

Université de Reims Champagne-Ardenne
U.F.R. des Lettres et Sciences humaines
CIRLEP E.A. 4299

École doctorale « Sciences de l'Homme et de la Société » – ED 462

Thèse de Doctorat en Sciences du langage

Mélisandre Caure

**Caractérisation de la transparence lexicale, extension de
la notion par ajustements graphophonologiques et
microsémantiques, et application aux lexiques de
l'anglais, de l'allemand et du néerlandais**

Thèse dirigée par Jean-Emmanuel Tyvaert et Éric Castagne

Présentée et soutenue publiquement le 2 décembre 2009

Devant un jury composé de :

M. Éric Castagne, Université de Reims Champagne-Ardenne
Mme Danièle Flament-Boistrancourt, Université Paris X
M. Bernard Laks, Université Paris X
M. Philippe Monneret, Université de Bourgogne
M. Jean-Emmanuel Tyvaert, Université de Reims Champagne-Ardenne

La préparation de cette thèse a été financée par le Ministère de l'Éducation nationale, de
l'Enseignement supérieur et de la Recherche.

Remerciements

Je tiens à remercier en premier lieu mes directeurs de recherche, Jean-Emmanuel Tyvaert et Éric Castagne. Ils ont su, au cours de nombreux rendez-vous et par mails, guider ma recherche et me donner de précieux conseils. Ce fut un véritable avantage, pendant ces trois années, de recevoir une double lecture/écoute et de bénéficier ainsi d'avis complémentaires. Ils m'ont également intégrée dans un programme de recherche, le programme ICE, ont tenu à ce que je présente des interventions dans des séminaires, et à ce que je publie mes premiers articles ; ces expériences ont participé à ma formation de chercheuse. Je les remercie pour leur disponibilité, leur aide et pour l'ambiance de travail très agréable qu'ils ont installée.

Je suis extrêmement reconnaissante à Mme Flament-Boistrancourt, à M. Laks et à M. Monneret d'avoir accepté de lire et d'évaluer mon travail.

Je remercie également le Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche d'avoir financé la préparation de cette thèse.

Je remercie enfin les membres de ma famille et mes amis, pour leur intérêt constant pour mon travail et son avancée, David pour ses conseils sur la langue anglaise, Joris pour ses conseils sur la langue néerlandaise, Philippe pour sa relecture, et Tilman, à qui je dois tant. Je lui suis à jamais redevable pour les heures de discussion que nous avons eues sur mon sujet en particulier et sur la linguistique en général, pour ses fructueuses remises en question et pour son soutien sans faille.

Résumé en français :

L'objectif de ce travail est de définir la transparence lexicale, c'est-à-dire d'établir quels sont les critères que doit respecter un mot d'une langue donnée pour pouvoir être considéré comme potentiellement compréhensible par un lecteur ayant une autre langue pour langue maternelle. Il s'agit ici de donner une définition objective et en synchronie de la transparence.

Nous portons notre attention sur la compréhension de langues dont la proximité avec le français est entendue comme une proximité géographique plus que génétique, pour rendre compte de l'intercompréhension rendue possible par les contacts qu'ont entretenus et que continuent à entretenir les langues. Nous nous intéressons ainsi aux langues germaniques que sont l'anglais, l'allemand et le néerlandais, pour en étudier la transparence vis-à-vis du français.

Une première étape de ce travail est consacrée à la caractérisation de la transparence en ce qu'elle représente la qualité des mots immédiatement compréhensibles par des lecteurs débutants. Dans un deuxième temps, nous nous attachons à donner une vision « étendue » de la transparence, de sorte qu'elle rende compte de tous les mots dont des apprenants qui cherchent à comprendre un texte peuvent, dans les faits, tirer profit. Pour ce faire, nous développons des méthodes qui permettent de jouer sur la forme et le sens de mots lors du processus de compréhension.

Nous terminons notre travail en appliquant les principes théoriques dégagés à l'analyse d'un corpus des 1 000 mots lexicaux les plus fréquents des trois langues germaniques citées comparées au français, et en tirons des données d'ordres quantitatif et qualitatif.

Mots-clés en français :

Transparence, lexique, phonologie, graphème, sémantique, champ sémantique, intercompréhension, multilinguisme, contrastivité.

