la manière dont le modèle représente le sens des mots soit identique. L'organisation du sens peut en effet être classée de deux façons : l'organisation « homonymique » et l'organisation « polysémique ». L'organisation homonymique d'un mot désigne une représentation selon laquelle la superposition des champs sémantiques des contexonymes est relativement réduite, alors que l'organisation polysémique désigne une représentation pour laquelle la superposition est plus importante. Par exemple, le recouvrement mutuel des trois sens de *conductor* (ÉLECTRIQUE, CHEF D'ORCHESTRE et RECEVEUR) est moins important que ceux de *match* dans la projection du plan principal de l'espace sémantique. Cela est dû au fait que certains contexonymes sont plus partagés pour ce dernier mot que pour le premier.

Il faudra noter que cet aspect homonymique/polysémique ne correspond pas systématiquement aux mots homonymiques/polysémiques. En ce sens, notre modèle propose d'établir des liens sémantiques d'un mot de façon systématique à partir des données au lieu de le faire d'une manière historique ou étymologique. Par exemple, les représentations de certains mots polysémiques peuvent être homonymiques et non polysémiques dans notre modèle. Cette position recouvre la position d'auteurs comme Palmer (1981) qui estiment que la distinction entre mots homonymiques et polysémiques a peu d'importance puisque certains liens historiques des mots polysémiques n'existent plus aujourd'hui.

Comme nous l'avons vu dans les résultats, le modèle reflète la variété des sens subtils du mot ainsi que son usage. Une première interprétation de cette observation est assez directe. Si un mot s'emploie afin d'exprimer l'attitude d'un locuteur telle que le caractère péjoratif, l'admiration, etc., il y a de fortes chances que ce mot accompagne d'autres mots qui sont employés pour le même effet. Par exemple, il est compréhensible et attendu que le mot *blunder* s'accompagne de mots

dépréciatifs tels que *stupid*. De la même manière, les contexonymes d'un mot qui s'emploient dans une situation spécifique peuvent refléter cette situation. Par exemple, les contexonymes {bell, prepared, bottle, breakfast, dinner, coffee, wine, supper, tea } de *order* mettent en évidence l'usage de *order* pour commander un repas, du café, etc. Ce type d'information est utile surtout pour une comparaison entre langues : par exemple, le mot équivalent français pour cet usage de *order* est *commander* et non son faux ami *ordre*.

Un des avantages de ce modèle est qu'il peut produire et proposer des contexonymes qui reflètent un usage ou une connotation difficiles à capter de manière explicite. En effet, une connaissance implicite est observée dans de nombreuses expériences en psychologie cognitive par exemple dans des expériences qui utilisent les effets d'amorçage. L'explication de Chomsky de la règle de la prononciation en anglais⁸ en sont un autre exemple : la plupart des anglophones prononcent correctement les mots en question sans connaître explicitement la règle. Ce type d'information sera utile pour des lexicographes ou des modèles comme celui de Edmonds et Hirst qui ont besoin du sens très précis des mots, comme par exemple la différence entre *forêt* et *bois*. L'approche descendante rencontre des limites théoriques concernant l'accès à cette information, car cette information est fréquemment implicite et non explicite.

7.4 Modèle informatique

Examinons maintenant les aspects informatiques du modèle à propos des quatre critères proposés dans le chapitre quatre. D'abord, en ce qui concerne la référence

⁸Par exemple, Chomsky explique qu'il existe trois règles concernant le son /i/ : a. une vélaire devient continue devant une nasale finale, b. voyelle + vélaire continue donne une voyelle tendue, c. /Ī/ devient [ay]. Ces règles qui expliquent les différentes prononciations d'une paire de mots comme *divine-divinity*, *ignite-ignition*, etc.

lexicale, ce modèle peut servir de référence disponible de manière pratique. Ce modèle propose un dictionnaire de contexonymes sur la page web⁹ pour plus de 100 000 mots en anglais et en français. Cette référence représente un intérêt pour un public qui va des gens qui apprennent une langue jusqu'aux traducteurs professionnels. La capacité du modèle à fournir des représentations selon les ressources utilisées permet au modèle d'être utilisé dans une étude comparative des sens des mots dans différents domaines. Comme le fait remarquer Habert (à paraître) dans son travail qui montre un changement du sens selon des ressources qui diffèrent d'une période à l'autre, ce type de modèle automatique permet de suivre le changement du sens des mots. Par exemple, les contexonymes de *vache* renvoient à la maladie de « vache folle » qui a été un sujet très vif dans les médias (Ji et Ploux 2003) :

vache : folle, maladie, ESB, bovine, spongiforme, encéphalopathie, agriculture, lait, agent, viande, bovins, farines, animaux, épidémie, transmission, humaine, britannique, britanniques, alimentaire, Creutzfeldt-Jakob, animal, Grande-Bretagne, boeuf, éleveurs, consommateurs, sang, animales, dossier, risques, fièvre¹⁰

Dans un sens, ces contexonymes mettent en évidence une relation de collocation ou de lien « pragmatique » d'après la terminologie de Lyons (1977).

Ce modèle peut aussi être utilisé comme dictionnaire ou système de représentation de connaissances. Par exemple, pour un voyageur voulant savoir sur quelle partie du corps s'applique une crème marquée « conditionner » disponible dans un hôtel, il sera possible de trouver la réponse dans les résultats du modèle : le modèle propose *hair, shampoo, intensive* comme contexonymes suggérant que le « conditionner » est un crème pour les cheveux et non pour le visage. C'est un exemple

⁹<http://dico.isc.cnrs.fr/dico/context/search>

¹⁰ $\alpha = \beta = \gamma = 0,01$ pour le corpus *Le Monde*.

pour l'apprentissage automatique aussi, car le modèle crée un lien entre *conditionner* et *hair* à partir d'un texte brut de traitement. De même, les résultats pour les quasi-synonymes, comme l'exemple cité précédemment, peuvent être interprétés comme des exemples pour l'apprentissage automatique. Ainsi, ce modèle montre qu'il est capable d'apprendre et d'organiser les contexonymes de manière acceptable.

Comme le constate Jackendoff (1990), la psycholinguistique ne s'intéresse qu'aux processus et la structure sémantique du lexique est supposée donnée au préalable. Notre modèle au contraire s'intéresse plus à la structure du lexique mental qu'aux processus. La question des procédures liées à la représentation sémantique d'un mot reste cependant ouverte : l'organisation graduelle d'un mot polysémique par exemple peut être comparée avec certains résultats de la psycholinguistique. Plus précisément, l'existence ou non du sens noyau d'un mot polysémique peut être comparée à l'existence ou non d'une représentation générique d'un sens dans la représentation de ce mot ; le temps de réaction dans l'effet d'amorçage d'un mot par un autre peut être comparé à la distance entre les contexonymes et le mot dans la représentation de ce mot.

La possibilité que le modèle puisse s'appliquer au TAL comme par exemple à la traduction automatique est aussi une perspective intéressante. Comme le montrent les exemples pour l'effet de fusion, plusieurs mots créent une représentation spécifique suggérant une réponse pour la désambiguïsation (voir chapitre 5). Il est vrai que la performance du modèle pour la désambiguïsation n'est pas aussi autonome que d'autres aspects du modèle comme la constitution d'une ressource ou la modélisation sémantique. Cependant la production explicite des contexonymes pourra servir de première étape pour différents modèles semi-automatiques qui nécessitent une intervention humaine.

7.5 Conclusion et perspective

Le modèle que nous avons développé dans cette étude présente un intérêt à la fois théorique et pratique pour la notion de concept et pour la représentation du sens des mots ainsi que pour le traitement automatique des langues (TAL) et la simulation psycholinguistique. Les résultats du modèle montrent une validité prometteuse pour chacun de ces aspects.

On peut toujours se poser la question de la nécessité pour un modèle du concept, du sens des mots ou du TAL de respecter ou de simuler la manière dont l'homme réalise les fonctions qui s'y rapportent. En général, il existe deux voies possibles : soit des modèles construits uniquement pour l'un des objectifs notamment : un modèle de simulation¹¹ ou un modèle de TAL, soit un modèle qui tend à répondre de manière synthétique à différents aspects. Il ne nous semble pas nécessaire de choisir l'une des deux voies au préalable. Un modèle mérite d'être considéré s'il fonctionne de manière productive sans simuler le fonctionnement ou la structure des humains. D'autre part, un modèle théorique ne doit pas être dévalorisé, comme le dit Jackendoff, par la simple absence d'un lien entre modèle théorique et modèle pratique. En ce sens, le simple fait qu'un modèle soit capable d'expliquer plusieurs choses de manière synthétique ne garantit pas en soi sa validité.

Prenant en compte ce fait, il sera profitable de résumer et de valoriser les choses que notre modèle a apportées.

A notre avis, un modèle qui simule de mieux en mieux les caractéristiques humaines a plus de chances d'avoir du succès et d'être utilisé. Cette anticipation

¹¹Il s'agit de modèles psycholinguistiques ou bien linguistiques. C'est-à-dire, de modèles qui simulent le processus psycholinguistique ou l'organisation du lexique mental d'une part et des modèles qui reflètent les connaissances théoriques établies au sein de la linguistique comme par exemple le modèle théorique de Jackendoff d'autre part.

provient de la nature du concept et du langage humain — comme le montre notre étude, l'approche définitionnelle ou logique échoue à expliquer plusieurs aspects à l'égard du concept.¹² Cette difficulté provient de la nature complexe et non logique du concept et du sens des mots. Par exemple, l'arbitraire de la répartition du sens des mots est tel qu'il est difficile d'établir de manière exhaustive l'organisation, du sens et des concepts. La nature arbitraire et non logique du langage humain, qui contraste avec le langage artificiel, du concept et du sens des mots devrait être prise comme une caractéristique essentielle et non comme le résultat d'une accumulation historique.

Nous avons montré par ailleurs que cette nature complexe et arbitraire des concepts et du sens des mots est difficile à capter, car d'une part il existe une difficulté pratique (par exemple le nombre des concepts est trop important et elle évolue rapidement) et d'autre part, ce type de connaissance est fréquemment implicite, ce qui ajoute un obstacle théorique.

Si l'organisation du concept et du sens des mots est construite autour de la vie d'un individu et de la vie sociale en général, les ressources textuelles sont sans conteste une ressource très limitée, en principe. Bien sûr, il est toujours possible d'intégrer des informations visuelles ou sensorielles pour compléter le modèle. Mais nous avons renoncé à développer un modèle théorique multi-sensoriel, car il existe déjà un nombre important de modèles qui tentent d'intégrer ce type d'information. Puisque notre objectif dans cette étude n'était pas que montrer un modèle possible de manière théorique mais un modèle capable de manifester une capacité effective, nous pensons qu'un modèle d'intégration qui n'est pas démontrable nuirait à notre objectif initial.

¹²L'approche par des sous-ensembles flous qui sont une sorte d'extension de la logique, ne peut pas surmonter tout à fait ce problème, car cette approche est basée aussi sur la logique définitionnelle.

En effet, il serait tout à fait possible d'intégrer des informations multi-sensorielles dans notre modèle s'il existait une méthode adéquate. Nous pensons notamment que des informations prosodiques seraient très utiles afin de discerner ou d'intensifier les relations entre mots. Par exemple, les énonciations adressées à un enfant et à un public adulte ne sont pas les mêmes ainsi distinguent-elles deux types de contexte différents.¹³

Nous utilisons le terme *refléter* pour la signification de la représentation du modèle. En effet, la représentation comporte elle-même des limites pour désigner tous les sens différents des mots. Par exemple, le sens de *mettre* n'est pas identique selon qu'on considère les cas suivants : *mettre un gant*, *mettre un chapeau* et *mettre un pantalon*. Malgré cette différence, les contexonymes de *gant*, *chapeau* et *pantalon* ne peuvent pas fournir la moindre information qui puisse aider à deviner les différents sens de *mettre*, car la différence des sens n'est pas marquée de façon linguistique en français ainsi qu'en anglais alors qu'elle existe dans le concept implicite construit à partir des informations moto-sensorielles. Ce n'est pas un problème négligeable, car dans certaines langues comme le coréen trois verbes différents s'appliquent très rigoureusement pour *mettre*, et pour la traduction, cette information devient essentielle. De ce fait, pour distinguer les sens de *mettre*, il peut y avoir deux solutions possibles. D'une part, comme l'a suggéré Dagan (Dagan et Itai 1994), d'autres langues qui distinguent les sens subtils peuvent être utilisées afin de classer les sens. D'autre part, des informations multi-moto-sensorielles pourraient être utilisées comme ressources.

Pour conclure, notre modèle montre qu'il peut servir de cadre pour la simulation psycholinguistique et de modèle linguistico-informatique ; il propose aussi

¹³Dans certains cas, l'intégration des informations prosodiques ne restera pas comme un ajout supplémentaire. Par exemple, dans une langue comme le coréen, le vocatif est marqué par une postposition, alors qu'en anglais ou en français, le vocatif ne s'identifie que par la prosodie.

une réponse concernant la discussion théorique sur le concept et le sens des mots.

Annexe A

Résultats du test sur des mots choisis de manière aléatoire

TAB. A.1: Résultats du test sur des mots choisis de manière aléatoire dans le corpus GP. Les contexonymes obtenus sont regroupés automatiquement par ACOM.

alliance	{ entered, formed, powerful, england, france, union, french, states, secret, alliance, peace, emperor, concluded, friendship, prince, austria, duke, holy, marriage } { proposed, prussia, treaty, political, powers, spain, germany, russia } { defensive } { offensive } {triple}
bits	{ pick, picked, bits, walls, dried, pieces, wood, dust, broken, stones } { dirty, stuff, straw, bread, furniture, color, cloth, silk, torn, china, glass, metal, bacon, choice, cut, paper } { gravel, scraps, fragments, scattered, shells, tiny, bones, tore } { stuck, stick, sticks, bark, sorts, string, ends, odd } {tit}

blasted	{ prospects, blighted, ruined, blasted, tree, rock, withered, heath, barren, crops } { damned, fame, hell, destroyed, hopes, fool, curse, shame } { lightning, gnarled, trunk, oak, pine }
brains	{ blow, knock, skull, beat, brains, inside, instead, mad, blood, heads, calf, cat, dash, em, ideas, minds, dull, hearts, clever, looks, working, bodies, fool, puzzled, brain, vain } { pistol, knocked, blown, blew, dashed, blowing, pink } { rack, racked } { racking }
canterbury	{ stephen, henry, york, england, pope, chancellor, london, thomas, william, library, robert, canterbury, university } { consecrated, christ, clergy, bells, cathedral, bishops, archbishop, ecclesiastical, cardinal, winchester, bishop, rome, abbot, church } { edition, pilgrims, poems, poets } { chaucer, tales } { prologue } { zealand }
contradiction	{ statement, assertion, involved, principle, necessity, individual, argument, contradiction, theory, terms, social, principles, relation, conception, non, opposition } { contrary, causes, clearly, direct, facts, explain, apparent, proposition, doctrine, exist } { possibility, existence, spirit, ideas, manifest } { involves } { flat }
devise	{ wit, saving, rid, solution, plans, schemes, execute, devise, scheme, tax, whereby, device, ingenuity } { apparatus, plot, plan, method, methods, punishment, ways, contrive, escaping } { efficient, expedient, measures, recommend, congress, earnestly } { didst, wisdom }

-
- dipping { lifting, rising, basin, dipping, stream, dip, solution, valley, rocks, sides, surface, drying, horizon, waves, sail, sailing } { liquid, fingers, pot, cool, dipped, needle, spoon, handkerchief, bowl, pen, cup, finger, bread, ink } { forming, beds, angles, inclined, eastward, degrees, angle, steep } { sandstone, strata, stratified, conglomerate, cleavage, inwards } { gulls, paddle, sails, paddles, spray }
- dug { natives, holes, inches, dig, wells, built, ashes, bodies, garden, planted, dug, depth, mud, lake, soil, stones, graves, hole, buried, pit, sand, snow, deeper, roots, bones, clay, bottom, grave } { ditch, canal, potatoes, pits, root, digging, spade } { heels, knife } { nails } { trench, trenches }
- eaves { steep, overhanging, projecting, porch, thatch, roofs, wooden, beneath, houses, barn, eaves, wind, leaves, hung, rain, built, windows, roof, straw, walls } { sparrows, swallow, birds, nests, nest, swallows } { dripping, drip, dripped }
- end { length, middle, cut, line, north, south, miles, east, west, farther, road, led, passage, piece, street, hair, foot, forward } { hours, minutes, rope, view, window, opposite, wall, placed, standing, hall, walked, front, garden } { necessary, comes, italic, further, journey, lower, upper, act, beginning, begin, end } { century, nearly, third, latter, week, month, months }
- fact { proved, proof, living, simple, mere, merely, existence, whatever, real, aware, knowledge, further, attention, necessary, subject, view, fact, character, curious, evidence, remarkable, history, natural, act, cause, interest } { nearly, probably, account, greater, important, position, age, england, generally, states, due, latter } { remains } { spite }

grape	{ purple, vines, apples, wine, garden, growing, fruits, soil, sugar, wild, grape, trees, branches, leaf, vine, fruit, leaves, corn, sweet, seed, apple, drink } { juice, clusters, jelly, grapes, ripe } { arbor, planted } { varieties, climate, cultivation } { shot } { guns }
harmonious	{ features, color, taste, beauty, variety, proportions, individual, perfection, colours, graceful, song, perfect, harmonious, harmony, emotions, lines, music } { complete, powers, action, development, physical, relation, mental, relations, principle, elements, union } { sounds, tones, voices } { notes, discord, discordant }
inexplicable	{ instinct, vague, conscious, sudden, impulse, dread, terror, ordinary, altogether, circumstance, utterly, otherwise, wholly, unknown, mysterious, explain, inexplicable, extraordinary, curious, explanation, events, mystery, apparently, conduct } { incident, explained, facts, species, theory } { phenomena, phenomenon }
loyalty	{ union, virtue, respect, religion, political, loyalty, personal, affection, confidence, devotion, friendship, sympathy, courage, proof } { oath, faith, obedience, subjects, loyal, majesty } { interests, patriotism, sovereign, honor, duty, faithful, fidelity } { gratitude, sentiment }
occur	{ instances, plants, occasionally, changes, forms, species, important, occurred, parts, animals, period, occur, natural, generally, cases, conditions, similar, various } { opportunity, probably, naturally, frequently, circumstances, especially, chance, change, likely } { events } { variations } { rarely }

-
- perspiration { starting, streaming, wiped, clammy, wiping, hair, trembling, breath, thick, faces, poured, bathed, covered, pale, hot, streamed, skin, terror, burst, violent, rolled, cold, heat, broke, trembled, bath, fever, patient, dripping, damp, handkerchief, perspiration, drops, beads, shirt, cheeks, hat, dust, forehead, brow, wet } { trickled, wipe } { pore } { profuse }
- threat { threat, terrible, terror, danger, effect, reply, threatened, kill, refused, violence, attack, carrying, storm, awful, alarmed, move, empty, anger, warning, blow, repeated } { uttered, fulfilled, loud, rage, frightened, torture } { execution, security, conveyed, terms, carry, force, promise, idle, menace } { ning } { soviet }
- uneasiness { cause, uneasiness, pain, increased, degree, caused, account, anxiety, slight, situation, secret, satisfaction, passion, presence, causes, suspicion } { considerable, pleasure } { vague, lest, growing, gives, filled, conscious, expression, signs, occasion, showed } { occasioned }
-

Annexe B

Résultats du test sur les exemples d'Edmonds et Hirst

TAB. B.1: Résultats du test sur les exemples d'Edmonds et Hirst dans le corpus GP.

Mot	α	β	γ	Contexonymes
blunder	0,07	0,15	0,05	[commit, committed, mistake] [stupid]
	0,35	0,50	0,35	{blunder, mistake, corrected, unpardonable, commit, committed, grievous, fatal, frightful, mistakes, repair} {stupidity, gross, stupid} {joke, unlucky}
	0,50	0,50	0,50	{stupidity, commits, gross, stupid, blunders, corrected, detected, mistake, blunder, repair} {indiscretion, unpardonable, committing, commit, mistakes, grievous, committed, fatal} {calamity, frightful} {awkward, joke, unlucky} [howells]

error	0,07 0,15 0,05	{argument, opinions, belief, faith, contrary, knowledge, source, greater, taught, easily, imagine, liberty, due, divine, former, understanding, experience, regard, merely, appears, authority} {error, admit, opinion, convinced, correct} {discovered, ideas, political, principle, causes, doctrine, degree, mere, religion, science, fundamental, modern, discover, method} {false, judgment, evil, virtue, conduct, ignorance, judge, lead, wise, fallen, makes, avoid, fall} {prove, wrong, errors, liable} {vulgar, sin, ways, commit, trial, fault, guilty, committed, lies} {grave, mistaken, , corrected, mistake, supposing} {serious, fatal} [clerical] [pointed] [text]
lapse	0,07 0,15 0,05	[{lapse, changes, slow} {centuries, ages, century} {geological, strata}] [memory]
	0,10 0,25 0,10	{strata, geological, organic, centuries, ages, century, lapse, slow, changes, species, rate, period, progress} {forgotten, memory} {minutes, interval, weeks, absence, months} {moments, recall}
slip	0,07 0,15 0,05	{foot, narrow, slip, fingers, hole, corner, quietly, coat, watch} {advantage, allowed, chance, letting, opportunity, reach, easily, easy, managed, try, lest, fall, caused, escape} {written, handed, wrote, tongue, pocket, paper, inside, lines} {fast, noose, rope, boat, knot, neck}

mistake	0,05 0,10 0,05	{discovered, wrong, error, instead, fatal, committed, supposing, mistake, serious} {impossible, correct, makes} {thinking, meaning, perceived} {possibility}
	0,07 0,15 0,05	[{{instead, mistake, opinion, mere, possibility, seeing, imagine, likely, unless, particular, supposed, discovered, easily, makes, easy, convinced, impossible} {slight, due, greatest, regard, greater, result, correct, false, giving, caused, committed, serious} {supposing, surely, error, marriage, possibly, mistaken, wrong, anybody, perceived, thinking} {intended, trying, meaning, explain, meant} {fatal, sad, fallen, terrible}}] [sorry]
enjoined	0,10 0,25 0,10	{secrecy, silence, multitude, commanded, enjoined} {priests, penance, perform} {instructions, obedience} [strictly]
	0,50 0,50 0,50	{strictly, instructions, obedience, code, priestly, piety, enjoined, silence} {prohibited, positively, expressly, forbidden, obey, priests, penance, punishment, penalty, commanded, commands, executed, execute, perform, governors, whatsoever} {prudence, advised, abstain, multitude} {injunction, observance, strict, despatch, injunctions, caution, secrecy, strictest} {commended, earnestly, solemnly} [directors]

prescribed	0,07 0,15 0,05	{prescribed, regulations, rules, terms} {assembly, authority, duties, oath, thereof, provided, conditions, according, required, constitution, laws} {section} {mode, treatment} {patient, physician, medicine, medicines} {remedies, remedy} [limits]
<hr/>		
ordered	0,05 0,05 0,05	{officer, officers, orders, royal, report, prisoner, prisoners, duke, ships, charge, emperor, appointed, governor, majesty, carry, ship, palace, send, arrived, horses, accordingly, follow, placed, immediately, ordered} {lieutenant, command, guard, commanded, join, camp, army, troops, move, division, regiment, soldiers, battle, enemy, advance, attack, brigade, cavalry, march, colonel, corps} {informed, board, finding, council, remain, receive, post, refused} {thrown, carriage, twelve, proceed, instantly, drive, paid} {clothes, servant, clock, servants} {bell, prepared, bottle, breakfast, dinner} {coffee, wine, supper, tea}
<hr/>		
forest	0,03 0,04 0,03	{(wind, woods, green, fields, forest, covered, trees, depths, snow, blue, sky) (narrow, village, hill, stream, broad, below, wide) (mountains, hills, mountain, plain, valley, distance, mile) (field, grass, wood, beasts, dense, beneath, tree) (east, vast, shore, west, lake, north, distant, plains) (oak, grew, leaves, birds, flowers, forth, tall, branches, thick, pine, summer, golden, spring) (path, edge, foot)} {deer, hunting} {knight, castle, rode}

woods	0,03 0,04 0,03	{snow, pine, sky, mountain, hills, meadows, forest, trees, woods} {leaves, flowers, bird, birds, green, tree, grass, fields, wood, winter, blue, gray, yellow, spring, thick, golden, summer} {rivers, streams, lake, valley, distant, rocks, forests, places, mountains, waters} {mile, shore, hill, edge, stream} {path, walk, walking}
-------	----------------	---

Annexe C

Test sur les résultats de Hirsh et Tree

TAB. C.1: Les résultats du modèle ACOM sur les exemples de Hirsh et Tree. Les mots identiques entre les résultats des sujets humains et les résultats de l'ACOM sont marqués en gras. Les contexonymes d'ACOM sont ordonnés des sens plus génériques vers les sens plus spécifiques. Les éléments équivalents du point de vue de l'ordre sont regroupés sur une même ligne et séparés par des virgules.

Sujets âgés	Sujets jeunes	Résultats d'ACOM
ACROBAT		$\alpha=0,40$ $\beta=0,20$ $\gamma=0,20$
Circus 0,40	Circus 0,31	Rope
Trapeze 0,09	Trapeze 0,16	Circus
Tumble 0,07	Gymnast 0,07	Adobe, Performance
Athlete 0,04	Dancer 0,04	Figure, Clown
Clown 0,04	Flying 0,04	Professional, Huge, Somersaults, Stretched, Physical, Acrobats, Graceful
Exercise 0,04	Swinging 0,04	Caught, Leap, Sun, Swings, Acrobatic, Temple, Macintosh, Falling, Likely, Cut, System, France, Sailor, Flat, Artist, Author, Windows, Floor, Skill, Shoulders, Actress, Flying

Gymnast 0,04	Trampoline 0,04	Format, Easily, Users, Earth, Exist, Ed, Feats, Equally, Fall, File, Except, Flint, Dwarf, Conjuror, Document, Bar, Beds, Attention, Beer, Big, Bolted, Beside, Associative, Across, Actor, Yielded, Write, Application, Dried, Breathlessly, Window, Cracked, Crowd, Dancer, Somersault
Tumbler 0,04 Agility, Athletic, Daring, Daring, Jump, Jumping, Lithe, Performer, Somersault, Wire 0,02	Aerobics, Agile, Balance, Ball, Clown, Energetic, Gym, Jester, Juggler, Leotard, Plump, Swing, Tightrope 0,02	

ALLEY		$\alpha=0,05$ $\beta=0,05$ $\gamma=0,05$
Lane 0,18 Cat 0,13 Passage 0,11 Dark 0,09 Gulley 0,09 Street 0,09 Way 0,09 Bowling 0,07 Narrow 0,04 Skittles 0,04 Blind, Marbles 0,02	Cat 0,38 Way 0,20 Bowling 0,07 Dark 0,09 Road 0,04 Street 0,04 Cul-de-sac, Back, Bin, Damp, Home, Mobile, Lane, Side, Town 0,02	Street Narrow Blind Dark Corner Wall Led Across, Bowling Ran Houses Window, Garden Lane Walked Tin, Pan Shop Front Road Gate, Yard Church Dead Entrance City Close, Steps Turn, Green Entered, Walk, Streets, Leading, Ground, Stopped Opened, Walking Passing, Main, London, Lived Running, Low, Square, Opposite Turning, Toward, Stone

ANGEL

Heaven 0,31
 Cake 0,09
 God 0,09
 Cherub 0,04
 Church 0,04
 Gabriel 0,04
 Heavenly 0,04
 Sky 0,04
 Wings 0,04
 Christmas, Fiction, Good-
 ness, Guardian, Hotel, Je-
 sus, Knee, Nice, Peace,
 Seraph, Winged 0,02

Heaven 0,16
 Christmas 0,09
 Halo 0,09
 Wing 0,09
 Wings 0,09
 God 0,07
 Devil 0,04
 Dust 0,04
 Mercy 0,04
 Arch, Beautiful, Eyes,
 Fairy, Flying, Gabriel,
 Gambler, Guardian,
 Heaven, Light, Peace,
 Perfume, White 0,02

 $\alpha=0,03 \quad \beta=0,03 \quad \gamma=0,03$

Heaven
Guardian
 Earth
 Soul
 Bright
Wings
 Sweet
Gabriel
 Human
Devil

Angels
 Spirit, Pure
 Fair
 Holy
 Golden
 Beauty
 Blue
 Form
 Recording
Peace
 Jacob, Loved
 Music, Evil
 Art
 Comes
 Blessed
 Hair, **Heavenly**
 Happy
 Fallen
 Book
 Daughter, Mary, Gold,
 Dream
 Glory, Tears, Divine,
 Sword
 Sun
 Saint
 Rose
 Dark, Spoke, Joy
 Save, Smile, **Church**
 Story
 Moses, Picture
 Dead, Strong
 Blood, Figure
 Sight, Grace
 Ye
 Sister
 Lips, Standing, Close
 Ministering, Soft, **Mercy**
 Sleep, Christ
 Led

ANKLE

 $\alpha=0,10 \quad \beta=0,10 \quad \gamma=0,10$

Foot 0,40	Foot 0,20	Deep
Bone 0,20	Bracelet 0,18	Foot
Sock 0,13	Chain 0,09	Sprained
Leg 0,11	Bone 0,07	Leg
Sprain 0,04	Shoe 0,07	Knee
	Sock 0,07	Mud
	Sprain 0,07	Red
Joint, Knee, Shoe, Strap,	Broken, Fat, Football, In-	Pain
Thin 0,02	jury, Joint, Knee, Thigh,	
	Walk, Wrist 0,02	
		Legs
		Across
		Bed, Arm, Green
		Covered
		Bones
		Shoulder, Walk
		Horse, Wore, Bone
		Pretty, Hurt, Nearly, Bro-
		ken
		Brown, Blue, Stocking,
		Silk
		Floor, Ground, Swollen
		Fall, Stand, Cloth, Skin,
		Sand, Stone
		Standing, Laid, Crossed,
		Struck
		Straight, Dropped, Fin-
		gers, Iron, Sudden, Soft
		Tree, Step, Ring, Shoe
		Captain, Wrist
		Accident

ARROW		$\alpha=0,03$ $\beta=0,03$ $\gamma=0,03$
Bow 0,67	Bow 0,53	Bow
Darts 0,07	Head 0,13	Shot
Head 0,07	Dart 0,09	Straight
Shoot 0,04	Point 0,04	Heads
Fire, Flight, Heart, Mis-	Band, Board, Dagger,	Struck
sile, Quiver, Sharp, Target	Eye, Heart, Kill, Norman,	
0,02	Sling, Stick 0,02	
		Drew
		String
		Pierced
		Shaft
		Shoot
		Wound
		Direction
		Blue
		Hit
		Wounded
		Red
		Spear
		Flew
		Fly
		Stone
		Mark
		Arrows

Swift, Poisoned
Eye
 Cut
 Golden, **Flight**
 Shooting
 Keys, Using
 Sped, Blood, Flint, Breast
 Horse, Across, Arm
 Broad
 Shield, Gold
 Aim, **Quiver**
 Past
 Area, War, Boat

BALL

Bat 0,51
 Games 0,11
 Play 0,09
 Bounce 0,04
 Football 0,04
 Round 0,04

Chain, Cricket, Dance,
 Pen, Racket, Rounders,
 Throw 0,02

Bat 0,16
 Game 0,16
 Foot 0,11
 Play 0,07
 Football 0,07
 Children 0,04
 Control 0,04
 Soccer 0,04
 Tennis 0,04

Basket, Bounce, Boy,
 Cannon, Dress, Encom-
 passing, Football, Goal,
 Racket, Room, Stick,
 Street 0,02

$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$

Cannon
 Shot
Game
Play
Dress
 Struck
 Red
 Hit
Foot
Dance

Court
 Rolled
 Powder
 Ground
 Captain, Grand
 Playing
 Sun, Fired
 Big
 Earth
 Rolling
 Eye
 Dinner
 Dancing
 Gold
 Fancy
 Arm
 Ladies
 Played
 Close
 Base
 Rifle, Entered, Madame
 Ran
 Cut

BASKET

Shopping 0,47
 Fruit 0,11

Ball 0,24
 Case 0,22

$\alpha=0,03$ $\beta=0,03$ $\gamma=0,03$

Table
 Filled

Weave 0,11	Eggs 0,07	Carrying
Carrier 0,07	Weave 0,07	Flowers
Bread 0,04	Flowers 0,04	Paper
	Fruit 0,04	Arm
	Shopping 0,04	Waste
	Wicker 0,04	Fruit
Ball, Cane, Carry, Chair,	Apples, Bike, Bread,	Big
First, Hold-all, Wicker,	Box, Countryside, Han-	
Willow, Woven 0,02	ging, Mother, Nutcase,	
	Vegetables, Woven 0,02	
		Bread
		Clothes, Carry
		Eggs
		Fish
		Covered
		Placed
		Wicker
		Red
		Chair
		Beside
		Lunch
		Laid
		Empty
		Green
		Rose
		Market
		Sitting
		Opened
		Packed, Corner
		Bottle
		Heavy
		Pretty
		Front
		Floor, Provisions, Food
		Fresh
		Window
		Garden
		Book, Hung
		Ground, Carefully, Walked
		Picked, Putting
		Silver, Across
		Blue
		Hanging
		Sight, Piece, Bottom,
		Containing
		Contents, Wood, Road
		Threw, Close
		Hat

 BAT

Ball 0,56
 Cricket 0,13
 Belfry 0,07
 Cave 0,04

Ball 0,31
 Man 0,16
 Cricket 0,13
 Hell 0,09
 Cave 0,07
 Black 0,04
 Fly 0,04
 Wing(s) 0,04

 $\alpha=0,10 \quad \beta=0,10 \quad \gamma=0,10$

Larry
Wings
Ball
 Jimmie
 Dale
Wing
 Blind
 Red

Creature, Dull, Innings, Loft, Roof, Runs, Vam- pire, Wall, Wings 0,02	Cat, Cow, Fruit, Hit, Tower 0,02	Bird, Gray
		Owl Flying Seal, Big Flew, Cricket Dead, Ground, Fly Leg, Tree Horse, Toward Birds, Sanctuary Form, Across, Ye Bones, Hit Low, Game, Play, Shadow, Caught Baseball, Street, Molly, Brown, Natural, Similar, Flies, Catch, Below, Cave Cry, Clothes, Vampire Character, Hung, Corner, Forth, Common, Cat
<hr/>		
BATTERY		$\alpha=0,03$ $\beta=0,03$ $\gamma=0,03$
Torch 0,31 Car 0,13 Charger 0,09 Hen 0,07 Clock 0,04 Power 0,04 Toys 0,04	Power 0,13 Walkman 0,09 Chicken 0,07 Energy 0,07 Hen 0,07 Torch 0,07 Acid 0,04 Car 0,04 Electricity 0,04 Test(s) 0,04 <Pass>, Bicycle, Card, Charge, Clock, Farm, Flash, Flat, Low, Meter, Operated, Radio, Vibrator, Watch, Zinc 0,02	Guns Artillery Storage Electric Current Assault, Enemy Colonel Horse Field Line, Captain, Gun Opened
Acid, Cell, Change, Charge, Clockwork, Creamy, Dynamo, Flat, Juice, Light, Radio, Watch 0,02		Infantry Batteries, Charge Position Front Regiment, Twenty Cavalry, Heavy Six, Cannon, Force Fort Wire
<hr/>		
BEER		$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$
Drink 0,22 Ale 0,20 Pub 0,13 Glass 0,11 Drunk 0,09	Drink 0,24 Lager 0,11 Pint 0,11 Wine 0,11 Glass 0,07	Wine Drink Glass Drinking Bottle

Alcohol 0,04 Barrel 0,04 Bitter, Froth, Keg, Lager, Mat, Mug, Stout 0,02	Ale 0,04 Brown, Festival, Flat, Fun, Goggles, Hangover, Head, Jug, Mat, Monster, Mug, Pub, Refreshing, Yeast 0,02	Bread Pint Bottles Drank Table Ginger Ale Spirits Bar Drunk Per Public Glasses, German Strong Tobacco Six Food, Houses, Beef Cheese Whisky Cold, Tea Pub Street Alcohol Lager Real, English Brewed Cent Barley Meat Dinner, Brewing, Market
---	--	--

BIRD		$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$
Flying 0,20 Robin 0,18 Nest 0,11 Fly 0,09 Sparrow 0,07 Table 0,07 Cage 0,04 Blackbird, Cockerel, Eagle, Feathers, Hand, Seed, Sing, Sky, Song, Sore, Tweet 0,02	Fly 0,27 Feathers 0,11 Brain 0,09 Nest 0,04 Owl 0,04 Robin 0,04 Sky 0,04 Wings 0,04 Blue tit, Breasts, Chick, Chicken, Cricket, Drop- ping, Fish, Flight, Flying, Girl, Gull, Pigeon, Prey, Tree 0,02	Species Birds Nest Prey Wings Eye Animal Song Feathers Flight Rare, Fly Big, Ground Sergeant Wild Food Cage Fish