Titre en anglais :

« Characterization of lexical transparency, extension by graphophonological and microsemantic adjustments, and application to English, German and Dutch lexicon »

Résumé en anglais :

The aim of this study is to define lexical transparency, establishing the criteria that a word from a given language must fulfill in order to be considered as potentially understandable to a non-native speaker. The aim is to give a definition of transparency which is both objective and synchronic.

We focus our attention on the understanding of languages whose proximity to French is considered as more geographic than genetic. This way, it is possible to study the intercomprehension which arises due to contacts that have existed and still exist between the languages. Thus, we focus on three Germanic languages, English, German and Dutch, and examine their transparency with respect to French.

Firstly, we characterize transparency as the property of words which are immediately understandable by beginners. Secondly, we propose an « extended » definition of transparency, including all words which may be understood by learners who would try to understand a text. To this end, we develop methods which allow to play with the form and the meaning of words during the comprehension process.

To end this study, we use the theoretical principles to analyse a corpus that contains the 1 000 most frequent lexical words of the three Germanic languages with respect to French, and draw from this analysis quantitative and qualitative conclusions.

Mots-clés en anglais :

Transparency, lexicon, phonology, grapheme, semantics, semantic field, intercomprehension, multilingualism, contrastivity.

Liste des abréviations et symboles utilisés

- La mention EN devant un mot signifie qu'il s'agit d'un mot du lexique de l'anglais.
- La mention DE devant un mot signifie qu'il s'agit d'un mot du lexique de l'allemand.
- La mention NL devant un mot signifie qu'il s'agit d'un mot du lexique du néerlandais.
- La mention FR devant un mot signifie qu'il s'agit d'un mot du lexique du français.
- Les mots étrangers sont cités en italique et les mots français entre guillemets.

- Conformément à la norme en vigueur, les sons sont notés entre crochets (ex. : [p]), les phonèmes entre barres obliques (ex. : /p/) et les graphèmes entre chevrons (ex. : <p>).

SOMMAIRE

Introduction	8
1. Prolégomènes	15
1.1. Une première approche de la notion de transparence	16
1.1.1. La notion de mot	16
1.1.2. Les définitions traditionnelles	23
1.1.3. Tentatives de caractérisation précise	25
1.2. Une nouvelle conception de la transparence : les programmes d'intercompréhension.	33
1.2.1. Le multilinguisme et l'intercompréhension.....	33
1.2.2. La formation à l'intercompréhension	41
1.2.2.1. Les différents programmes européens.....	41
1.2.2.2. Le programme ICE	44
1.2.3. Aménagements de la définition traditionnelle de la transparence.....	64
2. La transparence lexicale.....	74
2.1. Pour une révision de la notion de transparence	75
2.1.1. De nouveaux objectifs	75
2.1.2. De l'intersystème à l'hypersystème.....	82
2.1.3. Du statut inégal de la forme et de la signification	91
2.2. La transparence directe	95
2.2.1. La proximité formelle directe	95
2.2.1.1. Détermination des critères définitionnels	95
2.2.1.2. Importance des consonnes	106
2.2.1.3. Importance du début du mot.....	112
2.2.2. La proximité sémantique directe	117
2.3. La transparence indirecte.....	128
2.3.1. La proximité formelle indirecte : l'ajustement formel	128
2.3.2. La proximité sémantique indirecte : l'ajustement sémantique	153
2.4. Bilan.....	175
3. De l'interlexique à l'hyperlexique	180
3.1. Présentation du corpus.....	183
3.2. L'interlexique	189
3.2.1. Détermination de l'interlexique anglais/français.....	198
3.2.2. Détermination de l'interlexique allemand/français	201
3.2.3. Détermination de l'interlexique néerlandais/français.....	204
3.2.4. Bilan comparatif des interlexiques	206
3.3. L'hyperlexique.....	210

3.3.1. Détermination de l'hyperlexique anglais/français	237
3.3.2. Détermination de l'hyperlexique allemand/français.....	241
3.3.3. Détermination de l'hyperlexique néerlandais/français	245
3.3.4. Bilan comparatif des hyperlexiques	248
3.4. Bilan.....	255
Conclusion.....	260
Bibliographie	266
Annexes	293
Tableau de l'interlexique anglais/français	293
Tableau de l'interlexique allemand/français.....	307
Tableau de l'interlexique néerlandais/français	313
Tableau de l'hyperlexique anglais/français	319
Tableau de l'hyperlexique allemand/français.....	344
Tableau de l'hyperlexique néerlandais/français	368
Tableau des mots anglais opaques.....	392
Tableau des mots allemands opaques.....	396
Tableau des mots néerlandais opaques.....	403
Index	410