BLUE

Sky 0,40
 Colour 0,29
 Sea 0,09
 Mood 0,04
 Red 0,04

Depressed, Green, La-
 goon, Lamp, Miserable,
 Upset 0,02

Sky 0,20
 Red 0,11
 Green 0,07
 Cold 0,04
 Cross 0,04
 Moon 0,04
 Sad 0,04
 Sea 0,04
 White 0,04
 Bird, Bottle, Boy, Depres-
 sed, Grass, Hair, Ice, Kit-
 chen, Mood, Nile, Police,
 Sailor, Shirt, Train, True,
 Water 0,02

Tree

Beak, International, Bree-
 ding
 Sea
 Tail, Different, Male, Mi-
 nister, Red

$\alpha=0,01$ $\beta=0,01$ $\gamma=0,01$

Sky
Red
Green
 Dark
Hair
 Yellow
Sea
 Pale
 Bright
 Sun

Gold
 Clear
 Brown
 Coat
 Gray
 Trees
 Soft
 Rose
 Wore
 Silver
 Across
 Fair
 Mountains
 Golden
 Big
 Eye
 Pink
 Flowers
 Dress
 Hills
Colour
 Silk
 Hat
 Clouds
 Broad
 Smoke
 Purple
 Beyond, Grey
 Eyed
 Wide
 Tall
Grass
 Lips
 Pretty
 Fine
 Dressed

Color
River
Figure
Beneath
Stars, Thin
Covered
Low, Cloud
Wind
Line

BOY

$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$

Girl 0,60

Girl 0,82

School

Lad 0,13

Ball 0,04

Big

Child 0,04

Young 0,04

Age

Male 0,04

Brown

Son 0,04

Master

Brother, Energy, Grand-
son, Sex, Teenager, Young
0,02

Friend, Gender, Hood, Lit-
tle 0,02

Boys

Horse

Farmer

Hair

Ran

Care

Office

Ten

Bad

Red

Brother

Pretty

Story

Street, Dead

Uncle

Play

Bed

Fine, Strong

Family

Run

Twenty

Across

Bright

Sister

Six

Fellow

Lad

Strange, Blue

Youth

Loved

Happy, Caught

Show

Born

Turn

Sight

Book

John, Lived, Grew

Spoke

Wild, Dark

Captain, Beside

Gentleman, Close

Hold
Fair
Table
Standing
Dog, Eye, Mere
Met
Tried

BUBBLE

 $\alpha=0,10 \quad \beta=0,10 \quad \gamma=0,10$

Soap 0,36
Bath 0,31
Squeak 0,13
Foam 0,04
Gum 0,04

Bath 0,24
Burst 0,22
Squeak 0,11
Pop 0,09
Gum 0,07
Fish 0,04
Float 0,04

Blow, Burst, Liquid, Pipe,
Washing 0,02

Balloon, Blow, Blowing,
Camping, Cauldron, Neck-
lace, Trouble, Wheels
0,02

Burst
Soap
Madam
Sea
Glass
Blown
Shells
Sun

Surface, Earth, Deep
South, Snails, Forth
Bursting, Spirit

Blow
Clear, **Foam**
Pricked, Filled, Green,
Break, Wild
Wave, Dryad, **Float**

BUN

 $\alpha=0,20 \quad \beta=0,20 \quad \gamma=0,20$

Currant 0,24
Cake 0,16
Butter 0,09
Bread 0,07
Chelsea 0,07
Fight 0,07
Food 0,07
Hot-cross 0,04
Baker, Easter, Eat, Ice,
Jam, Oven, Sweet, Tea,
Toasted 0,02

Oven 0,29
Cake 0,11
Currant 0,09
Hair 0,07
Baker 0,04
Bakery 0,04
Bread 0,04
Hot 0,04
Bake, Bap, Breasts, Bur-
ger, Cream, Eat, Egg,
Fight, Food, Ice, Rack,
Roll 0,02

Hair
Hill
Shop
Cross, **Hot**
Eat
Penny, Wa
Cake
Ta
Ra, Ate, **Tea**

Buy, **Oven**
Bath, Big, Na, **Currant**
Piece, Cook, Glass
Rabbit, Eating, Walked,
Ha
Runner, Blue, Bit, Ca,
Chief, Mouth, Top, Tied,
Pow, Meat, Story, Dress,
Drawn
Pulled, Nape, London,
Hungry, Nice, **Baker**

Tom, Indian, Peter,
Cheese, **Chelsea, Burger,**
Butter

BUNCH

$\alpha=0,03$ $\beta=0,03$ $\gamma=0,03$

Flowers 0,49
Grapes 0,20
Bananas 0,11
Group 0,04
Lots 0,04
Bouquet 0,04
Crowd, Handful, Keys
0,02

Flowers 0,42
Banana(s) 0,16
Grapes 0,13
Hair 0,07
Fives 0,04
Group 0,04
Carrots, Crowd, Hold,
Lots, Punch, Rose 0,02

Flowers

Keys

Big

Roses

Grapes

Violets

Hair

Wild

Red

Blue, Grass

Green

Cut, Pocket

Table, Fresh

Leaves

Cattle

Pretty

Tied

Top

Huge

Hung

Gathered, Em, Hat, Key

Yellow

CARROT

$\alpha=0,10$ $\beta=0,10$ $\gamma=0,10$

Vegetable 0,38
Donkey 0,11
Rabbit 0,09
Cake 0,07
Eat 0,07
Lunch 0,04
Red 0,04
Stew 0,04
Dinner, Fly, Peas, Potato,
See, Soup, Stick 0,02

Orange 0,36
Vegetable 0,11
Cake 0,09
Stick 0,09
Rabbit 0,07
Donkey 0,04

<Pass>, Bribe, Bunny,
Crunch, Eat, Pea, Peel,
Raw, Snowman, Soup,
Vimes 0,02

Juice, Onion

Stick

Celery

Fruit

Salt

Butter, Salad

Cut

Green

Vegetables

Parsley, Stock

Fresh

Cabbage, Chopped, Piece

Oil

Cup, Turnip

Raw

Orange

Pepper

Ham, Grated, **Vegetable,**

Eat

Bacon, Mineral, Sauce

Apple, Ground, Lemon,
Potatoes, Page, Add, Pepper-
corns
Fry, Leaf, Wheat, Glasses,
Lettuce, **Potato**
Cake
Plus, Ingredients, Herbs,
Boil, Carrots, Beet, Pars-
nip
Beef, Top

CART

Horse 0,84
Ride 0,04
Farm, Milk, Push, Trap,
Wheel 0,02

Horse 0,53
Wheel(s) 0,27
Carry, Go, Motor, Ox,
Pull, Sledge, Trolley, Van,
Wagon 0,02

$\alpha=0, 10$ $\beta=0, 10$ $\gamma=0, 10$

Horse
Road
Track

Gabriel
Horses
Wheels
Across
Pony
Coal, **Farm**
Past, Yard
Mr
Load, Drove, Racing,
Apple
Driver, Village, Top, Don-
key, Hay, Street
Ran, Golf, Sitting, Woo-
den, Pulled, Miles
Seat, Drive, Station, Dark,
Covered, Corner
Helped, Journey, Outside,
Jumped, Standing, Mar-
ket, Ready, Beside, Inside,
Wagon
Tail, **Carry, Pull**
Reins, Caught, Fruit, Ca,
Arrived, London, Painted,
Sight, Afternoon, Couple,
Valley, Erm, Rough, Sun
Wheel

CAVE

Hole 0,20
Cavern 0,16
Man 0,16
Dark 0,07
Rock 0,07
Seaside 0,04

Dark 0,22
Man 0,22
Hole 0,07
Bat 0,04
Bear 0,04
Dweller 0,04
Stone 0,04

$\alpha=0, 03$ $\beta=0, 03$ $\gamma=0, 03$

Mouth
Rock
Entrance
Mountain
Dark
Stone
Deep

Darkness, Deep, Dweller, Dwelling, Grotto, Hide, People, Pothole, Prehisto- ric, Sea, Seaside, Under, Underground, Water 0,02	Beach, Cavern, Cold, Ex- plore, Hill, Monster, Nick, Open, Rescue, Robber, Rock, Seaside, Stalactites, Vagina 0,02	Earth Wall Wild Sea Bones, Bear Adam Sun, Dead Floor, Led, Cliff, Entered Beneath Lived Forth Rocks Close, Top Across Red Hill Beyond Ran, Hole Narrow, Eve Front Discovered River Ground, Secret, Animals, Big, Human, Trees Cut, Hidden, Land, Ope- ning Mammoth Roof, Wide, Laid, Dark- ness
---	--	---

CHAIR

Seat 0,29
Sit 0,18
Rest 0,11
Comfortable 0,09
Table 0,09
Arm 0,07
Sit 0,07
Lift 0,04
Easy, Rocking, Stool 0,02

Table 0,38
Sit 0,16
Seat 0,11
Rocking 0,07
Cushion 0,04
Legs 0,04

Desk, Electric, High, Lift,
Pad, Rest, Sofa, Stool,
Wood 0,02

 $\alpha=0,01 \quad \beta=0,01 \quad \gamma=0,01$

Arm
Table
Sitting
Seated
Easy
Sit
Rose
Beside
Sank

Leaned
Drew
Window
Placed
Bed, Leaning
Forward
Dropped
Rocking
Close
Floor
Low
Front

Seat
Pushed
Standing
Legs
Big
Across
Threw
Moved

CHEESE

Biscuit 0,16
Bread 0,13
Cheddar 0,11
Mouse 0,09
Crackers 0,07
Pickle 0,07
Sandwich 0,07
Food 0,04
Toast 0,04
Board, Butter, Cake,
Cow, Dairy, Dish, Knife,
Onions, Protein, Tasty
0,02

Wine 0,13
Biscuit(s) 0,07
Board 0,07
Bread 0,07
Cheddar 0,07
Grater 0,07
Mouse 0,07
Chalk 0,04

Answer, Bacon, Blue,
Crackers, Edam, End,
Food, Fridge, Knife, Meat,
Mouse, Music, Nine,
Plate, Port, Smell, Straws,
Tomato, Worm 0,02

$\alpha=0,03$ $\beta=0,03$ $\gamma=0,03$

Bread
Milk
Cream
Butter
Meat
Wine
Cottage
Eggs
Grated
Fat

Food
Fruit
Fish
Salad
Sandwich
Fresh
Eat
Tomato, Biscuits
Cheeses
Ham
Sauce, **Cheddar**
Coffee
French, Foods
Top, **Toast**
Lunch
Tea, Chopped
Spread, Sandwiches, Pepper
Low
Soft, Oz, Beef
Chicken, Onion, **Blue**
Eating
Tomatoes, Parmesan, Cut,
Meal
Served
Potatoes, Add
Dish

CIGARETTE

Smoke 0,49
Ash 0,11

Smoke 0,41
Lighter 0,14

$\alpha=0,03$ $\beta=0,03$ $\gamma=0,03$

Lit
Smoke

Lighter 0,11
Smoking 0,04
Cancer, Card, Cigar, Dirty,
Ends, Fag, Holder, Illness,
Lungs, Match, No 0,02

Cigar 0,11
Butt 0,07
Addiction, Ash, Ashtray,
Cough, Desire, Die, Fag,
Light, Matches, Pint,
Smelly, Tobacco 0,02

Smoking
Lighted
Smoked

Lighting
Mouth
Lips
Match
Ash
Table
Tobacco
Box
Chair, Fingers, Threw
Rolled
Pocket
Paper
Drew
Silver
Blue
Gold
Sitting
Across, Coffee
Dropped
Leaned, Watched
Window
Rose, Slowly
Floor, Corner
Red
Cigar, Lighter
Offered
Finished, Beside
Rolling
Lights
Walked

CIRCLE

Round 0,38
Ring 0,29
Theatre 0,11
Clown 0,07
Ball, Hoop, Magic, Roun-
dabout, Seance, Square,
Stones 0,02

Square 0,56
Round 0,24
Friends 0,04
Big, Centre, Entrap, Light,
Red, Ring, Shape 0,02

$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$

Family
Centre
Outside
Wide, Formed
Inner

Beyond
Form
Line
Sun
Close
Social
Narrow
Ground, Drawn
Earth
Red
Vicious
Stone
Blue

Dark
Magic
Square
 Across
 Group
 Drew
 Standing, Society
 Members
 Number, City
 Surrounded
 Outer, Faces
 Table
 Twenty
 Turn
 Ladies
 Arctic, **Ring**
 Complete, Eye, Miles
 Diameter, North
 Human
 Party
 Middle, Slowly
 Figure
 Inside
 Ten
 Placed
 Different
 Sea, Gathered
 Rose, Moving
 Domestic, Midst, South
 Cut, Trees, Described,
 Move
 Forward, Six
 Distance
 Charmed
 Front, Top
 Semi, Green, Dead,
Stones
 Stand, Land
 Space, Sitting, Perfect
 Person
 Broken, Draw, **Big**
 Bright, Member
 Edge, Acquaintance
 Position, Company
 Camp, Low, Court
 Forming, Closed, Ac-
 quaintances, Wall
 Foot

CIRCUS

Clown 0,51
 Lions 0,09
 Animals 0,07
 Horse 0,04
 Ring 0,04
 Tent 0,04

Clown 0,42
 Acrobat 0,13
 Act 0,09
 Tent 0,07
 Performance 0,04

$\alpha=0,05$ $\beta=0,05$ $\gamma=0,05$

Boys
 Show
 Piccadilly
 Big
 Phil
Horse

Acrobat, Act, Entertainment, Negative, Ringmaster, Round, Square, Top, Trapeze 0,02

Animal, Babe, Big top, Clothes, Fair, Freak, Lion, Master, Nose, Ring, Trampoline 0,02

Tent

**Ring
Clown**

Street

Animals

Ha

Teddy

Lot, Flying

Horses, Rider

Performance

London

Top

Act

Across

Maximus, Sparling, Oxford, Train

Travelling

Animal

City, Run, Parade, Road

Theatre, Business, Car,

Roman

Idea, Family, Tom, Real

CLOAK

Dagger 0,31

Coat 0,20

Cape 0,16

Room 0,13

Garment 0,04

Black, Clothing, Cover, Price, Wear, Winter, Wrap 0,02

Dagger 0,44

Room 0,11

Coat 0,09

Witch 0,09

Black 0,07

Hook 0,04

Bag, Dark, Dracula, Evil, Gown, Hood, Magic 0,02

$\alpha=0,03$ $\beta=0,03$ $\gamma=0,03$

Hat

Wrapped

Shoulders

Red

Wore

Dark

Threw

Hood

Hair

Blue

Sword

Dress, Drew

Figure

Gold

Arm, Beneath

Fur

Thrown, Bonnet

Covered

Silk

Dressed

Cap

Velvet

Scarlet

Heavy

Coat

Laid

Wearing

Grey, Rose, Cold

Chair

Across
Shoulder
Gray
Tall
Wear
Flung, Brown
Fine, Hung
Wind, Broad
Throwing, Person, Folds
Close
Standing
Gentleman, Aside, Purple
Neck, Thick, Lined, Muffled, Caught, Drawn
Horse

COUCH

Settee 0,36
Rest 0,16
Sit 0,09
Sofa 0,09
Cushions 0,04
Potato 0,04
Sleep 0,04
Chair, Doctor, Furniture,
Lie, Lizard, Relax, Seat,
Smooch 0,02

Potato 0,44
Sofa 0,13
Chair 0,04
Seat 0,04
Settee 0,04

Cold, Comfort, Comfy,
Counseling, Kneel, Laze,
Lie, Old, Psychiatrist,
Sit, Sleep, Television,
Vegetate 0,02

$\alpha=0,03$ $\beta=0,03$ $\gamma=0,03$

Bed
Lying
Beside
Sleep
Laid
Sitting
Rose
Table

Stretched
Covered
Lie
Threw, Soft
Chair
Floor
Low
Window
Seated
Foot
Chamber
Placed, **Sit**
Corner
Across
Sank
Slept
Close
Flung
Arm, Spread
Wall
Deep, **Cushions**
Beneath
Hair

CROSS

Angry 0,20

Angry 0,22

$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$

River

Church 0,20
 Word 0,09
 Religion 0,07
 Road 0,07

Anger, Annoyed, Chain,
 Christ, Crucifixion, Crucifix,
 Jesus, Malta, Patch, Patch,
 Robe, Sign, Stitch, Suffering,
 Swords, Tick 0,02

Road 0,13
 Church 0,07
 Jesus 0,07
 Bun 0,04
 Nought 0,04
 Religion 0,04
 Stitch 0,04
 Tick 0,04
 Word(s) 0,04
 Bad, Cemetery, Christianity,
 Country, Crucifix, Double,
 Happy, Mother, Red, Square,
 Vampire 0,02

Red
Road
Sign
 Roads
 Self
 Plants
 Street
 Holy
 Examination
 Bridge

Sea
 Charing
Christ
 Flowers
Church
 Southern
 Miles
 Legged
 Fertilised
 Fertilisation
 Foot
 Land, Line
 Ground
 North
 Bear
 Across
 Pass
 Gold
 Turn
 Front
 Horse
 Path, Hall
 Cut
 Stone, Form
 Heavy
 Able
 Blue
 Crossed, Examined
 Number, Stream
 Different
 Bow
 War
 Blood
 City, Reach
 Fresh
 Dead
 Laid
 Top
 Army, Piece
 Dark, Deep, Hill
 Wood, Close
 Follow
 South
 Boat
 Species, Beyond
 Sight

Mountains, Trees, English,
 Breast
 West, Green
 Meet
 Opposite, Raised
 Horses
 Run
 Distance, Straight
 Silver, Main
 Lines, Ten
 Big, Wide
 Bearing, Golden, Walk
 Rose, Attempt, Led, Wild
 Threshold
 East
 Plant
 Earth, Twenty
 Table, Bar
 Nearly
 Below, Hold
 Village, Stand, Low, Cap-
 tain
 Figure, Strong, Enemy,
 Purposes
 Show, Length
 Tree, Past
 Six, Narrow
 Fall, Streets, Sun
 Fine, Arm, Valley, Marked
 Bed, Sword
 Pollen, Hung, Broken,
 Wall
 Party
 Corner
 Toward, Sitting

CUP

Saucer 0,62
 Tea 0,29
 Coffee 0,04
 Drink 0,04

Saucer 0,51
 Tea 0,22
 Drink 0,09
 Mug 0,09
 Coffee, Size, Spoon, Water
 0,02

$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$

Tea
Coffee
 Wine
Drink
 Teaspoon

Drank
 Table
 Tablespoons
 Fresh
 Filled
 Lips
 Chicken
 Ground
 Taste
 Drinking
 Bread
 Silver
 Salt
 Hot
 Pepper

Poured
 Minced
 Cups
 Butter
 Gold
 Red
 Cut
 Golden
 Drop
 Tablespoon, Oil
 Bitter
 Chopped
 Sugar
 Sweet, Cold
 Tin, Milk
 Fill, Drained
 Breakfast
 Breast
 Handed
 Empty
 Glass
 Plate
 Bed, Drunk
 Placed
 Rose
 Sliced
 Egg
 Ounces
 Green