Introduction

Dans le cadre de l'apprentissage d'une langue étrangère, le lexique bénéficie d'une attention particulière, tant pour les didacticiens et les linguistes que pour les apprenants. Nous ne faisons pas référence ici, lorsque nous évoquons la didactique des langues, aux nombreux manuels de vocabulaire¹, qui proposent de passer en revue les principaux mots et locutions d'une langue, classés par thème, et de donner à l'apprenant un bagage considéré comme nécessaire et suffisant pour la maîtrise de l'expression et de la compréhension dans une langue donnée. Ces manuels trouvent leur pendant syntaxique dans les tout aussi nombreuses grammaires², qui proposent d'expliquer le fonctionnement d'une langue et de donner à l'apprenant les moyens de produire des énoncés correctement construits (c'est-à-dire construits selon la norme exposée).

On peut voir que le lexique occupe une place particulière pour les didacticiens et les linguistes par les travaux qu'ils consacrent à l'exposé de son *rôle* et de la façon dont il est *mis à profit* lors de la compréhension d'un texte rédigé dans une langue étrangère. Il n'est pas seulement consigné par écrit en tant qu'objet à apprendre mais aussi étudié pour lui-même en tant qu'outil. Nous en voulons pour preuves les ouvrages et articles consacrés à l'apprentissage du lexique, parmi lesquels on peut citer *Le vocabulaire dans l'apprentissage des langues étrangères*, de P. Bogaards, publié en 1994, « Quelques remarques sur l'apprentissage du lexique d'une langue étrangère », de B. Py, publié en 1997 et « Approche de l'apprentissage lexical et du rôle du lexique dans plusieurs méthodes et manuels de catalan », de P. Reynés, publié en 2004. De même, tous les travaux traitant des faux amis constituent des méta-données sur l'attention qu'il faut porter à certains mots. C'est dans une moindre mesure que la syntaxe a fait l'objet d'études ayant pour but de décrire comment les apprenants peuvent s'appuyer sur des indices syntaxiques dans leur processus de compréhension³. Le lexique, au contraire, dans ses aspects morphologique, formels (graphique et phonologique) et sémantique, constitue un objet d'étude préférentiel. Signalons que nous distinguons, comme J. Picoche (Picoche 1997), le

¹ Citons par exemple pour la langue anglaise : Bouscaren & Rivière, *Anglais : les 1 800 mots de base*, 1993 ; Rey et al., *Le mot et l'idée : anglais*, 2, 2003 ; Fromonot et al., *Vocabulaire de l'anglais contemporain*, 2009.

² Gabilan, *Grammaire expliquée de l'anglais*, 2006 ; Larreya & Rivière, *Grammaire explicative de l'anglais*, 2007 ; Malavieille & Rotgé, *Anglais : la grammaire*, 2008.

³ Citons néanmoins l'article de G. Noizet et M. Vion, « Les stratégies de compréhension dans le traitement des relations fonctionnelles de base », publié en 1983.

lexique du vocabulaire en considérant le premier comme l'ensemble des mots d'une langue et le second comme un sous-ensemble du premier (on peut alors parler du vocabulaire d'une profession, d'un auteur, etc.) :

*On conviendra d'appeler **lexique** l'ensemble de mots qu'une langue met à la disposition des locuteurs, et **vocabulaire** l'ensemble des mots utilisés par un locuteur donné dans des circonstances données. Le **lexique** est une réalité de langue à laquelle on ne peut accéder que par la connaissance des **vocabulaires** particuliers qui sont une réalité de discours. (Picoche 1997 : 45-46)*

Un numéro récent de la revue *Les langues modernes* (le n° 1/2009) a ainsi pour titre « Enseigner et apprendre le lexique ». Dans l'introduction à ce numéro, F. R. Garnier affirme en quoi, selon lui, le lexique est « primordial » :

Or, si la maîtrise approximative de la grammaire permet quand même assez souvent de comprendre, au moins à l'oral (si l'on fait une faute de genre en anglais, on sera quand même souvent surpris, ce qui vaut également pour des fautes de déclinaison en allemand), une mauvaise compréhension du lexique peut davantage provoquer des malentendus (dites à un anglophone que vous avez trop chaud au soleil et ajoutez : « Why don't we go into the shadow ? »³ - la réaction risque d'être amusante). L'enseignement du lexique est primordial. Dans son Tractatus logico-philosophicus, Wittgenstein écrit que les limites de notre langage sont les limites de notre monde⁴. Certes, nos élèves parlent une langue, voire deux, en dehors de celles qu'ils apprennent à l'école. Mais c'est bien le lexique qui permet à l'apprenant de réellement saisir les nuances et les subtilités de la langue et aussi de saisir la façon de penser de l'autre. C'est à travers les mots et leur sens que l'élève apprend les différences culturelles qui existent entre les différents pays et les différentes traditions. Ce sont ces mots qui lui permettront de faire connaissance avec la culture de l'autre. (Garnier 2009 : 13)

3/. « Why don't we go into the shadow ? » - « Et si on se mettait à l'ombre ? », mais l'ombre est ici l'ombre projetée d'une personne ou d'un objet, contrairement à l'endroit qui n'est pas éclairé par la lumière du soleil (shade).

4/. WITTGENSTEIN, Ludwig (1921). Tractatus logico-philosophicus. Logisch-philosophische Abhandlung. Suhrkamp : Francfort sur le Main, 115 p. (5.6).

Le lexique, qui fait donc l'objet d'études portant sur son rôle dans le processus d'apprentissage, est également considéré comme un élément majeur par les apprenants eux-mêmes, qui ont, semble-t-il, tendance à construire leur représentation d'un texte qu'ils cherchent à comprendre en mettant côte à côte les informations lexicales dont ils disposent, ou à concentrer leur motivation sur la découverte du vocabulaire d'un texte. C. Blanche-Benveniste indique que, pour les apprenants qui suivaient la méthode de formation à l'intercompréhension EuRom⁴, la compréhension du sens d'un texte coïncidait souvent avec la compréhension du sens des mots : l'« intérêt [des participants], pendant plusieurs semaines, ne portaient [sic] [...] que sur le lexique » (Blanche-Benveniste 2004 : 42).

On considère volontiers, même si cela pourrait évidemment être discuté, que la connaissance d'une langue se mesure par l'étendue du vocabulaire d'un locuteur, et que la syntaxe n'intervient que secondairement, pour lier entre eux les éléments lexicaux :

Apprendre une langue, c'est 1° apprendre des mots, 2° construire des phrases. [...] Les règles de grammaire seraient de peu d'utilité sans la connaissance d'un certain nombre de mots. Ce sont les matériaux indispensables avec lesquels ont [sic] doit construire des phrases, comme un architecte doit rassembler pierres et poutres pour construire une maison. (Geysen 1990 : 7)

On se représente volontiers la langue comme un ensemble de mots organisé par des règles grammaticales. Lorsque, dans un groupe de personnes profanes, il est question de langue, la discussion porte volontiers sur les mots et leur prononciation (ou leur orthographe), et sans doute plus rarement sur la morphologie ou la syntaxe. (Py 1997 : 176)

Dans les études portant sur les moyens utilisés par les apprenants pour accéder au lexique d'une langue, l'attention est notamment portée sur la façon dont ils mettent en relation les mots de la langue étrangère avec ceux de leur langue maternelle. Nous entendons par « langue maternelle » ce que d'autres entendent par « langue première », c'est-à-dire que nous n'incluons pas dans sa définition la figure de la mère. La langue maternelle d'un locuteur est « tout simplement celle qu'il a acquise en premier, chronologiquement, au moment du développement de sa capacité de langage » (Martinez

⁴ Nous consacrerons un développement aux programmes d'intercompréhension dans la section 1.2.2. de ce travail.

2004 : 18). Une langue étrangère sera pour un locuteur toute langue autre que sa langue maternelle, qu'elle soit seconde ou apprise pendant une courte période dans le cadre scolaire.