DAISY

Chain 0,53
 Flower 0,38
 Marguerites 0,04
 Lawn, Weed 0,02

Chain 0,49
 Flower 0,31
 Cow 0,11
 Buttercup, Fields, Hazy,
 Yellow 0,02

$\alpha=0,10$ $\beta=0,10$ $\gamma=0,10$

Patty
 Nan
 Aunt
 Miller
 Flowers
 Demi, Pretty
 Blue
 Red, Nat
 Dora
 Green, Rose
 Winterbourne, Fresh
 Bunting
 Perdita
 Boys, **Flower**
 Sister, Drew, Play
 Sweet, Mutlar
 Sun, **Chain**
Yellow
 Joe
 Bill, Lupin, Hair
 Jo, Soft
 Bit, Dinner, Bess, Spring,
 Dow
 Girls, Garden
 Bright, Tears, Big, Fred,
 Stopped, Alice, Across

Chainsaw, Gay, Wild,
 Nearly, Maud
 Chandler, Summer, Table,
 Beneath, Tiny, Petals, Mrs,
 Won't, Bell, Ricky, Gol-
 den, Leaves
 Idea, Walk, Car, Driving,
 Quiet, Meg, Deep, Tree,
 Ethel, Farnsworth
 School

DOG

$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$

Pets 0,18
 Animal 0,13
 Cat 0,11
 Bark 0,07
 Bone 0,07
 Lead 0,07
 Friend 0,04
 Kennel 0,04
 Lead 0,04
 Bad, Bat, Bone, Cur,
 Fight, Meat, Mongrel,
 Poodle, Retriever, Walk,
 Woof 0,02

Cat 0,64
 Pet 0,11
 Walk 0,04
 Woof 0,04

Bone, Boy, Fleas, Hair,
 Shit, Stroke, Style 0,02

Master
 Big
Cat
 Horse
 Dogs
 Wild
 Tail
 Red
 Wolf
 Ran

Sheep
Animal
 Barking
 Watch
Bark
 Bull
 Cart
 Dead
 Run
 Mad
 Yellow
 Teeth, Legs
 Nose
 Close
 Human
 Faithful, Ground
 Ears
 Rose
 Eye
 Running
 Barked
 Bit
 Killed
 Sound, Across
 Show
 Mouth
Hair
 Strange, Turn
Bad
 Animals, **Fight**
 Road
 Front
 Low

Beside, Bite, Kill
 Fellow
 Toward
 Caught, Foot
 Sea
 Blood
 Walked
 Tree
 Fine
 Gray
 Hold, Comes
 Bed
 Street
 River
Meat
 Bear, Care, Except
 Lying
 Die
 Food, Hang
 Deep
 Shot, Sitting, Eat
 Bird, Brown, Fox, Collar
 Ye, Forward
 Corner, Tried
 Sprang
 Stand
 Tired
 Led, Cold, Dark
 Started, Piece, Lap
 Creature, Neither
 Laid
 Cut

DRAGON

Fire 0,24
 Wales 0,18
 Fly 0,13
 Welsh 0,13
 George 0,07
 Fierce 0,04
 Red 0,04
 Animal, Blocked, China,
 Fictitious, Legend, Roar,
 Wife 0,02

Fire 0,29
 Wales 0,11
 Slayer 0,07
 Fly 0,07
 Cave 0,04
 Knight 0,04
 Welsh 0,04
 Breath, Damsel, Dinosaur,
 Emblem, Exotic, Flames,
 George, Green, Lance,
 Monster, Mother, Myth,
 Red, Slay, Witch 0,02

$\alpha=0,03$ $\beta=0,03$ $\gamma=0,03$

Fly
Green
George
 Teeth
Red
 Golden
 Sea
 Flies, Blue

Tail, **Knight**
 Gold
 Earth
 Mouth
 Sun
 Serpent
 Ship
 Wings
 Blood, Heads
Monster
 Fiery
 Sword, Story

Fight
Dragons, Mr
Cave
Princess

DRAIN

Pipe 0,20
Sewer 0,18
Water 0,18
Sink 0,13
Smell 0,09
Blocked 0,04
Sieve 0,04
Away, Electric, Plug,
Pong, Storm, Strain 0,02

Pipe 0,45
Water 0,14
Block 0,05
Gutter 0,15
Sink 0,05

Blockage, Dirty, Filth,
Hair, Hole, Plug, Plunger,
Road, Smell, Strain, Tired
0,02

$\alpha=0,05$ $\beta=0,05$ $\gamma=0,05$

Cup
River
Resources
Blood
Brain
Paper
Pipe
Land

Dry
Minutes, System
Wine
Cut, Deep, Turn, Main,
Kitchen
Surface
Cold
Glass, Heavy
Ground, Add
Pipes, Gold, Joy
South, Bank
Clear, Butter, Waters
Bottom, Dregs, Hot
Outside, Rivers, Level
Sea, Remove, Lips
Drink, Dish, Waste
Across, City
Run, Allow, Comfort, Bit-
ter
Supply, War, **Road, Hole**
Sauce, Cook, Pan, East

DRILL

Army 0,16
Bore 0,16
Exercise 0,09
Hole 0,09
Bit 0,07
Soldier 0,07
March 0,04
Sergeant 0,04

Bit 0,27
Hole 0,07
Noise 0,07
Screw 0,07
Oil 0,04

$\alpha=0,05$ $\beta=0,05$ $\gamma=0,05$

Holes
Military
Hole
Sergeant
Bit
Company, Officers
Electric
Army

Carpentry, Discipline,
Electric, Hand, Marching,
Mend, Noise, Parade,
Power, Seed, Wood, Work
0,02

Army, Bore, Bosch, Den-
tist, Electric, Engine,
Hammer, Head, Hurt,
Jump, Kill, Machine,
Master, Nut, Piece, Ser-
geant, Seed, Tool, Up,
Wall, Work, Workman
0,02

Discipline

Drilling
Wall
Bits
Soldiers
Using
Machine, Soldier
War
Drills, Practice
Hammer
Training
Master, Exercise
Turn, System
School, Squad
Week, Command

DUCK

Pond 0,51
Drake 0,20
Quack 0,04

Quack 0,16
Bird 0,09
Pond 0,07
Swan 0,07
Swim 0,07
A-l'orange 0,04
Dive 0,04
Drake 0,04
Egg 0,04
Toilet 0,04
Water 0,04

Animal, Cider, Dodge,
Eat, Fowl, Geese, Hen,
Orange, Swimming,
Water, Wet 0,02

Bath, Beak, Billed, Bread,
Cricket, Dog, Face, Fea-
thers, Fly, Fuck, Golden,
Goose, Wader 0,02

$\alpha=0,05$ $\beta=0,05$ $\gamma=0,05$

Drake
Bill
Quack
Tufted
Donald
Mallard
Wild, Stephen
Eye
Dead
Dog
Species, Eider
Green, Ducks

Lame, Birds, Yellow
Sauce, Forceps, P., Neck
Including
Flight, Teal, Tail, Specu-
lum, Gybe
Low, Push, Chicken,
Crate, Breast, Roast,
Break, Eggs
Sitting, Fish, Grey, Onto,
Big, Century

EAGLE

Bird 0,44
Soar 0,09
Prey 0,07

Bird 0,27
Eye 0,20
Fly 0,11

$\alpha=0,04$ $\beta=0,04$ $\gamma=0,04$

Owl
Golden
Star

Bald 0,04	Bald 0,04	Mr
Eye 0,04	Sky 0,04	Species, Spotted
Fly 0,04	Star 0,04	European
Flying 0,04		Assemblage, Nest, Wings
Nest 0,04		Owls
Beak, Beautiful, Golden,	America, Bat, Bird, Dares,	Tawny
Hunter, Ledge, Mountain,	Egg, Flight, Hawk, Lan-	
Powerful, Wings 0,02	ding, Owl, Predator, Sky,	
	Talons, Wing 0,02	
		Company
		Creggan
		Prey
		Red, Fox
		Trust, Group, Verreaux,
		Sea, Barn, Insurance
		Cage, Digestion
		Mill, Eye
		Direct, American, Birds
		Imperial, Medium, Birdies
		Short-eared, Bird
		Centre, Talons
		Long-eared

FINGER

Hand 0,31	Thumb 0,31	$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$
Digit 0,20	Nail(s) 0,13	Ring
Nail 0,07	Point 0,09	Pointed
Point 0,07	Fish 0,07	Tips
Ring 0,07	Hand(s) 0,07	Pointing
Thumb 0,07	Food 0,04	Thumb
Puppet 0,04		Lips
Ring 0,04		Laid
Cake, Cut, Fish, Handle,	Biscuit, Gold, Joint, Li-	Nails
Index, Paint, Sore, Stall,	cking, Long, Steal, Toe,	Gold
Tip, Toe 0,02	Up yours, Woman 0,02	
		Touched
		Drew
		Touch
		Mouth
		Placed
		Eye
		Cut
		Raised
		Arm
		Hair
		Tip
		Lift
		Shook
		Middle
		Fingers
		Nose
		Holding
		Ran
		Red, Hold
		Lifted
		Table
		Third
		Blood

Across
Close
Nail
Pressed
Rings, Trigger
Laying
Slowly
Book

FISH

 $\alpha=0,01 \quad \beta=0,01 \quad \gamma=0,01$

Pond 0,22
Chips 0,20
Sea 0,11
Cod 0,04
Food 0,04
Gold 0,04
Meal 0,04
Tank 0,04

Chips 0,14
Swim 0,11
Finger(s) 0,07
Tank 0,07
Gold 0,05
Pond 0,05
Rod 0,05
Salmon 0,05
Scale(s) 0,05
Water 0,05

Cake, Catch, Fresh, Plaice,
Point, River, Rod, Salmon,
Steak, Tails, Water 0,02

Bowl, Bread, Bubble,
Cake, Cat, Film, Fishing,
Fresh, Marmalade, Net,
Sea, Smell, Trout, Waiter,
Wanda 0,02

Tank
Species
Food
Chips
Sea
Meat
Fishing
Caught
Big
Fresh
Pond

Eat
River
Catch
Feed
Birds
Aquarium
Shop
Different
Salmon
Feeding
Oil
Live
Size, Waters
Swim
Eggs
Surface
Marine
Chip

FLASH

 $\alpha=0,02 \quad \beta=0,02 \quad \gamma=0,02$

Lightning 0,29
Light 0,29
Torch 0,11
Camera 0,09
Bang 0,07
Flood 0,04

Light 0,22
Camera 0,18
Lightning 0,13
Gordon 0,09
Dance 0,04

Lightning
Sudden
Quick
Dark
Eye
Across

Apparition, Bulb, Cleaner, Explosion, Photography 0,02

Bang, Bulb, Car, Card, Coat, Flood, Git, Pan, Quick, Smile, Spark, Storm, Torch, White, Willy 0,02

Blue

Thunder
Red
Instant
Caught
Bright
Shot
Blinding, Darkness
Struck
Swift
Sun
Smile
Teeth
Past, Showed, Glance
Sound
Memory
Sky
Lips
Clear

FORMAL

Dress 0,20
Polite 0,13
Correct 0,07
Reception 0,07
Smart 0,07
Stiff 0,07
Suit 0,07
Informal 0,04
Occasion 0,04

Address, Attire, Casual, Chemical, Distant, Invitation, Just so, One, Proper, Reserved, Staid 0,02

Informal 0,13
Dress 0,11
Smart 0,09
Wear 0,09
Casual 0,07
Suit 0,07
Declaration 0,04
Dinner 0,04
Posh 0,04
Occasion 0,04
Tuxedo 0,04
Wedding 0,04

Clothes, Evening, Letter, Party, Respectable, Stiff, Straight, Strict 0,02

$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$

Informal
Education
Language
Training
Social
System, **Party**
Form
Public, Law
Different
Structure
Rules, Political
Process, Meeting
School

Relations
Act
Terms, Council
English, Authority, Written
Information
Local
Agreement
Organisation
Group
Control
Business
Legal, Court
Sense, Subject, Family
Merely
Groups
Number

Level
 National, Garden
 Committee, Individual,
 Recognition
 Example, War, States,
 Practice, Staff
 Members
 Teaching
 Development
 Policy
 Society
 Parties
 Application, Early
 Procedures, Position, Care
 Action, Mr
Dinner
 Stage, Study, Report
 Procedure
 Erm
 Statement, Speech, Expe-
 rience, Basis, Visit, Sys-
 tems
 Cold, Approach, Commu-
 nity, Usually
 President
 Knowledge, Writing
 British, Company, Particu-
 lar, Treaty
 Evidence, Chapter, Clear,
 United, Assessment

GAS

Fire 0,18
 Cooker 0,16
 Heat 0,11
 Stove 0,11
 Light 0,07
 Cook 0,04
 Mask 0,04
 Pipe 0,04
 Bang, Cork, Danger, Elec-
 tric, Employment, Explo-
 sion, Fuel, Heater, Jet, Poi-
 son, Power 0,02

Fire 0,16
 Cooker 0,11
 Bill 0,07
 Electric 0,07
 Mask 0,07
 Petrol 0,07
 Leak 0,04
 Oven 0,04
 Bag, Burner, Canister,
 Cooking, Die, Electricity,
 Fart, Flame, Fumes, Hea-
 ting, Light, Mains, Man,
 Meter, Smell, Stove, Water
 0,02

$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$

Lighted
 Balloon
Jet
 Lamps
 Lit
 Lamp
 German
Bag
 Coal

Street
 Acid
 Company
 Hydrogen
 Jets
 Mustard
 Big
Electric
 Dark
 Lighting
 Shell

Lights, Oil
 Carbonic, Front
 Burning
 Cloud
Heat
 City
Stove

GRAND

 $\alpha=0,02 \quad \beta=0,02 \quad \gamma=0,02$

Hotel 0,24
 Piano 0,13
 Great 0,07
 Staircase 0,07

Piano 0,22
 Stand 0,09
 Large 0,07
 Posh 0,07
 National 0,04
 Slam 0,04
 Thousand 0,04

Duke
 Master
Daughter
 Fine
 Court
 Prince
 Army
 River

Affair, Best, Big, Canyon,
 Child, Children, Circle,
 Daughter, Fare, Fire,
 Good, Large, Mall, Mar-
 vellous, Money, Mother,
 National, Palatial, Parade,
 Posh, Regal, Royal 0,02

Big, Canal, Canyon, Char-
 ming, Day, Gesture, Hall,
 Hotel, Money, Opening,
 Prize, Racing, Refinery,
 Scale, Small, Smart, Sport,
 Style, Total 0,02

Le
Stand
 Jury
 City
 Noble
Canyon
Hotel
 Opera
 Palace
 Emperor
 France
Hall
 Duchess
Piano
 War
 Dinner
 Music, French
 Family
 Public
 Council, Chief
Royal
 Party, Twenty
 Idea
 Church
 La, Queen
 Sight
 Gentleman, English
 Paris, Brother
 Gold, **Big**
 Ladies
 Lake
 Louis
 Six, Street
 Show
 Madame, Red, Marshal

Sea
 Central
 Majesty
Style
 Company
 Beauty, Scene
 Ball
 Mountains
 Table, Land, Wild, Deep
 Magnificent, Human,
 Across
 Green, Rich
 Line, Entered, Trees
 Past, Soul, **Scale**
 Arrived
Canal
 Beyond, Person
 Number
 Age, Front
 Occasion, Eye
 Common, **Staircase**
 Sun, Vizier, Honor
 Horse, Officers, Art, Seven
 Ten, Count
 Spirit, Object
 North
 Form, United, Close, His-
 tory
 Met, Law
 Splendid, Road, Uncle,
 Presence, Ground
 Entrance, Led, Visit, Inter-
 est, Business
 Mountain, South, Knights,
 Stone, Simple
 Except
 March, Monsieur, Dark
 Ancient, Steps, Broad,
 Standing, Captain
 Able, John, Strong, Prin-
 cess, Natural
 Character, Meet

GRILL

Bacon 0,20
 Toast 0,18
 Cook 0,09
 Food 0,09
 Fry 0,09
 Steak 0,09
 Bacon 0,04
 Steak 0,04

 Burn, Fish, Meal, Metal,
 Oven, Plate, Roast, Stove
 0,02

Bacon 0,16
 Pan 0,13
 Steak 0,11
 Bbq 0,09
 Fry 0,07
 Meat 0,07
 Sausages 0,04
 Burger 0,04
 Cook 0,04
 Food 0,04
 Burn, Char, Charcoal,
 Chef, Chips, Cooker,
 Metal, Oven, Stove 0,02

 $\alpha=0,10 \quad \beta=0,10 \quad \gamma=0,10$

Minutes
 Hot
 Cooked
Oven
 Inches
 Coals
 Medium
 Turning
Pan
 Top
 Bar

Brown

			Heat Microwave Cook Cooking Turn Fat, Meat Bread Dish, Toast Bake, Table, Fish, Bacon Food Iron, Slices, Lunch, Fry
<hr/>			
HAMSTER			$\alpha=0, 20$ $\beta=0, 20$ $\gamma=0, 20$
Pet 0,24	Cage 0,18		Dna
Animal 0,18	Gerbil 0,16		Cell, Eggs, Human, Cells
Cage 0,18	Mice/Mouse 0,09		Cage
Guinea pig 0,07	Pet 0,07		Rh, Pancreatic, Sfi, Cat,
			Lanes, Line, Live
Mouse 0,07	Wheel 0,07		Kb, Total, Ip, Mice,
			Mouse, Food, Pet,
			School, Rat
Rodent 0,07	Animal 0,04		Gel, Cheek, Wheel
Rat 0,04	Rat 0,04		
Wheel 0,04			
Brown, Cheeks, Rabbit,	Balcony, Brown, Cheeks,		
Small, Straw 0,02	Death, Fluffy, Food,		
	Fur, Guinea pig, Hairy,		
	Lettuce, Night, Rab-		
	bit, Rodent, Rummage,		
	School, Teeth 0,02		
<hr/>			
JELLY			$\alpha=0, 10$ $\beta=0, 10$ $\gamma=0, 10$
Blancmange 0,27	Ice-cream 0,31		Fish
Custard 0,16	Baby 0,16		Currant
Wobbly 0,16	Fish 0,13		Chicken
Cream 0,07	Party(ies) 0,09		Sea
Mould 0,07	Belly 0,07		Cake
Party 0,07	Wobble 0,07		Wine, Red
Baby 0,04	Blancmange 0,04		Glass
Desert 0,04			Dish, Cold
Sweet 0,04			Fruit
Birthday, Eat, Shake,	Bean, Dessert, Eel, Gela-		Bread
Trifle 0,02	tine, Plate, Spoon 0,02		
			Tea
			Big
			Green
			Roast
			Send, Foot
			Butter, Table, Apple
			Soup, Meat, Jam, Fishes
			Cup, Eggs, Cut, Form
			Pot, Boiled, Belly, Cream
			Hot, Clear, Yellow, Beef,
			Ate, Guava, Paper
			Soft, Sweet

KENNEL $\alpha=0,20$ $\beta=0,20$ $\gamma=0,20$
 Dog 0,98
 Maid 0,02

Dog 0,91
 Club, Cough, Doggy, Lead
 0,02

Dog
 Dogs

Club
 Hounds
 Master
 Yard
 Top
 Show, Street, Stable
 Hound, Outside
 Silver, Across
 Streets, Chained, Chain,
 Public, Staff, Horses
 Pack, Garden, Mr, Stables,
 Court
 Bark, Lived, Empty, Cor-
 ner, Blood
 Animals, Sound, Kitchen,
 Houses, Huntsman, Run
 Paper, Inside, Breeding,
 Ears, Owned, Party, Big,
 Boarding, Hunt, Husband,
 Front, Close, Lie, Nui-
 sance, Loose, Grand, Li-
 ving, Live, Won, Mas-
 tiff, Guard, Stand, Walk,
 Beside, Ground, Business,
 Ran, Bred, Road
 Knows, Window, Deep,
 Try, Floor, Beyond,
 Slowly, Manager, Free,
 Tree, Owners, Wolf,
 Slept, Wall, Mud, Mrs,
 Six, Forth, Waiting, Wal-
 ked, Noise, Built, Foot,
 Bed, Iron, Dark, Crept,
 Mistress, Rottweiler,
 Coach, Strong, Entrance,
 Pleasure, Hunting

KNIFE
 Fork 0,73
 Cut 0,18
 Sharp 0,05
 Carve, Stab 0,03

Fork 0,47
 Cut 0,07
 Kill 0,07
 Spoon 0,07
 Dagger 0,04
 Sharp 0,04
 Stab 0,04
 Blade, Cutlery, Eat, Edge,
 Pain, Pocket, Sandwich,
 Stanley, Sword 0,02

$\alpha=0,03$ $\beta=0,03$ $\gamma=0,03$

Cut
Sharp
Fork
Blade
 Using
Edge
 Mr
 Kitchen

Police
Throat
Carving
Attack
Table
Threatened, Palette
Wound, Stabbed
Plate, Top
Handle
Cutting
Across
Neck
Gun, Pulled, Armed,
Bread
Blood, Paper, **Pocket**
Cake, Carrying
Holding
Weapon

NECKLACE

Pearl 0,31
Jewellery 0,18
Beads 0,11
Neck 0,11
Diamond(s) 0,07
Bracelet 0,04

Adorn, Brooch, Chain,
Decoration, Lady, Pen-
dant, Present, Women
0,02

Jewellery 0,18
Bracelet 0,13
Chain 0,13
Gold 0,09
Pearl(s) 0,09
Neck 0,07
Bead(s) 0,04
Diamond, Gems, Golden,
Head, Headless, Jealous,
Pendant, Present, Ring,
Scarf, String, Wear 0,02

$\alpha=0,05$ $\beta=0,05$ $\gamma=0,05$

Gold
Pearl
Diamond
Pearls, Neck
Wore
Beads
Golden
Hair

Rings
Diamonds
Bracelets
Queen
Dress
Silver
Blue
Coral
Jewels
Red
Earrings
Throat
Green
Wear
Hung, Ear
Bracelet
Ring
Stones, Madame

NEEDLE

Thread 0,53
Sew 0,20
Sewing 0,16
Cotton 0,09
Pin 0,02

Thread 0,52
Pin 0,09
Haystack 0,07
Sew 0,07
Eye 0,05

$\alpha=0,05$ $\beta=0,05$ $\gamma=0,05$

Thread
Eye
Magnetic
North
Compass

Sewing, Cotton, Hypo-
dermic, Injection, Match,
Pain, Point, Prick, Stick
0,02

Sharp

Fine
Pole
Points
Sewing
Pointed
Fingers
Piece
Darning
Iron
Book
Red
Der, Table
Camel
South, Current, Silk, Rich
Die, Hollow, Knitting
Sea
Degrees, Finger, Working,
Cut
Wire, Stuck, Direction
Magnet, Drew, Ballade
Tailor, **Sew**
Arm, Blood
Pen, Pretty, Pricked, **Prick**
Box

ORGAN

Music 0,33
Church 0,24
Grinder 0,09
Heart 0,07
Play 0,07
Mouth 0,04
Piano 0,04
Body, Ear, Instrument,
Monkey, Musical 0,02

Heart 0,27
Music 0,16
Piano 0,13
Church 0,09
Donor 0,07
Grinder 0,07
Play 0,04
<Pass>, Big, Body, Do-
nation, Grandfather, Key-
board, Liver, Sexual 0,02

$\alpha=0,03$ $\beta=0,03$ $\gamma=0,03$

Church
Music
Played
Sound
Playing
Play
Barrel
Choir

Brain
Organs
Mouth
Piano
Grinder
Sense
Voices, Singing, Notes
Human
Instrument
Eye, Soul
Species
Street
Developed, Natural
Different
Function
Form
Cathedral, Service
Pipe, Solemn, Pipes

Parts
 Tones, Blood
 Party
 Cases, System
 Structure, Early
 Forth, Animal
 Loft, Common, Special
 Sweet
 Instruments, Chapel
 Mighty, Number
 Note
 Organist, Rose, Boys,
 Fine, Public
 Sing, Formed

PARCEL

Post 0,49
 Gift 0,11
 Present 0,11
 Birthday 0,04
 Package 0,04
 Packet 0,04
 Postman 0,04
 Bundle, Open, Shelf,
 String, Surprise 0,02

Post 0,13
 String 0,13
 Package 0,11
 Present 0,11
 Force 0,09
 Paper 0,04
 Baggage, Birthday, Box,
 Brown, Christmas, De-
 livery, Envelope, Open,
 Part, Pipe, Postman, Send,
 Surprise, Unknown, Van,
 Wrap, Wrapping 0,02

$\alpha=0,05$ $\beta=0,05$ $\gamma=0,05$

Paper
Brown
Post
 Land
 Opened
 Wrapped
 Pocket
 Containing

Carrying
 Tied
 Handed
Box
Send
Delivery
 Red
 Bag, Book, Arm, Contai-
 ned
 Office, Table
String
 Carefully
 Arrived, Letters
 Delivered
 Cut
 Gold, Laid, Shop, Blue
 Big
 Mean
 Mr

PARTY

Dress 0,11
 Enjoyment 0,09
 Fun 0,09
 Celebration 0,07
 Celebrate 0,04

Fun 0,16
 Animal 0,11
 Balloon(s) 0,07
 Birthday 0,07
 Drink 0,07

$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$

Dinner
Political
 War
 Republican
 River

Hat(s) 0,07
 Music 0,07
 Alcohol 0,04
 Games 0,04
 Acid, Celebrate, Dress,
 Girl, Group, Happy,
 House, Labour, Line,
 Loud, Pervert, Streamer,
 Table, Throw 0,02

Ball, Balloon, Bat, Birth-
 day, Birthday, Booze,
 Cake, Christmas, Conser-
 vative, Dance, Drink, En-
 joy, Family, Feast, Fete,
 Function, Gaiety, Games,
 Goer, Happiness, Happy,
 Invitations, Labour, Poli-
 tical, Poppers, Tea, Time
 0,02

Members
 Met
 Public
 Third
 Captain

Horses
 States
 Camp
 Strong
 Miles
Family
 Arrived
 Number
 Led
 Joined
 Democratic, Parties
 Court
 City
 Leader
 Neither
 Ladies
 Twenty
 United
 Chief
 National
 Ready
 Join
 Company, President
 Member
 Across
 English
 Person, Church
 Law
 Leaders
 Six
 Early, Late
 Natives, Spirit
 Gentlemen
Tea
 Leaving, **Table**
 Supper, Ten
 Meet, Close
 Interest
 Except
 Question
 Opposite
 French
 Immediately
 Free, Entered
 Road
 Horse

Forward, John
North
Hunting
Toward, Distance, Formed
Ground, Indians, Turn
None
Village, Soldiers, Support,
South
Cause, Brother
Gentleman, Broke
Peace, Occasion
Act, Form
Boat, Enemy
Force
Opposition, Union, Office
Nearly, Vote
Persons
Purpose
Account
Garden, **Line**
Prince
Started
Remained, Position
Different
Invited
Pass
American
Duke, Election
Arrival
Majority
Liberal
Journey
Land
Sight
Pleasure, Engaged
Determined
Ran, Expected, Service
Entire, Cut
Giving, Real
Royal, Feeling
Business
Governor, Officers
Conversation
Search
Character
Afterwards, Private, Pro-
ceeded
Hold
Attack, Front, Common
Influence
Week
Eight, Declared
Assembled, Waiting
Queen, Afternoon, For-
mer, Show
March
Carry
Beyond, Boys, Moved
Visit
Happened, Usual
Mr

Send
Sea, Killed, Opinion
Dead, Rode, Seeing
Secret, Command

PAVEMENT

Walk 0,47
Uneven 0,09
Curb 0,04
Pedestrian 0,04
Slab 0,04
Street 0,04
Street 0,04
Walkway 0,04
Artist, Broken, Kerb, Litter, Path, Road, Roman, Stone 0,02

Road 0,31
Walk 0,22
Stone(s) 0,11
Crack(s) 0,09
Sidewalk 0,04
Street 0,04

Concrete, Council, Curb, Fall, Footpath, Sick, Step, Wet 0,02

$\alpha=0,03$ $\beta=0,03$ $\gamma=0,03$

Street
Road
Across, Outside, **Stone**
Front
Window
Standing
Walked, Mosaic
Steps
Walking, Streets, Windows

Foot
Walls
Walk
Red
Houses
Marble
Shop
Stones
City
Below, Beneath
Cold, Wall
Car, Dark, Square
Opposite
Edge
Carriage
Hall
Past
Green
Ran
Rain, Slowly, Church
Corner
Stopped
Ground
Horses, Broad
Middle, Crowd
Close, Crossed
Beside, Lying, Narrow, Figure, **Step**
Dead, Court

PEANUT

Butter 0,60
Nut 0,09
Monkey 0,07
Shells 0,07
Eat 0,04
Salt 0,04

Butter 0,20
Allerg(ies) 0,07
Eat 0,07
Food 0,07
Monkey 0,07
Salt 0,07
Cake 0,04
Jam 0,04

$\alpha=0,20$ $\beta=0,20$ $\gamma=0,20$

Butter
Oil
Cup
Tablespoons
Teaspoon, Cut
Sauce, **Salt**
Chopped, Chicken, Pepper
Minced, Taste, Peanuts

Birds, Brittle, Chocolate, Dish 0,02	Oil 0,04	Roaster, Soy, Ground, Tablespoon, Vegetable, Shells Sliced, Stand
	Brittle, Cashew, Die, Dry, Elephant, Fattening, Joke, Kill, Pub, Sandwich, Satay, Shell, Snoopy, Walnut, Yogurt 0,02	Fresh, Food Jam Popcorn, Teaspoons, Peas, Jelly, Add, Brittle Fruit, Crowds, Olives

PHONE

Ring 0,29
Talk 0,11
Number 0,09
Call 0,07
Conversation 0,07
Speak 0,07
Box 0,04
Message 0,04
Answer, Bell, Bill, Book, Contact, Family, House, Ringing, Sister, Speaking 0,02

Ring 0,22
Call 0,20
Number 0,07
Speak 0,07
Bill 0,04
Talk 0,04

Answer, Answer machine, Bone, Book, Communicate, Computer, Contact, Conversation, Cord, Dog, Help, Home, Line, Phone, Me, Wire 0,02

 $\alpha=0,20 \quad \beta=0,20 \quad \gamma=0,20$

Calls
Number
Office, **Line**
Par
Telephone
System, Fax
Company, Phoenix
Usa, Address, Wayne
Jimmy, Send, Getting,
Calling, **Computer**

Conversation
McMahon, Lines, **Bell**,
Message, Ringing
Messages

PLANT

Flower 0,29
Garden 0,22
Pot 0,16
Bulbs 0,07
Daffodil, Dig, Geranium, Green, Greenery, House, Life, Machines, Place, Sew, Shrub, Violet 0,02

Green 0,29
Flower 0,11
Grow 0,09
Life 0,09
Tree 0,09
<Pass>, Bury, Bush, Cannabis, Extract, Food, Greenery, Indoors, Leaves, Nuclear, Pot Shrub, Soil, Spy, Weed 0,02

 $\alpha=0,01 \quad \beta=0,01 \quad \gamma=0,01$

Nuclear
Plants
Company
Production
Animal
Species

Growth
Leaves
Garden
Material
Per
Soil
Manufacturing
Workers
Million

Area
 Equipment
Grow
 Process
 Waste
 Local
 Example
Food
 Machinery
 Car
 British, Working
 Jobs
 Investment
 System
 Flowers
 Energy, Development,
 Cost
 Trees, Growing

PLATE

 $\alpha=0,02 \quad \beta=0,02 \quad \gamma=0,02$

Dinner 0,27
 Food 0,20
 China 0,04
 Cup 0,04
 Dish 0,04
 Knife 0,04
 Meal 0,04
 Rack 0,04
 Silver 0,04

Food 0,18
 Cup 0,09
 Dinner 0,09
 Bowl 0,07
 Dish 0,07
 Knife 0,07
 China 0,04
 Cutlery 0,04
 Fork 0,04
 Saucer 0,04
 <Pass>, Break, Chips, Dishwater, Eat, Put, River, Steel, Table, Tablecloth, Teeth, Washing 0,02

Bread, Communion, Crockery, Decorate, Glass, Sandwich, Serve, Supper, Table, Tea 0,02

Glass
Silver
Table
 Gold
 Brass
 Piece
Bread
Dinner
 Placed
Cup
 Iron

Windows
 Copper, Furniture
 Soup
 Window
 Laid
Knife
Fork
 Tin
 Front
Eat
 Surface
 Red
 Jewels, Figure
 Beside
 Cut, **Food**
 Pieces
 Breast, Pushed
 Hot
 Street, Wine, **Steel**
 Fine

POND

 $\alpha=0,03 \quad \beta=0,03 \quad \gamma=0,03$

Fish 0,27
Duck 0,22
Water 0,16
Frogs 0,09

Dirty, Golden, Lake,
Lilies, Meadow, Newt,
Ponds, Pool, Puddle,
Tiddler, Village, Weed
0,02

Duck(s) 0,27
Water 0,18
Fish 0,16
Life 0,09
Frog(s) 0,07
Lake 0,07
Weed 0,07
Bug, Goldfish, Stream,
Scum, Toad 0,02

Fish
Mill
Garden
Across
Green
Surface
Deep
Bottom

Trees
Road
Ice
Summer, **Lake**
River
Big
Duck
Stream, Ducks
Bank
Village
Clear
Mud, Woods, Filter
Middle, Shore
Stone
Plants
Ground
Paddy, Covered
Sun, Lily, Dam
Horse
Sea
Top, Frog
Hill
Dark
Area, Below

PROTECTION

Safety 0,13
Police 0,11
Care 0,09
Guard 0,07
Armor 0,04
Cover 0,04
Help 0,04
Insurance 0,04
Racket 0,04
Safe 0,04
Shield 0,04
Apron, Asbestos, Coat,
Covering, Husband, Lock,
Money, Mother, Police,
Safeguard, Scheme,
Security, Self 0,02

Condom(s) 0,18
Safety 0,13
Safe 0,09
Guard 0,07
Security 0,07
Helmet 0,04
Insurance 0,04
Racket 0,04
Safety 0,04

Comfort, Armed, Arrest,
Care, Coat, Cover, Father,
Gangster, Law, Police,
Rune, Secret, Strength
0,02

$\alpha=0,01$ $\beta=0,01$ $\gamma=0,01$

Act
Law
Environmental
Public
States
Rights
Society
United
Property
National
Court
Trade

Care
Laws
System
Provide

British, Special
 Afforded, Legal
 Council
 Consumer
 Data, Person
 Interests, **Safety**
Security
 Policy
 Force
 Social
 Copyright
 War
 Environment
 Free, Form
Police
 Health, Support, Employ-
 ment
 Interest
 Provided, Service
 Citizens
 Foreign, Land
 Family
 Local
 Subject

PUMA

Animal 0,31
 Cat 0,24
 Leopard 0,13
 Lion 0,09

Cat 0,38
 Leopard 0,13
 Lion 0,09
 Trainers 0,09
 Tiger 0,09
 Cheetah 0,04
 Sport(s) 0,04

Black, Dangerous, Feline,
 Forest, Jungle, Lynx, Run-
 ning, Sleek, Tiger, Zoo
 0,02

Adidas, Black, Car, Moun-
 tain, Panther, Tree 0,02

$\alpha=0,40$ $\beta=0,20$ $\gamma=0,20$

Cat
 Monkey
Lion
 Welwyn, Animals
 Big, **Panther**
 Jaguar, American, **Adidas**
 Watching, Tennis, Sunder-
 land, Spring, Final, Spe-
 cies, **Animal, Mountain**
 Helicopter, Super, States,
 Centre, Degrees, **Sports**

Deal, Killed, Famous,
 Cougar, Leather, Largest,
 South, Blood, Book, Prey,
 Area, Places, School,
 North, Challenge, Stag,
 Club, Carcass
 Games, Antelope, Yel-
 low, Athletics, Yesterday,
 America, Cordillera, Flew,
 Win, Cattle, Direction,
 Champions, Creature,
 Common, Company,
 Christie, **Tree**

RETURN

$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$

Back 0,29
 Home 0,20
 Ticket 0,11
 Send 0,07
 Away 0,04

<Pass>, Absence, Come,
 Distance, Give, Go,
 Going, Happy, Post, Reap-
 pear, Retreat, Visit, Yield
 0,02

Ticket 0,18
 Go 0,16
 Home 0,13
 Back 0,07
 Book 0,04
 Journey 0,04
 Leave 0,04
 Post 0,04
 Again, Boomerang, Call,
 Depart, Give, Library,
 Key, Loan, Receipt, Ro-
 ver, Send, Sender, Train
 0,02

Journey

Able
 Land
 Party
Visit
 London
 Immediately
 Family
 Expected

War
 City
 Paris, Early
 France
 Company
 Months
 Court
 Free, Husband
 Promised
 Pay
Send
 Service
 Former
 Business
 Receive
 Interest
 Form
 Rate
 Ready
 Brother, Per
 Ten
 Remain
 Desire
 Allowed
 Waiting
 Position
 Safe
 Leaving
 River
 Captain
 Meet, Wait
 Public
 Stay
 Week
 Decided
 Camp, Subject, Late
 Question, Longer
 Master
 Promise
 Account
 Number, Sea
 Obligated
 Six
 Person, Road
 Met
 Care
 Mr
 Peace