Pour comprendre un texte rédigé dans une langue étrangère, il est conseillé – par les enseignants lors de l'apprentissage ou par les didacticiens dans leurs travaux de recherche – de prendre appui sur les mots qui peuvent le plus facilement être mis en relation avec des éléments de leur langue maternelle, à savoir les « mots transparents » (qui connaissent également, nous aurons l'occasion d'y revenir, d'autres appellations). Le terme de transparence est parfois utilisé à d'autres fins que la description du lexique, par exemple ici par É. Castagne :

Nous appelons « séquences transparentes » les séquences, qu'elles soient graphiques, phoniques, morphologiques, ou syntaxiques, qui peuvent être jugées lisibles par un lecteur « débutant » dans une langue étrangère (LE) à partir des seules connaissances extraites dans sa langue maternelle (LM). (Castagne 2007a : 157)

Chez la majorité des auteurs, la transparence est néanmoins invoquée pour décrire les propriétés de mots d'une langue donnée immédiatement compréhensibles par des locuteurs ayant une autre langue pour langue maternelle. Elle est définie comme une identité formelle et sémantique ou une quasi-identité formelle et sémantique entre deux mots de langues différentes. Cette définition, pour le moins imprécise, permet certes aux auteurs qui l'utilisent de qualifier ponctuellement un mot de « transparent » s'ils l'évaluent comme suffisamment proche de forme et de signification d'un mot d'une autre langue. Elle interdit cependant toute qualification objective et, partant, toute quantification. En effet, si l'on définit la transparence comme une (quasi-)similitude formelle et sémantique entre deux mots, le choix de l'appartenance d'un mot donné à la catégorie des mots transparents reste entièrement subjectif, et susceptible de différer d'une personne à l'autre. Ainsi, s'il semble irréfutable que le nom néerlandais *moment* est transparent par rapport au nom français « moment », qu'en est-il pour l'adjectif *lang* (« long »), qui diffère d'une lettre, puis pour *plaats* (« place »), et à plus forte raison pour *natuurlijk* (« naturel ») ou *kamer* (« chambre ») ? Si la limite à fixer entre la transparence et la non-transparence (ou opacité) est affaire de choix personnel, il devient impossible de délimiter la part des mots

transparents d'un lexique. Le premier but que nous avons fixé à notre travail est ainsi de caractériser précisément la notion de transparence lexicale, telle qu'elle est utilisée par tous. Il s'agit donc de définir quels sont les critères formels et sémantiques qu'un mot doit respecter pour pouvoir être dit identique ou quasi-identique à un mot d'une autre langue, et ainsi de définir jusqu'où peut aller la quasi-identité avant de se transformer en altérité. H. Walter, dans son article « Les 'faux amis' anglais et l'autre côté du miroir », publié en 2001, s'interrogeait déjà sur la possibilité de fixer une limite entre la transparence (identité ou quasi-identité) et l'opacité :

Le problème à résoudre est double, à la fois sur le plan du signifiant (« à partir de quand peut-on considérer que deux formes sont semblables, sinon identiques ? ») et sur le plan du signifié (« le sens d'une forme donnée se retrouve-t-il en tous points identique dans l'autre langue ? »). (Walter 2001b : 103)

Elle insiste ensuite sur la limite qu'il y aurait à fixer pour pouvoir déterminer si une ressemblance, sur le plan du signifiant, est « suffisante » :

Mais à partir de quand peut-on considérer cette ressemblance comme suffisante ? Des formes comme biology et biologie peuvent sans doute être considérées comme analogues : différence minime du signifiant, à laquelle s'ajoutent l'équivalence répétée dans analogy/analogie, anatomy/anatomie, philosophie/philosophy, etc., et la confirmation apportée par un étymon identique. Mais peut-on considérer la forme anglaise considerate comme l'équivalent du français considérer, tout comme l'angl. baton et le fr. bâton, ou encore l'angl. ancient et le fr. ancien et même l'angl. literature et le fr. littérature, attendu que les différences formelles y sont minimales ? (ibid. : 103-104)

Elle conclut de ces interrogations que « [l]es discussions sur les limites toujours floues de cette ressemblance risquent d'être interminables » (*ibid.* : 104). Nous nous proposerons, dans ce travail, de mettre au jour des critères qui, précisément, permettent de fixer une limite, sur le plan formel, entre une ressemblance suffisante et une ressemblance insuffisante. Une réflexion sur l'identité ou la quasi-identité de signification nous permettra ensuite de définir la transparence dans son ensemble, c'est-à-dire considérée dans ses aspects formel et sémantique.

Il sera ensuite temps de s'interroger sur les éventuelles insuffisances des critères formulés pour définir la transparence. En effet, plusieurs équipes de recherche travaillant sur l'intercompréhension en Europe ont montré que des mots étrangers qui, à première vue, ne sont ni strictement identiques ni presque identiques à des mots appartenant au lexique de la langue maternelle d'apprenants, donc qui ne répondent pas aux critères de transparence, sont tout de même utilisés par ces apprenants pour accéder au sens de textes. Il semble donc que la définition de la transparence qui aura été fixée dans un premier temps demande à être étendue pour pouvoir rendre compte de ces mots qui, sans être immédiatement compréhensibles par des débutants, sont tout de même identifiables moyennant un jeu sur leur forme et/ou leur sens.

Notre travail s'intègre précisément dans une réflexion sur l'intercompréhension. Les membres du programme de recherche dans le cadre duquel notre étude prend place, à savoir le programme ICE (InterCompréhension Européenne), développé à l'université de Reims, travaillent sur l'acquisition de langues étrangères. Cette recherche est constituée d'un volet théorique (réflexion sur les processus à l'œuvre lors de la compréhension de textes), et d'un volet applicatif, qui se nourrit du volet théorique (méthode de formation à l'intercompréhension et observations en séances). Cette méthode a pour but de former des apprenants à lire des textes rédigés dans des langues qu'ils ignorent éventuellement, mais qui ne diffèrent pas trop de celle(s) qu'ils connaissent déjà : ils peuvent par exemple identifier une partie du lexique et des structures syntaxiques, et les alphabets et conventions d'écriture sont pratiquement identiques. La formation à l'intercompréhension est intégrée, depuis la rentrée 2005, aux enseignements dispensés dans le cadre d'un Master de l'université de Reims, le Master GMI⁵ (Gestion Multilingue de l'Information). Les séances de formation sont organisées autour de deux groupes de langues appartenant à la même famille⁶, si bien que l'on demande à un apprenant de comprendre des textes rédigés dans des langues appartenant au même groupe que celui de sa langue maternelle (ou d'une langue autre que sa langue maternelle, mais qu'il maîtrise également), et d'autres textes rédigés dans des langues appartenant à un autre groupe que celui de sa langue maternelle (ou d'une langue qu'il maîtrise). Ainsi, on propose à un apprenant francophone

⁵ Pour des précisions sur cette formation, voir la page d'informations : http://www.univ-reims.fr/index.php?p=107&art_id=89#sommaire (consultée le 16/10/09).

⁶ Nous réservons le terme « famille » aux plus grands regroupements de langues (par exemple la famille indo-européenne), et celui de « groupe » aux différentes subdivisions de ces familles (par exemple les groupes des langues romanes, germaniques ou slaves). Nous suivons en cela la terminologie présentée dans l'arbre de parenté des langues indo-européennes de R. Breton, reproduit dans ce travail p. 77. Pour une étude sur les relations historiques entre langues, voir notamment Bader 1994, Cavalli-Sforza 1996 et Ruhlen 1996.

des textes rédigés dans les langues romanes que sont le portugais, l'espagnol, l'italien et le roumain, et dans les langues germaniques que sont l'anglais, l'allemand et le néerlandais. Le choix de faire travailler des francophones sur des langues germaniques est spécifique à ce programme (la proximité entre langues romanes étant, dans les autres programmes, la seule exploitée pour un public de francophones). L'expérience prouve que ce choix fut pertinent, car l'intercompréhension entre groupes de langues distincts s'est révélée effectivement possible. Nous allons donc concentrer notre attention sur les trois langues germaniques citées ci-dessus pour définir et étendre la notion de transparence, à la lumière des principes de lecture approximative prônés par les différents programmes d'intercompréhension.

La première partie de ce travail (1. Prolégomènes) sera consacrée à l'exposé des différents travaux qui ont pour objet une réflexion sur la notion de transparence. Nous y présenterons les travaux qui utilisent ou qui cherchent à définir la transparence dans son acception restreinte (c'est-à-dire celle qui rend compte des mots immédiatement compréhensibles par des lecteurs débutants), puis présenterons l'intercompréhension, ses principes, et les travaux d'équipes qui se sont donné pour but de caractériser les différents moyens permettant d'accéder à la compréhension de mots qui ne sont pas immédiatement compréhensibles. La deuxième partie (2. La transparence lexicale) sera dévolue au développement de notre apport théorique à la définition de la transparence. Nous y présenterons les critères formels et sémantiques qui permettent de donner une caractérisation objective de la transparence, définie « en texte », et exposerons comment doit être, selon nous, pensée l'extension de la notion. Enfin, la troisième partie (3. De l'interlexique à l'hyperlexique) sera le lieu de l'application des principes théoriques dégagés. En transposant les critères de la deuxième partie à une analyse « hors texte », nous y déterminerons le nombre de mots transparents dans les lexiques de l'anglais, de l'allemand et du néerlandais par rapport au français, en prenant pour base de travail un corpus constitué des mots considérés comme les plus fréquents de ces langues. Nous aurons ainsi l'occasion d'exploiter la théorie pour l'analyse d'un corpus et de tirer les conclusions des résultats qui auront été obtenus.