Native
 Paid
 Determined
 Agreed, Army
 Past
 Change
 Law
 Period, **Happy**
 Capital
 English
 Show
 South
 Wished, Prince
 Live
 Ship
 States
 Future
Absence
 Office
 Twenty, Thanks
 Horse
 Daughter, Offered
 Weeks, Queen
 Ordered
 North
 Force, **Post**
 Pass
 Seeing, Offer, Letters
 News, French
 Remained, John
 Miles
 Dead
 Church
 Turn
 Cent
 Complete, Forth
 Act
 Earth, Bed
 Duty
 Dinner, Gold

ROPE

Skip 0,18
 Ladder 0,14
 Skipping 0,11
 Cable 0,07
 Hang 0,07
 Hemp 0,05
 String 0,05
 Tie 0,05
 Ascend, Climb, Cord, Gal-
 lows, Knot, Line, Noose,
 Ship, Tied, Tight, Towing,
 Trick, Washing 0,02

Hang 0,16
 String 0,11
 Climb 0,09
 Ladder 0,09
 Pull 0,09
 Burn 0,04
 Noose 0,04
 Tie 0,04
 Bondage, Bridge, Clim-
 bing, Cord, Film, Hanging,
 Knit, Knot, Old, Pope,
 Sail, Stick, Swing, Twine,
 Wood 0,02

$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$

Tied
 Neck
Ladder
 Fastened
 Cut
 Boat
 Hold
 Across
 Fast

Bell
 Attached
 Tree

Pulled
 Ground
 Horse
Pull
 Top
 Coil
 Strong
 Caught
Tight
 Hung
 Piece
 Length
 Window
Line
Hanging
 Holding, Foot
 Weight
Ship
Tie
 Heavy
 Legs
 Ran
 Dragged
 Bound, Front

SAFETY

 $\alpha=0,02 \quad \beta=0,02 \quad \gamma=0,02$

First 0,36
 Careful 0,11
 Crossing 0,09
 Net 0,09
 Harness 0,07
 Belt 0,04
 Care 0,04
 Home 0,04
 Catch, Chain, House, Locked, Risk, Road, Security 0,02

Pin 0,20
 Net 0,16
 Harness 0,07
 Belt 0,04
 Security 0,04

Boring, Car, Care, Cushion, Danger, Dangerous, First, Gate, Guard, Harm, Health, Helmet, Home, Hospital, Lock, Match, Protection, Road, Rope, Seatbelt, Sure, Warm 0,02

Public
 Committee
 Peace
Danger
 Personal
 City
Care
 Ship
 Perfect

War
 Greater
 Escape
 Law
 Honour, Land
 Sought
 Sea, Army
 Chance
 Secure, Seek
Protection
 Party, Reach
 Enemy
 Valve, **Security**
 Anxiety, Common
 Force
 Regard
 Master

		Placed Free Arrived, Captain, Strength Cause, Ground, Sense Strong, Able, Across
--	--	--

SCREW		$\alpha=0,10$ $\beta=0,10$ $\gamma=0,10$
Driver 0,27 Nail 0,20 Nut 0,16 Screwdriver 0,13 Loose 0,04	Nail 0,24 Driver 0,11 Nut 0,09 Cork 0,07 Hammer 0,07 Shag 0,07 Screwdriver 0,04 Sex 0,04	Turn Propeller Loose Top Engine Machine Driver Iron Turning, Ship
Carpentry, Door, Fasten, Fix, Mend, Sex, Tighten, Twist, Wood 0,02	Ball, Bang, Bolt, Door, Hard, Hole, Jar, Loose, Loosen, Tighten, Tool, Up 0,02	Pitch Horse, Courage Wood Vessel, Big, Speed Plate, Form Wall, Steamer, Steam Thumb, Erm, Box, Cutting Piece, Holes Guns, Fitted, Position, Driven, Inch, Cork Stern, Twin, Glass, Board, Frame, Bottom Applied, Hole Number, Line, Steel, Fit, Mr, Hold, Inside, Try, Bot- tle Propelled, Fixed, File, Spring, Tried, Twenty Bow, Ships, Wheel, En- gines, Screwed, Plastic, Front, Floor, Nut, Ham- mer Angle, Surface, Type, Holding, Mouth, Motion, Handle, Nose, Metal, Ran Below, Bolt, Tool Machines, Machinery, Inches, Common, Heavy, Six, Edge, Size, Nail Driving, Forward, Connected, Parts

SILVER		$\alpha=0,01$ $\beta=0,01$ $\gamma=0,01$
Gold 0,33 Polish 0,07	Gold 0,51 Bullet 0,07	Gold Blue

Clean 0,04	Fish 0,04	Table
Coin 0,04	Plate 0,04	Golden
Cup 0,04	Ring 0,04	Moon
Metal 0,04		Hair
Precious 0,04		Green
Shine 0,04		Red
Spoon 0,04		Plate
Bars, Brooch, Cleaner,	Bracelet, Cloud, Cross,	Rich
Cleaning, Cutlery, Dish,	Hoard, Jewellery, Lining,	
Dollar, Money, Ornate,	Necklace, Platter, Rich,	
Ring, Shoes, Ware,	Salver, Service, Ware,	
Wedding 0,02	Watch 0,02	
		Value
		Dark
		Fine
		Sea
		Bright
		Sun
		Copper
		Coin
		Covered
		Rose
		Hung
		Precious
		Silk
		Quantity
		Trees
		Paper
		Glass
		Sky
		Shining
		River
		Pieces
		Box
		Across
		Piece
		Mines, Watch
		Iron
		Land
		Cup
		Cloth
		Velvet, Gray
		Wore

SINK		$\alpha=0,03$ $\beta=0,03$ $\gamma=0,03$
Dishes 0,22	Kitchen 0,13	Kitchen
Wash 0,16	Drain 0,11	Rise
Basin 0,09	Tap 0,11	Deep
Drain 0,09	Swim 0,09	Sea
Washing 0,09	Washing 0,09	Beneath
Drown 0,07	Plug 0,07	Earth
Swim 0,11	Wash 0,07	Bottom
Kitchen 0,04	Dish(es) 0,04	Below
Work 0,04	Float 0,04	Ground
	Plunger 0,04	Surface, Swim
Drop, Trough, Unit, Wash-	Beer, Blocked, Boat, Cold,	Sun
up 0,02	Down, Fall, Metal, Toilet,	
	Water 0,02	

Lower
 Ship
 Ready
 Low
 Weight
 Slowly
 Level
 Floor
Fall
 Deeper
 Soul
 Sight, **Boat**
 Die
 Sleep, **Cold**
 Table
 Land
 Chair
 Heavy, Bed, Ocean
 Big, Red, Turn
 River, Across, **Float, Washing**
 Hot
Wash
 Allowed
 Board
 Hold, Human
 Top

SKELETON

 $\alpha=0,05 \quad \beta=0,05 \quad \gamma=0,05$

Bones 0,71
 Body 0,11
 Key 0,04

Bone(s) 0,53
 Closet 0,09
 Skull 0,07
 Cupboard 0,04
 Doctor 0,04
 Skinny 0,04

Anatomy, Cupboard,
 Death, Framework, Grave,
 Human 0,02

Biology, Brain, Dead,
 Halloween, King, Muscle,
 Staff, White 0,02

Bones
Human
 Flesh
 Living
 Form
Skull
 Parts, **Dead**

Museum
 Bare, Animal, Skin
Key
 Thin, **Bone**
 Covered
 Mere
 Trees
 Remains
 Figure, Cases, Arguments,
 Gold, Ribs
 Lying, Reduced
 Bony, Discovered, Horse
 Foot
 Number, Dark, Keys
 Family
 Placed, Wild, Huge, Rose
 Nearly, Complete, **Closet**
 Showed, Animals

SKIRT

Blouse 0,49
Clothes 0,16
Long 0,07
Jumper 0,07

Female, Knee-length, Ladies, Length, Mini, Pleats, Short, Tight, Waist, Woman 0,02

Dress 0,20
Trouser(s) 0,18
Blouse 0,16
Short 0,09
Legs 0,07
Clothes 0,04
Annoying, Beauty, Flirt, Gap, Girl, Jeans, Jumper, Look, Mini, Shirt, Stockings, Woman 0,02

 $\alpha=0,03 \quad \beta=0,03 \quad \gamma=0,03$

Dress
Blue
Blouse
Waist
Silk
Hair
Coat

Wore
Jacket
Red
Hat
Brown
Wearing
Shoes
Dressed
Dark
Green
Wear
Riding
Top, **Legs**
Worn, Front
Bodice, **Shirt**
Figure
Pretty, Pink
Velvet, Shoulders
Grey, Knees
Cut, Neck
Lace

SLEDGE

Snow 0,80
Hammer 0,18
Toboggan 0,02

Hammer 0,53
Snow 0,38
Reindeer, Santa, Ride, Toboggan 0,02

 $\alpha=0,10 \quad \beta=0,10 \quad \gamma=0,10$

Snow
Hammer
Dogs

Journey
Horse
Wind
Runners, Heavy
Iron, Dog, Miles
Drove
Ice
Drawn, Big
Nikita, Winter, Six, Front
Whip
Wood, Party, Driver
Struck, Fur, Hammers,
Blows
Drive
Load, Cross
Distance
Sea, Magic, Frozen, Quickly, Wisting, Maiden

Philip, Started, Foot, Coat,
 Harness, Vasili, North
 Meter, Andreevich, Ready,
 Break, Driving, Horses,
 Blow
 Boat, Ran, Strong, Claus,
 Forward
 Journeys, Steed, Racer,
 Courser, Road, Broke,
 Flew, Placed, Getting,
 Forest
 Running, Ancient, Run,
 Provisions, Sledges, Cove-
 red, Across
 Team, Anvil, Standing,
 Third, Hero, Warm,
 Ground

SLY

Fox 0,20
 Cunning 0,09
 Dog 0,09
 Devious 0,07
 Sneaky 0,07
 Crafty 0,04
 Deceitful 0,04
 Evil 0,04
 Mean 0,04
 Nasty 0,04
 Shifty 0,04
 Artful, Cheat, Creepy, De-
 ceit, Dishonest, Person,
 Scheming, Snide, Spite,
 Untrustworthy 0,02

Fox 0,44
 Devious 0,07
 Crafty 0,04
 Cunning 0,04
 Evil 0,04
 Sneaky 0,04

Bad, Cruel Deceit, Deceit-
 ful, Horse, Mean, Scraper,
 Shy, Skunk, Smooth, Stal-
 lone, Stone, Wicked, Witty
 0,02

$\alpha=0,10$ $\beta=0,10$ $\gamma=0,10$

Smile
 Glance
 Eye
Fox
 Fellow
Cunning
 Expression, Glances, **Dog**
 Gentleman, Wink
 Looks
 Girls, Added
 Pretty, Grog, Laugh
Stone

Cat, Grin, Creature
 Master
 Boys, **Mean**
 Sally, Mouth, Lips, John,
 Mr, **Bad**
 Brother, Ye, Laughed,
 Forth
 Cast, **Person**
 Trick, Laughing, Close,
 Caught, Soft, Play, False,
 Sense, Watch, Red
 Ladies, Giving, Truth,
 Try, Hooja, Fun, Merry,
 Daughter, Nose, Met,
 Lucy, Story, Sweet,
 Business, Puss, Live
 Ways, Knowing, Art,
 Sharp, Makes, Ah, Meant,
 Real, Dark, Fine, Tone,
 Law, Mrs, Laid, Captain

Shop, Slow, Care, Sister, Wit, Chance, Court, Green, Madame, Bit, Ill, Rose, Treacherous, Quick Attention, Tricks, Window, Spoke, Played, Hold, Hint, Cut, Devil, Age, Show, Observed, Stole, Uncle, Free, Low, Spirit, Run, Especially Dirty, Self, Malicious, Meet, Quiet, Neither, Comes

 SPRING

Flowers 0,24
 Summer 0,16
 Season 0,11
 Jump 0,07
 Daffodils 0,04
 Leap 0,04
 Snowdrops 0,04

Roll 0,04

Chicken, Clean, Coil, Easter, Fever, Forward, Lambs, Love, Onions, Recoil, Warm, Water, Weather 0,02

Summer 0,24
 Water 0,09
 Flowers 0,07
 Jump 0,09
 Autumn 0,04
 Board 0,04
 Bounce 0,04
 Chicken 0,04
 Coil 0,04

Winter 0,04
 <Pass>, Cleaning, Lamb, Metal, Season, Seesaw, Time, Trampoline, Well 0,02

$\alpha=0,01$ $\beta=0,01$ $\gamma=0,01$

Winter
 Early
Summer
Flowers
 Green
Autumn
 Trees
 Earth
 Sun
 Sweet
 Fresh
 Fall, River

Cold
 Ground
 Leaves
 Tree
 Wind
Season
 Wild
 Grass
Warm
 Birds, Clear
 Forth, Late
 Bright
 Snow
 Blue
 Comes
 Song, Ready
 Garden
 Deep, Soft

 STAR

Sky 0,44
 Heaven 0,16
 Shine 0,09
 Moon 0,04

Sky 0,16
 Moon 0,11
 Shine 0,09
 Bright 0,07

$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$

Sun
Bright
Sky
 Stars, Earth

Bethlehem, Films, Freddy, Howl, Light, Night, North, Planet, Singer, Twinkle, Venus, View 0,02	Gaze 0,07 Twinkle 0,07 Heaven 0,04 Light 0,04 Points 0,04 Shooting 0,04 Sun 0,04 Chaser, Crib, Crossed, Ga- zer, Gazing, Night, Signs, Sparkle, Super, Wars 0,02	Heaven Moon Red Sea Dark Blue Shone, Shining North
		Rose Soul, Spangled Golden Eye Fair Rising, Silver Pale Clouds Beauty, Darkness, Clear Sweet, Heavens Across Pole Beyond Gold Forth, Banner Shine Wind Line, Millennial, Chamber Green, Born Breast Fixed Deep, East Shines Glory Space Single Lit, Land Spirit West

SUIT

Clothes 0,27
Man/Men 0,16
Dress 0,07
Wear 0,07
Jacket 0,07
Lounge 0,04
Shirt 0,04
Best 0,04
Bridge, Cards, Cloth, Clo-
thing, Dress, Dressy, Grey,
Hearts, Out, Outfitters,
Smart, Tidy, Tie 0,02

Tie 0,24
Smart 0,18
Man 0,09
Armor 0,07
Jacket 0,04

Case, Clothes, Coat,
Dinner, Hat, Hire, Inter-
view, Men, No, Office,
Pin-stripe, Sharp, Tailor,
Waistcoat, Wardrobe,
Wear, Work 0,02

 $\alpha=0,02 \quad \beta=0,02 \quad \gamma=0,02$

Clothes
Dressed
Blue
Hat
Dress
Wore
Dark
Law
Grey

Court
Wearing

Brown
 Hair
Shirt
 Follow
Wear
 Business, Red
Tie
 Different
 Cut
 Needs
 Shoes
 Silk
 Person
 Gray
Coat
 Worn
 Figure, Gold, Yourself
 Company
 Tweed
 Taste
 Purpose
 Ready
 Particular, Armour
 Green
 Change
 Pair
 Party
 Bought
 Gentleman
 Big, Clad
 Pay
 Fit
 Style, Boots
 Buy
 Collar
 Husband, City
Tailor
 Linen, Public
 Ten, **Cloth**
 Cap, Velvet
 Tall, Handsome
 Plain, Form
 Brother
 Changed, Front, Exactly,
 Top
 Summer, Likely
 Horse
 Street
 Service

 SWAN

Lake 0,49
 River 0,13
 Bird 0,11
 Signet 0,09
 Down 0,04

White 0,24
 Lake 0,16
 Bird 0,11
 Swim 0,09
 Goose/Geese 0,07
 Duck 0,04
 Elegant 0,04

 $\alpha=0,04 \quad \beta=0,04 \quad \gamma=0,04$

Lake
River
Song
Neck
 Hunter
 Wild
 Maiden

Black, Dive, Duck, Neck, Upping, Water 0,02	Neck 0,04 Song 0,04 Beautiful, Cold, Cy- gnet, Dive, Dying, Egret, Matches 0,02	Goose Silver Hotel John Shot, Feathers, Street, Birds David Gold, Wings Red Blue Form Story Major, Theatre, Ship, Creek, Hair Fair, Bird
<hr/>		
SWEDE		$\alpha=0,20 \quad \beta=0,20 \quad \gamma=0,20$
Vegetable 0,58 Carrot 0,04 Eat 0,04 Lunch 0,04 Mashed 0,04 Stew 0,04 Turnip 0,04	Turnip 0,22 Vegetable 0,16 Carrot 0,09 Orange 0,07 Parsnip 0,07 Potato 0,07 Blonde 0,04 Cabbage 0,04	Olaf Big German, Th Swedish Red Dutch, Ship Sailor, Cut, Captain, John, English, Twenty, Play Named, Ago, Top, Camp, Ships, Carl Malbihn, Dane
Blonde, Cabbage, Ho- nesty, Norway, Orange, Potato, Root 0,02	Cord, Dane, Eat, Ground, Jacket, Leather, Moun- tains, Norwegian, Radish, Sunday, Yellow 0,02	Onions, Race, Eat Broken, Anders, Bar, Blond, Birth, Beet
<hr/>		
TEA		$\alpha=0,02 \quad \beta=0,02 \quad \gamma=0,02$
Cup 0,20 Drink 0,18 Break 0,11 Cake 0,07 Coffee 0,07 Cup 0,04 Cuppa 0,04 Time 0,04 Bag, Caddy, Earl Grey, Meal, Milk, Party, Pot, Sandwiches, Strong, Sweet, Teacake 0,02	Coffee 0,38 Bag 0,09 Drink 0,09 Cup 0,07 Biscuit 0,04 Drink 0,04 Leaf 0,04 Sugar 0,04 Cakes, China, Pot, Relax, Rest, Room, Tetley, Toast, Urn 0,02	Cup Coffee Drink Cups Afternoon Table Milk Hot Sugar, Pot Breakfast Kitchen, Mrs Drinking, Erm Lunch

Nice, Mug, Biscuits

Party

Drank, Bread

Poured

Cake

Shop

Dinner

Eat, **Sandwiches**

Bit

Sit

Ready

Fruit, Tray

Strong

Bed

Toast

Cold

Early, Kettle, Served

Week

Finished

Bags

TOAD

$\alpha=0,10$ $\beta=0,10$ $\gamma=0,10$

Frog 0,40

Stool 0,09

Amphibian 0,09

Pond 0,07

Hole 0,04

Animal, Countryside,
Croak, Hall, Hop, Mouth,
Noisy, Pool, Reptile, Sly,
Swamp, Tadpole, Ugly,
Warts 0,02