1. Prolégomènes

L'objet de cette première partie est d'examiner les différentes études dans lesquelles ont été proposées des réflexions sur la notion de transparence lexicale. Les travaux présentés dans le premier chapitre (1.1. Une première approche de la notion de transparence) se rapportent à la transparence en tant que propriété des mots identiques ou quasi-identiques entre langues, tandis que les travaux présentés dans le deuxième chapitre (1.2. Une nouvelle conception de la transparence : les programmes d'intercompréhension) considèrent la transparence comme une notion plus vaste, qui peut s'appliquer à un nombre de mots plus important, c'est-à-dire à tous les mots effectivement mis à profit par les apprenants lorsqu'ils cherchent à comprendre un texte. Cette première partie vise donc à présenter une synthèse des travaux portant sur la transparence ; nous serons ensuite en mesure de situer notre réflexion par rapport à ces travaux.

1.1. Une première approche de la notion de transparence

Nous allons, dans ce premier chapitre, présenter des études portant sur la notion de transparence. Nous divisons ces études en deux catégories : celles qui emploient cette notion dans le cadre de leur réflexion, et donc qui précisent selon quelle définition elles l'utilisent, et celles qui se proposent de caractériser précisément la notion, et donc d'apporter une nouvelle définition. Avant de présenter ces deux types d'études, nous allons consacrer un développement à la notion de mot. En effet, il est entendu par transparence « lexicale » transparence au niveau du mot. Avant même de présenter la transparence, il nous faut donc définir ce que nous entendons précisément par ce terme.

1.1.1. La notion de mot

Nous nous concentrerons ici sur le mot tel qu'on peut le définir pour les langues qui nous intéressent, c'est-à-dire le français, l'anglais, l'allemand et le néerlandais.

Ce terme, d'un emploi pourtant extrêmement courant, ne possède pas de définition qui mette tous les linguistes d'accord, et nombreux sont parmi eux ceux qui refusent d'y recourir. Le sens commun consiste à le caractériser comme une suite de lettres entre deux blancs. On sait cependant que ceci, d'une part, ne serait valable qu'à l'écrit, et, d'autre

part, ne résout pas les problèmes posés par les segments du type « l'arbre » (où « l » et « arbre » ne sont pas séparés par un blanc) ou « pomme de terre » (il y a ici trois unités, séparées par deux blancs, mais le tout a dans sa classe de substituts des unités uniques, telles que « poireau » ou « navet »). Le mot ne se laisse pas définir non plus par des critères purement phonétiques, car, si le mot français est oxitonique, l'accent de mot disparaît au profit de l'accent de groupe.

Les critères d'inséparabilité et de commutation (Picoche 1997 : 15, Lehmann & Martin-Berthet 2005 : 183-184, Kahane 2008) semblent permettre de segmenter un énoncé en mots de façon sûre et objective et ainsi de déterminer des unités « à signifié unique et constant » (Benveniste 2004 : 171). Le premier critère consiste à appeler « mot » toute unité à l'intérieur de laquelle on ne peut insérer aucun morphème et le second consiste à appeler « mot » toute unité complexe à laquelle on peut substituer une unité simple. Ces deux critères ne résolvent toutefois pas le problème posé. J. Picoche souligne elle-même qu'une unité peut répondre à un critère et pas à l'autre : « *Faire peur* est commutable en bloc avec *effrayer*, mais le critère d'inséparabilité ne joue pas : *cela me fait très peur*, *affreusement peur*, *une peur affreuse* [...] » (*ibid.* : 16). Ajoutons à l'encontre du critère d'inséparabilité que personne ne songerait à dénier à « insoluble » le statut de mot sous prétexte qu'il existe « indissoluble » dans le lexique du français, et à l'encontre du critère de commutation qu'il est circulaire, puisque, pour définir le mot, il s'agit de pouvoir remplacer un segment par un mot.