Frog 0,51

Hole 0,11

Hall 0,09

Fish 0,04

Pond 0,04

Stool 0,04

Animal, Green, Leek,
Man, Princess, Slime,
Wart 0,02

Hall

Horned

Rat

Mole, Tree

Snake

Badger

Animal

Stone

Animals, **Hole**

Table, Big

Road, Across, Skin, Toads,

Horse

Dog, Species, Lizard, Red

Spadefoot, Dead, Jewel,

Eye, Past

Ugly, Frog, Mouth

Car, Eater, Egg

TOKEN

$\alpha=0,05$ $\beta=0,05$ $\gamma=0,05$

Gift 0,53

Book 0,11

Ticket 0,07

Label 0,04

Gesture 0,18

Gift 0,11

Money 0,09

Love 0,04

Record 0,04

Ring 0,04

Ring

Gratitude

Sign

Gold

Respect

Send

Chitty, Coin, Friendship,
Game, Meter, Present,
Prize, Respect, Spend,
Trust, Voucher 0,02

<Pass>, Arcade, Book,
Bus, Chip, Coin, Cou-
pon, Crust, Economy,
Exchange, Fair, Game,
Machine, Monopoly,
Offering, Reinforcement,
Representative, Reward,
Speech, Stamp, Ticket,
Voucher 0,02

Friendship

Accept
Ethernet, **Book**
Hair
Remembrance
Company
Network, Person
Ye, Show
Presented
Queen
Affection, Raised, Sup-
port, **Fair**
Kissed, Spoken, Earth,
Message
Red, Wear
Arm, **Gift**
Waved, Golden
Blood, Bear, Regard, Me-
mory
Giving, Master, Cut
Broken, Meaning, Finger,
Laid, Dead, Hath
Captain, Past, Service,
None, Public, Joy, Pre-
sence, Form
Kiss, Offer, Forth, Mark,
Happy, Lips, Deep
Receive, Heaven, Bowed,
Soul, Land, Human
Sun, Shook, Peace, Prince,
Bore, Sword, Ground
Truth, Visible, Piece,
Heads, Sea, Single, Met,
City, Free
Husband, Breast, Holy, Fa-
mily, System, Stand, Mine

TOOTH

Ache 0,27
Brush 0,25
Mouth 0,09
Molars 0,07
Dentist 0,07
Filling 0,05

Brush 0,24
Ache 0,13
Fairy 0,11
Nail 0,09
Mouth 0,07
Decay 0,04
Dentist 0,04

$\alpha=0,05$ $\beta=0,05$ $\gamma=0,05$

Nail
Eye
Teeth
Brush
Sweet
Front
Broken, Dog

False, Bite, Clean, Decay,
Loose, Paste, Sharp, Teeth,
Wisdom 0,02

Dagger, Eye, Hurt, Mug,
Paste, Pick, Rotten, Shark,
Sore, Teeth, White, Wis-
dom 0,02

Mouth

Big
Fight, Red, Gold
Jaw
Claw
Fine, Blue
Fought
Bone, Cut
Comb, Hair
Pain
Dentist, Decay
Ache
Horse, Brushes
Powder, Lower
Enamel
Blood, Beaver, Surface
Sugar, Serpent, Loss,
Sharp
Extracted

TRAFFIC

Lights 0,27
Jam 0,22
Car 0,20
Warden 0,07
Policeman 0,04

Chaos, Delay, Fumes,
Heavy, Hold-up, Hor-
rendous, Noise, Road,
Vehicles 0,02

Jam 0,38
Light(s) 0,20
Car(s) 0,11
Warden 0,09
Cone 0,04
Road 0,04
Accident, Cop, Frustra-
tion, Fumes, Noise, Stress
0,02

$\alpha=0,04$ $\beta=0,04$ $\gamma=0,04$

Street
Road
Streets
Trade
City
Public
Business

River
Car
Slave
Commerce
North
Roads
Lines, Line, States
Human
Main
Heavy
Across, Land
Commercial, West, Mer-
chants
York, Company
United, Foreign, Law,
Busy
Coast, Stream, Service,
London, Passing, Control
Liquor
Laws, Goods, Foot
Railway, American, Sea,
Ships
Opened, Sound, Slaves
Pass

Narrow, Passenger, Quiet

TRAP

Mouse 0,40
 Door 0,22
 Snare 0,18
 Net 0,04

Bought, Catch, Caught,
 Gin, Man, Pony, Rabbit
 0,02

Door 0,36
 Mice/Mouse 0,20
 Bear 0,04
 Rat 0,04
 Catch 0,04
 Cage, Capture, Caught,
 Cheese, Down, Fox, Ha-
 zard, Jump, Maze, Money,
 Pony, Scared, Snare, Stuck
 0,02

 $\alpha=0,05$ $\beta=0,05$ $\gamma=0,05$

Caught
 Laid
Mouse
Catch
Rat
 Rocks

Floor
 Fallen
 Opened
 Led
 Foot, Closed
 Fall
 Doors
 Line
 Steel
 Horse
 Ladder
 Shut, Spring
 Roof
 Walked
 Clap, Bait, Fish
 Escape, Opening
Fox
 Rock
 River
 Horses
 Cut, Close, Run
 Deep, **Bear**
 Beneath, Carefully, Ran
 Hole, Stone, Heavy, Hold,
 Across, Top, Passage
 Narrow, Entered, Slowly,
 Drove
 Traps, Iron, Getting, Se-
 cret
 Strong, Corner, Baited,
 Road
 Leading, Wall, Police,
 Covered, Free, Ready,
 Ground, Laying, Dark
 Fine, Setting, Sprung, Cei-
 ling, Game

TRENCH

Ditch 0,25
 Dig 0,20
 Coat 0,09
 Garden 0,09

Coat 0,33
 War 0,24
 Army 0,04
 Dig 0,04

 $\alpha=0,05$ $\beta=0,05$ $\gamma=0,05$

Deep
 Dug
 Earth
 Line

Pit 0,07
Furrow 0,05
War 0,05
Beans, Deep, Hole, Mud,
Pipes, Soak, Soldier, War-
fare, Work 0,02

Ditch 0,04
Foot 0,04
Warfare 0,04
Die, Dip, Dog, Lane, Mac-
kintosh, Muddy, Right,
Walk, Wet 0,02

Ground
Wall
Warfare
German, Across

Front
Inches
Top, **Dig, War**
Bottom
Laid, Digging
Wide, Cut
Yards, Covered, Field, Slit
Close
Enemy
Shallow, Shell, **Foot**

TUMMY

Ache 0,36
Stomach 0,24
Food 0,09
Belly 0,07
Fat 0,07
Bug 0,04
Trouble 0,04
Flat, Full, Pain, Tremble
0,02

Ache 0,29
Button 0,16
Stomach 0,16
Belly 0,07
Bug 0,07
Fat 0,04
Bellybutton, Big, Body,
Hungry, Illness, Navel,
Pain, Rumble, Tickle,
Upset 0,02

$\alpha=0,20$ $\beta=0,20$ $\gamma=0,20$

Relax
Muscles
Pull
Raise, Chest, Shoulders
Inch
Lying
Knees, Twist, Toning
Baby

Rock, Holding, Bottom,
Flattener, Cycle
Big
Pain, Ache
Fat
Spine, Exercise
Upset
Double, Eliminator, Pul-
ling, Stick
Floor, Tuck, Repeat, Exer-
cises
Comes, Upsets, Diet, Arm,
Ca, Legs, **Hungry**

VEIL

Wedding 0,40
Bride 0,13
Face 0,11
Cover 0,07
Hat 0,07
Nun 0,11
Covering 0,04
Eyes, Mask, Punch 0,02

Wedding 0,40
Bride 0,18
Cover 0,09
Marriage 0,07
Mask 0,04

$\alpha=0,02$ $\beta=0,02$ $\gamma=0,02$

Cover-up, Death, Deer,
Dress, Face, Funeral,
Hide, Threat, Secrecy,
Widow 0,02

Lifted
Dark
Thick
Hair
Blue
Dress
Hat
Covered

Sun
 Raised
 Aside, Thin
 Beneath, Drawn
 Hung
 Drew
 Draw
 Wore
 Green
 Soul, Fair
 Mist
 Sky, Pale, Lift, Hidden
 Threw
 Across
 Rose, Beauty
 Lace
 Cloud
 Wear

 WARDROBE

 $\alpha=0,05$ $\beta=0,05$ $\gamma=0,05$

Clothes 0,71
 Bedroom 0,11
 Cupboard 0,04
 Coat, Door, Dress, Full, Mistress, Suits 0,02
 Clothes 0,53
 Cupboard 0,13
 Door 0,07
 Close, Coat, Coat hanger, Drawers, Dress, Fire, Hammer, Ironing, Lion, Mirror, Shirts, Wooden 0,02

Clothes
 Bed
Dress
 Table

Opened
 Top
Bedroom
 Doors
 Wear
Drawers
 Master, Furniture
 Pair
 Dresses
Mirror
 Floor
 Wall, Front
 Corner
 Fitted, Hung
 Suit, Chair, Glass, Gold
 Clothing
 Big
 Blue, Articles, **Coat**

 WARRIOR

 $\alpha=0,03$ $\beta=0,03$ $\gamma=0,03$

Soldier 0,44
 Fighter 0,11
 War 0,11
 Brave 0,07
 Battle 0,04
 Fight 0,22
 Sword 0,09
 Battle 0,07
 Brave 0,07
 Princess 0,07
 Soldier 0,07
 War 0,07
 Fighter 0,04

War
Battle
Sword
 Chief
 Indian
Brave
 Red, Mighty
 Dead

Army, Chariot, Fighting,
Hero, Kill, King, Knight,
Spear, Tank, Tribe 0,02

Helmet 0,04
Chinese, Gladiator, Hero,
Hunter, Iron, King, Play,
Rogue, Shield, Slave,
Swordsman, Zulu 0,02

Fight
Spear

Blood
Bold
Noble
Knight
Spirit, Strong
Shield
City
Earth, Savage
Fierce
Fair, Close
Land, Sun, Horse, Single,
Field
Prince, Soul
Rose, Toward
Army, Fighting, Hunter
Beneath, Strength, Tall,
Wild, Drew
Slain, Eye, Sight, Strange
Age, Forward, Bear

WEDDING

Ring 0,20
Bells 0,18
Bride 0,11
Reception 0,09
Couple 0,07
Guests 0,07
Marriage 0,04

Anniversary, Celebration,
Church, Day, Dress, Fes-
tivity, Happy, Honeymoon,
Love, March, Pretty 0,02

Bell(s) 0,27
Cake 0,09
Dress 0,09
Funeral 0,09
Marriage 0,09
Ring 0,09
Church 0,07
Bride 0,04
White 0,04
Car, Celebrate, Divorce,
Joy, Present, Singer 0,02

$\alpha=0,03$ $\beta=0,03$ $\gamma=0,03$

Ring
Bride
Married
Dress
Church
Marriage
Feast
Party
Daughter
Family

Husband
Guests
Ceremony
Cake
Couple
Sister, Breakfast
Silver
Presents, Gold, **Anniver-
sary**
Bridegroom
Happy
Celebrated
Week
Brother, Princess
Prince
Journey
Clothes
Invited
Months

Fixed
 Queen
 Early
 Ready, Flowers, Gift
 Marry
 Dinner
 Wear
 Six, **Bells**
 Groom, Gown
Reception
 Quiet, John, Golden
 Parents, Royal
 Bed
 London
 Table, Weeks, Rings
 Finger
 Bridal
 Big
 Occasion
 Buy
 Pair
 Story, Ago
 Wore, Twenty
 Fine, Late, Grand, Attend,
 Send

 WINDOW

Pane 0,18
 Curtains 0,16
 View 0,16
 Glass 0,11
 Frame 0,09
 Sill 0,09
 Box 0,04
 Light 0,04
 Clean, Cleaning, Door,
 House, Look, Open 0,02

Glass 0,22
 Pane 0,11
 Frame 0,09
 Sill 0,09
 Door 0,07
 Light 0,07
 Look 0,04
 Open 0,04
 Break, Box, Cable,
 Church, Cleaner, Frame,
 Rain, Seat, See, Shopping,
 View, Watch 0,02

 $\alpha=0,01 \quad \beta=0,01 \quad \gamma=0,01$

Opened
 Street
 Sitting
Glass
 Across
 Standing
 Front
 Table
 Bed

Garden
 Chair
Sill
 Dark
 Outside
 Floor
 Rose
 Walked
 Sun
 Wall
 Close
 Closed
 Red, **Seat**
Curtains
 Corner
 Shop
 Opposite
 Threw
 Watching

Ran

WISE

Man/Men 0,20
 Clever 0,11
 Owl 0,11
 Foolish 0,07
 Sage 0,07

Bird, Dull, Elder, Gold,
 Know-all, Knowing,
 Knowledge, Learned, Old,
 Older, Scholar, Sensible,
 Smart, Solomon, Stupid,
 Thoughtful 0,02

Old 0,27
 Man/Men 0,25
 Clever 0,20
 Owl 0,09
 Stupid 0,05
 Know 0,04
 Wisdom 0,04
 Beard, Brainy, Courage,
 Intelligent, Knowledge,
 Sage 0,02

 $\alpha=0,02 \quad \beta=0,02 \quad \gamma=0,02$

Wisdom
Foolish
Learned
 Fool
 Strong, Fair
 Ye
 Human
 Soul

Truth
 None, War
 Earth
 Law
Knowledge
 Public, Art
 Neither
 Live, Land
 Advice
 Counsel
 Person
 Care
 Prince
 Self, Laws, Brave
 Noble, Evil, Spirit
 Hath
 Ought
 Master
 Sense
 Rich
 Age
 Happy
 Prudent, Youth
 Peace
 Hold
 Act
 Able, Free
 Fools, Show
 Heaven, Knows
 Ways, Opinion
 Queen, Makes, Unto
 Business
 Cause, City
 Question
 Beyond, Virtue, Policy
 Follow
 Sweet
 Ones, Experience
 Learn, Truly, According,
 Turn
 Future, Common, Grave,
 Brother

Books, Forth, Simple,
Subject
Book, Happiness, Plea-
sure, Thoughts
Rule, Desire
Living, Honest
Husband, Strength
Past, Comes
Character, Form, **Gold**
Daughter, Mighty, Clear
Fall, Regard
Bear, Seeing
Providence
Natural, States
Eye, Ancient
Worldly, Bad
Folk, Mean
Change, Loved
Dead, Matters
Fine, Lived, Seek, Family,
Sun
Ill, Gods, Purpose, Per-
sons, Worthy
Fortune

ZEBRA

Stripes 0,47
Crossing 0,29
Animal 0,09
Africa 0,04
Horse 0,04

Black+White, Jungle, Zoo
0,02

Crossing 0,33
Stripe(s) 0,27
Animal 0,09
Horse 0,09
Lion 0,04
Stripy 0,04
Africa, Desert, Elephant,
Elk, Kangaroo, Zoo 0,02

$\alpha=0,20$ $\beta=0,20$ $\gamma=0,20$

Crossing
Stripes
Giraffe
Buffalo
Ass, **Lion**
Striped, Legs, Species
Halictus, **Horse**

Animals
Animal
Area, Road
Traffic, **Africa**
Rhinoceros, **Elephant**
Quagga, Vehicle
Finches, Herd, Pedestrian,
Leopard, Game, Limits
Common, Regulation,
Lions
Herds, Killed, Eland,
Chief, Crossings, Lines,
Camp, Skin, Danios,
Flesh, Close, Meat
Plains, Stop, Ps., Across,
Uncontrolled, Barred,
Cross
South, Female, Hemionus,
Crab, Shoulder, Ground,
Dark, Deer, Trees, Front,
Tail, Sclusively, Ostriches,
Impala, Antelopes, Sha-
dows, Different

Centre, Manenko, Travel-
ling, Hartebeest, Masai,
Idea, Walls, Bold, Horses,
Hybrid, Hippopotamus,
Dog

ZIP		$\alpha=0, 40$ $\beta=0, 20$ $\gamma=0, 20$
Fastener 0,60	Trouser(s) 0,13	Coon
Trousers 0,13	Fly 0,11	Network
Dress 0,09	Button(s) 0,09	City
Skirt 0,09	Jacket 0,09	Street, Bullet, Secondary, Transmissions, Carrier, Satellite
	Clothes 0,07	Zenith, Code, Bullets, Subscribers, Makes, Address, List
	Up 0,07	Identifying, Past, County, Including, Primary, Sub- mit, Station
	Dress 0,04	Ledge, Lot, Transmission, Arrow, Zee, Snowshoes, None, Sharp
	Fastener 0,04	Ai, Affiliated, Coil, Clear, Bow, Flutter, Fellows, File, Commencing, Blaze, Alphabetical, Cut, Bent

Références

- AMSLER Robert A., *The Structure of the Merriam-Webster Pocket Dictionary*, Thèse de doctorat, University of Texas, Austin, Texas, 1980.
- ARMSTRONG Sharon L., Lila R. GLEITMAN et Henry GLEITMAN, « What some concepts might not be », *Cognition*, vol. 13, 1983, p. 263–308.
- AU T.K., « Language and cognition », Dans *Language perspectives : Acquisition, retardation, and intervention*, SCHIEFELBUSCH R.L. et L.L. LLOYD, dir., Pro-Ed, Austin, TX, 1988.
- AUSTIN J. L., *Quand dire, c'est faire.*, Seuil, Paris, 1962/1970.
- AYDEDE Murat, « Fodor on concepts and Frege puzzles », *Pacific Philosophical Quarterly*, vol. 79, n° 4, 1998, p. 289–294.
- BACHELARD Gaston, *La philosophie du non*, P.U.F., Paris, 1940.
- BAHL L. R., F. JELINEK et R. L. MERCER, « A Maximum-likelihood approach to continuous speech recognition », *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, vol. 5, n° 2, 1983, p. 179–190.
- BAKER James K., « The DRAGON system – An overview », *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-23, n° 1, 1975, p. 24–29.
- BAR-HILLEL Yehoshua, « A quasi-arithmetical notation for syntactic description », *Language*, vol. 29, 1953, p. 47–58, réimprimé dans Y. Bar-

- Hillel. (1964). *Language and Information : Selected Essays on their Theory and Application*, Addison-Wesley 1964, p. 61–74.
- BAUM L. E., « An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of a Markov Process », *Inequalities*, vol. 3, 1972, p. 1–8.
- BENNETT Jonathan, *Learning from six philosophers*, vol. 1, Oxford University Press, New York, 2001.
- BENZÉCRI Jean-Paul, *Pratique de l'analyse des données linguistique et lexicologique*, Dunod, Paris, 1981.
- , *Correspondence Analysis Handbook*, Marcel Dekker, New York, 1992.
- BLOOM Paul et Frank C. KEIL, « Thinking through language », *Mind and Language*, vol. 16, 2001, p. 351–367.
- BRÉAL Michel, *Essai de sémantique (Science des significations)*, Gérard Monfort, 1897/1982.
- BROWN P. F., S. DELLA PIETRA, V. DELLA PIETRA et R. L. MERCER, « Word Sense Disambiguation using Statistical Methods », Dans *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics, ACL*, 1991, p. 264–270.
- BURGE T, « Individualism and the mental », Dans *Midwest Studies in Philosophy IV : Studies in Metaphysics*, FRENCH P., T. UEHLING et H. WETTSTEIN, dir., Minneapolis : University of Minnesota Press, 1979.
- CARAMAZZA A. et E. GROBER, « Polysemy and the structure of the subjective lexicon », Dans *Semantics : Theory and application*, RAMEH C., dir., Georgetown University Press, Washington, DC, 1976.
- CHOMSKY Noam, « Three models for the description of language », *IRI*

- Transactions on Information Theory*, vol. 2, n° 3, 1956, p. 113–124.
- , *Syntactic Structures*, Mouton, The Hague, 1957.
- , *Knowledge of Language : Its Nature, Origin, and Use*, Praeger, New York, NY, 1986.
- , « Language and nature », *Mind*, vol. 104, n° 413, 1995, p. 1–61, Reprinted in Chomsky 2000.
- , *New Horizons in the Study of Language and Mind*, Cambridge University Press, Cambridge, 2000.
- , *Media Control : The Spectacular Achievements of Propaganda*, Seven Stories Press, New York, 2002.
- CHURCH Kenneth W. et William A. GALE, « A comparison of the enhanced Good-Turing and deleted estimation methods for estimating probabilities of English bigrams », *Computer Speech and Language*, vol. 5, 1991, p. 19–54.
- COLLINS A.M. et E.F. LOFTUS, « A spreading activation theory of semantic processing », *Psychological Review*, vol. 82, 1975, p. 407–428.
- DAGAN Ido et Alon ITAI, « Word sense disambiguation using a second language monolingual corpus », *Computational Linguistics*, vol. 20, n° 4, 1994, p. 563–596.
- DELEUZE Gilles, *Le pli*, Minuit, Paris, 1988.
- DELGADO Ana R, « Order in spanish colour words : Evidence against linguistic relativity », *British Journal of Psychology*, vol. 95, n° 1, 2004, p. 81–90.
- DENNETT Daniel C., *Kinds of Minds : Towards an Understanding of Consciousness*, Basic Books, New York, NY, 1996.
- DESCARTES René, *Les principes de la philosophie*, Gallimard, Paris, 1644/1953.

- EDMONDS Philip et Graeme HIRST, « Near-synonymy and lexical choice », *Computational Linguistics*, vol. 28, n° 2, 2002, p. 105–144.
- ENGELS, *Marx-Engels Correspondence*, International Publishers, 1895.
- FASS Dan et Yorick WILKS, « Preference semantics, ill-formedness, and metaphor », *Computational Linguistics*, vol. 9, n° 3–4, 1983, p. 178–187.
- FELLBAUM Christiane D., *WordNet : An Electronic Lexical Database*, MIT Press, New York, 1998.
- FILLMORE C. J., « Towards a descriptive framework for spatial deixis », Dans *Speech, Place, and Action*, JARVELLA R. J. et W. KLEIN, dir., Wiley, Chichester, 1982, p. 31–59.
- FODOR Jerry et Ernie LEPORE, « The red herring and the pet fish : why concepts still can't be prototypes », *Cognition*, vol. 58, 1996, p. 253–270.
- FODOR Jerry A., *The Language of Thought*, The Harvester Press Limited, Hassocks, Sussex, 1975, first published in the United States of America by Thomas Y. Crowell Company, Inc.
- , *The Elm and the Expert : Mentalese and Its Semantics*, MIT Press, Cambridge, MA, 1994.
- , *Concepts : Where Cognitive Science Went Wrong*, Oxford University Press, Oxford, 1998.
- FREGE Gottlob, « Function and concept », Dans *Frege Reader*, BEANEY Michael, dir., Blackwell, Oxford, 1891/1997.
- , « On sense and reference », Dans *Translations from the Philosophical Writings of Gottlob Frege*, GEACH P. T. et M. BLACK, dir., Blackwell, Oxford, 1892/1952.

- , « On concept and object », Dans *Frege Reader*, BEANEY Michael, dir., Blackwell, Oxford, 1892/1997.
- FURUI S., « On the role of spectral transition for speech perception », *The Journal of the Acoustical Society of America*, vol. 80, n° 4, 1986, p. 1016–1025.
- GALE W., K. W. CHURCH et D. YAROWSKY, « Estimating Upper and Lower Bounds on the Performance of Word Sense Disambiguation », Dans *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, ACL, 1992.
- GREFENSTETTE Gregory et Simone TEUFEL, « Corpus-based method for automatic identification of support verbs for nominalizations », Dans *Proceedings of the seventh conference on European chapter of the Association for Computational Linguistics*, Dublin, Ireland, 1995.
- GUTHRIE Joe A., Louise GUTHRIE, Yorick WILKS et Homa AIDINEJAD, « Subject-dependent co-occurrence and word sense disambiguation », Dans *ACL-91*, ACL, Berkeley, CA, 1991, p. 146–152.
- HABERT Benoît, Elie NAULLEAU et Adeline NAZARENKO, « Symbolic word clustering for medium-size corpora », Dans *Proceedings of the 16th conference on Computational linguistics*, Copenhagen, Denmark, 1996, p. 490–495.
- HABERT Benoît, Gabriel ILLOUZ et Helka FOLCH, « Des décalages de distribution aux divergences d’acception », Dans *Sémantique et corpus*, CONDAMINES Anne, dir., Hermès, Paris, à paraître.
- HEGEL Georg W. F., *Lectures on the History of Philosophy*, Univ of Nebraska, 1806/1995, traduit par E.S. Haldane.
- HIRSH Katherine. W. et Jeremy. J. TREE, « Word association norms for two cohorts of british adults », *Journal of Neurolinguistics*, vol. 14,

n° 1, 2001, p. 1– 44.

HIRST Graeme, *Semantic Interpretation and the Resolution of Ambiguity*, Studies in Natural Language Processing, Cambridge University Press, Cambridge, 1987.

IDE N. et J. VÉRONIS, « Introduction to the Special Issue on Word Sense Disambiguation : The State of the Art », *Computational Linguistics*, vol. 24, n° 1, 1998, p. 1–40.

INKPEN Diana Zaiu et Graeme HIRST, « Automatic sense disambiguation of the near-synonyms in a dictionary entry », Dans *Proceedings of the CICLing*, 2003, p. 258–267.

IWAŃSKA Lucja M., « Natural language is a powerful knowledge representation system : the UNO model », Dans *Natural Language Processing and Knowledge Representation : Language for Knowledge and Knowledge for Language*, IWAŃSKA Lucja M. et Stuart C. SHAPIRO, dir., AAAI Press, Menlo Park, 2000, p. 7–64.

JACKENDOFF Ray, *Semantic Structures*, MIT Press, Cambridge, MA, 1990.

JELINEK F., R. L. MERCER, L. R. BAHL et J. K. BAKER, « Perplexity – a measure of the difficulty of speech recognition tasks », *Journal of the Acoustical Society of America*, vol. 62, 1977, p. 63, supplement 1.

JI Hyungsuk et Sabine PLOUX, « Automatic contextonym organizing model », Dans *Proceedings of the 25th annual meeting of the Cognitive Science Society*, 2003, p. 622–627.

KAHNEMAN D. et A. TVERSKY, « On the psychology of prediction », *Psychology Review*, vol. 80, 1973, p. 237–251.

KANT Immanuel, *Critique of Pure Reason*, 1781/1929, URL <http://www.arts.cuhk.edu.hk/Philosophy/Kant/cpr>,

traduit par Norman Kemp Smith.

KAROV Y. et S. EDELMAN, « Learning Similarity-Based Word Sense Disambiguation from Sparse Data », Dans *Proceedings of the 4th Workshop on Very Large Corpora*, Copenhagen, Denmark, 1996.

KELLY Edward F. et Philip J. STONE, *Computer Recognition of English Word Senses*, North-Holland, Amsterdam, 1975.

KINTSCH Walter, « Predication. », *Cognitive Science*, vol. 25, 2001, p. 173–202.

KLEIN Devorah E. et Gregory L. MURPHY, « The representation of polysemous words », *Journal of Memory and Language*, vol. 45, 2001, p. 259–282.

KRIPKE Saul A., *Naming and Necessity*, Harvard University Press, Cambridge Massachusetts, 1972.

LAKOFF George, « Cognitive models and prototype theory », Dans *Concepts and conceptual development*, NEISSER U., dir., Cambridge University Press, Cambridge, 1987a.

———, *Women, Fire, and Dangerous Things*, University of Chicago Press, Chicago, 1987b.

LANDAUER Thomas K., Darrell LAHAM, Bob REHDER et M. E. SCHREINER, « How well can passage meaning be derived without using word order : A comparison of latent semantic analysis and humans », Dans *COGSCI-97*, Lawrence Erlbaum, Stanford, CA, 1997, p. 412–417.

LANDAUER Thomas K., Peter W. FOLTZ et Darrell LAHAM, « An introduction to Latent Semantic Analysis », *Discourse Processes*, vol. 25, 1998, p. 259–284.

LAURENCE Stephen et Eric MARGOLIS, « Concepts and cognitive science », Dans *Concepts-Core readings*, MARGOLIS Eric et Stephen

- LAURENCE, dir., MIT Press, Cambridge, MA, 1999.
- , « Radical concept nativism », *Cognition*, vol. 86, n° 1, 2002, p. 25–55.
- L'ECUYER P., « An efficient and portable combined random number generator », *Communications of the ACM*, vol. 31, n° 6, 1998, p. 742–749.
- LEE Lillian et Fernando PEREIRA, « Distributional similarity models : Clustering vs. nearest neighbors », Dans *Proceedings of the 37th Annual Meeting of the ACL*, 1999, p. 33–40.
- LEHRER A., « Polysemy, conventionality, and the structure of the lexicon », *Cognitive Linguistics*, vol. 1, 1990, p. 207–246.
- LESK Michael E., « Automatic sense disambiguation using machine readable dictionaries : How to tell a pine cone from an ice cream cone », Dans *Proceedings of the 1986 SIGDOC Conference, Toronto*, Association for Computing Machinery, New York, 1986, p. 24–26.
- LYONS John, *Semantics*, Cambridge University Press, Cambridge, 1977.
- MALINOWSKI Bronislaw, « The problem of meaning in primitive languages », Dans *The Meaning of Meaning*, OGDEN C.K. et I.A. RICHARDS, dir., Harcourt Brace, New York, 1923.
- MCCLELLAND James L. et Jeffrey L. ELMAN, « Interactive processes in speech perception : The TRACE model », Dans *Parallel Distributed Processing Volume 2 : Psychological and Biological Models*, MCCLELLAND James L., David E. RUMELHART et THE PDP RESEARCH GROUP, dir., MIT Press, Cambridge, MA, 1986, p. 58–121.
- MCCLOSKEY M. E. et S. GLUCKSBERG, « Natural categories : Well defined or fuzzy sets ? », *Memory and cognition*, vol. 6, 1978, p. 462–472.
- MERVIS C. B. et E. ROSCH, « Categorization of natural objects », *Annual Review of Psychology*, vol. 32, 1981, p. 89–115.

- MICHIELS A., J. MULLENDERS et J. NOËL, « Exploiting a large data base by Longman », Dans *Proceedings of the 8th International Conference on Computational Linguistics, COLING'80*, 1980.
- MILL John Stuart, *Système de logique*, Mardaga, Paris, 1843/1988, Œuvre original publié en 1843 sous le titre *A System of Logic*.
- MILLER George A. et Noam CHOMSKY, « Finitary models of language users », Dans *Handbook of Mathematical Psychology*, LUCE R. Duncan, Robert R. BUSH et Eugene GALANTER, dir., vol. II, John Wiley, New York, 1963, p. 419–491.
- MURPHY GL, « On metaphoric representation », *Cognition*, vol. 60, 1996, p. 173–204.
- NUNBERG Geoffrey, « The non-uniqueness of semantic solutions : Polysemy », *Linguistics and Philosophy*, vol. 3, n° 2, 1979, p. 143–184.
- OLNEY John C., « To all interested in the merriam-webster transcripts and data derived from them », Technical Report L-13579, System Development Corporation, Santa Monica, CA, October, 1968.
- OSHERSON Daniel N. et Edward E SMITH, « On the adequacy of prototype theory as a theory of concepts », *Cognition*, vol. 9, 1981, p. 35–58.
- PALMER F. R., *Semantics*, Cambridge University Press, 1981.
- PALMER John A., *Plato's Reception of Parmenides*, Oxford University Press, 1999.
- PLOUX Sabine, « Modélisation et traitement informatique de la synonymie », *Linguisticae Investigationes*, vol. XXI, n° 1, 1997, p. 1–28.
- PLOUX Sabine et Hyungsuk JI, « A model for matching semantic maps between languages (French / English, English / French) », *Computational Linguistics*, vol. 29, n° 2, 2003, p. 155–178.

- PLOUX Sabine et Bernard VICTORRI, « Construction d'espaces sémantiques à l'aide de dictionnaires informatisés des synonymes », *TAL*, vol. 39, n° 1, 1998, p. 161–182.
- PUSTEJOVSKY James, *The Generative Lexicon*, MIT Press, Cambridge, MA, 1995.
- PUTNAM Hilary, « Explanation and reference », Dans *Mind, Language and Reality*, Cambridge University Press, Cambridge, England, 1975a.
- , « The meaning of 'meaning' », Dans *Mind, Language and Reality*, Cambridge University Press, Cambridge, England, 1975b.
- QUILLIAN M. ROSS, « Semantic memory », Dans *Semantic Information Processing*, MINSKY Marvin, dir., MIT Press, Cambridge, MA, 1968, p. 227–270.
- REBOUL Anne, « Words, concepts, mental representations, and other biological categories », Dans *The Lexicon-Encyclopedia Interface*, PEETERS B., dir., Benjamins, Amsterdam, à paraître.
- RESNIK Philip, « Selectional preference and sense disambiguation », Dans *Proceedings of ACL SIGLEX Workshop on Tagging Text with Lexical Semantics*, ACL, Washington, D.C., 1997, p. 52–57.
- RICHENS R. H., « Interlingual machine translation », *The Computer Journal*, vol. 1, n° 3, 1958, p. 144–147.
- RIPS L., J. SHOBEN et E. SMITH, « Semantic distance and the verification of semantic relations », *Journal of Verbal Learning and Verbal Behavior*, vol. 12, 1973, p. 1–20.
- ROSCH Eleanor, « Natural categories », *Cognitive Psychology*, vol. 4, 1973a, p. 328–350.
- , « On the internal structure of perceptual and semantic categories », Dans *Cognitive Development and Acquisition of Language*, MOORE

- T. E., dir., Academic Press, New York, NY, 1973b.
- , « Cognitive representations of semantic categories », *Journal of Experimental Psychology*, vol. 104, 1975, p. 192–233.
- RUSSELL Bertrand, « Vagueness », *The Australasian Journal of Psychology and Philosophy*, vol. 1, 1923, p. 84–92.
- SAPIR Edward, « The status of linguistics as science », Dans *Selected Writings of Edward Sapir*, University of California Press, Berkeley, CA, 1929.
- SCHANK R. C. et R. P. ALBELSON, *Scripts, Plans, Goals and Understanding*, Lawrence Erlbaum, Hillsdale, NJ, 1977.
- SCHÜTZE Hinrich, « Automatic word sense discrimination », *Computational Linguistics*, vol. 24, n° 1, 1998, p. 97–123.
- SHANNON C. E., « A mathematical theory of communication », *Bell System Technical Journal*, vol. 27, n° 3, 1948, p. 379–423, continued in following volume.
- SINCLAIR John, « Lexicographic evidence », Dans *Dictionaries, Lexicography and Language Learning*, BRUMFIT C.J., dir., Pergamon Press, Oxford, UK, 1985.
- SMALL S. L., « A distributed word-based approach to parsing », Dans *Natural Language Parsing Systems*, BOLC L., dir., Springer, Berlin, Heidelberg, 1987, p. 161–201.
- SONG Jae Jung, *Linguistic Typology : Morphology and Syntax*, Pearson Education, Harlow and London, 2001.
- STICH Stephen P., *The Fragmentation of Reason*, MIT Press, Cambridge, MA, 1990.
- TVERSKY Amos, « Features of similarity », *Psychological Review*, vol. 84, n° 4, 1977, p. 327–352.

- VYGOTSKY L. S., *Mind in Society*, Harvard University Press, Cambridge, MA, 1930/1978.
- , *Thought and Language*, MIT Press, Cambridge, MA, 1934/1986.
- WEAVER Warren, « Translation », Dans *Machine Translation of Languages*, LOCKE William N. et A. Donald BOOTHE, dir., MIT Press, Cambridge, MA, 1949/1955, p. 15–23.
- WEBER Max, *The Methodology of the Social Sciences*, Free Press, New York, 1917/1949.
- , « Science as a vocation », Dans *From Max Weber : Essays in sociology*, GERTH H. et CW MILLS, dir., Oxford University Press, New York, 1919/1975.
- WEHRLI Eric, « Translating idioms », Dans *Proceedings of COLING98, Montreal*, 1998, p. 1388–1392.
- WEISS S., « Learning to disambiguate », *Information Storage and Retrieval*, vol. 9, 1973, p. 33–41.
- WHORF Benjamin Lee, *Language, Thought and Reality : Selected Writing of Benjamin Lee Whorf*, MIT Press, Cambridge, MA, 1956, edited with an introduction by John B. Carroll.
- WILKS Yorick A, « On-line semantic analysis of english texts », *Mechanical Translation*, vol. 11, n° 3–4, 1968, p. 59–72.
- WILKS Yorick A. et Dan FASS, « Preference semantics : A family history », Report MCCS-90-194, Computing Research Laboratory, New Mexico State University, Las Cruces, NM, 1990.
- WILKS Yorick A., Dan C. FASS, Chengming GUO, James E. MCDONALD, Tony PLATE et Brian M. SLATOR, « A tractable machine dictionary as a resource for computational semantics », Dans *Computational Lexicography for Natural Language Processing*, BOGURAEV Branimir K. et

Ted BRISCOE, dir., Longman Group Limited, Harlow, Essex, England, 1988.

WITTGENSTEIN Ludwig, *Philosophical Investigations*, Basil Blackwell, Oxford, 1953/1972.

YAROWSKY David, « Unsupervised word sense disambiguation rivaling supervised methods », Dans *Proceedings of the 33rd Annual Meeting of the ACL*, Cambridge, MA, 1995, p. 189–196.

ZIFF Paul, *Semantic Analysis*, Cornell University Press, 1960.