Étant donné les difficultés rencontrées pour donner une définition du mot qui serait valable aussi bien pour l'écrit que pour l'oral, les linguistes ont préféré lui substituer successivement plusieurs notions, dont la première fut celle de « morphème ». Rappelons que cette notion, introduite par A. Martinet sous le terme « monème », désigne la plus petite unité linguistique douée de signification⁷. Elle est ainsi la première unité de la première articulation du langage. Le langage s'articule en premier lieu grâce à des éléments significatifs, les monèmes (« **La première articulation** du langage est celle selon laquelle tout fait d'expérience à transmettre, tout besoin qu'on désire faire connaître à autrui s'analysent en une suite d'unités douées chacune d'une forme vocale et d'un sens »

⁷ En le désignant comme « la plus petite unité linguistique douée de signification », nous situons le morphème dans la tradition du monème d'A. Martinet. Il s'agit, par conséquent, d'une notion bien différente de celle déjà nommée « morphème » par J. Vendryes dans son ouvrage *Le langage, introduction linguistique à l'histoire*, publié en 1921, dans lequel il définit les morphèmes comme les « éléments linguistiques [...] qui expriment les rapports entre les idées » (Vendryes 1979 : 92) – par exemple le rôle des morphèmes, dans « Le cheval court », est « le fait que la course associée au cheval en général est rapportée à la troisième personne du singulier de l'indicatif » (*ibid.*).

(Martinet 2008 : 13)), qui sont à leur tour, en second lieu, articulés grâce à des éléments non significatifs, les phonèmes (« Grâce à la seconde articulation, les langues peuvent se contenter de quelques dizaines de production phoniques distinctes que l'on combine pour obtenir la forme vocale des unités de première articulation » (*ibid.* : 15)). Les monèmes sont constitués d'un assemblage de phonèmes, et les énoncés sont constitués d'un assemblage de monèmes. Toutefois, certains monèmes sont constitués d'un seul phonème, « à » par exemple, et certains énoncés d'un seul monème, « Feu ! » par exemple. A. Martinet distingue à l'intérieur de la classe des monèmes les morphèmes d'une part et les lexèmes de l'autre. Les morphèmes appartiennent à une classe fermée (« [ils] alternent, [dans le cadre du syntagme autonome], avec un nombre relativement réduit d'autres monèmes » (*ibid.* : 118)) et les lexèmes appartiennent à une classe ouverte (« [ils] appartiennent à des inventaires illimités » (*ibid.* : 118)). Par exemple, le syntagme autonome « jongleur » se découpe en deux monèmes, « jongl- » et « -eur ». On peut substituer au premier un nombre indéterminé d'autres monèmes – de lexèmes – tels que « chant- », « aviat- », « animat- », etc., alors que l'on peut dresser la liste des monèmes – des morphèmes – substituables au second : « -er », « -age » et « -erie » (et leurs flexions respectives). Le terme de morphème s'est généralisé pour désigner aujourd'hui chez la plupart des auteurs les monèmes d'A. Martinet. On parle dorénavant le plus souvent de morphèmes lexicaux et de morphèmes grammaticaux. Cette terminologie illustre le fait que les premiers portent la charge sémantique tandis que les seconds ont plutôt un rôle syntaxique :

Les morphèmes grammaticaux véhiculent des notions très générales souvent axées sur la situation d'énonciation (relations intersubjectives, temporalité, quantification, détermination, etc.) et qui contrastent avec le plus ou moins grand degré de spécificité du contenu descriptif des morphèmes lexicaux. (Riegel et al. 1998 : 537)

*[Les morphèmes lexicaux] renvoient aux types d'objets, d'êtres, de propriétés, d'événements, d'états et de processus qui catégorisent le monde tel qu'il est conçu par l'homme. Les représentations des choses véhiculées par les [morphèmes lexicaux] se trouvent **actualisées** dans les phrases par les [morphèmes grammaticaux] (affixes flexionnels du genre, du nombre, de la personne, du temps, etc.) et par des morphèmes grammaticaux tels que les déterminants, qui leur confèrent une référence discursive [...]. (*ibid.* : 563)*

Les morphèmes sont donc des unités qui ne sont pas obligatoirement autonomes (« feu » est autonome, « jongl- » et « -eur » ne le sont pas), et il revient aux morphologues d'étudier les différents moyens de combiner les morphèmes entre eux (Corbin 1991, Apothéloz 2002, Tournier 2007).

L'autre notion principale forgée à cause de la difficulté de proposer une segmentation en mots est celle de « lexie ». Elle est définie par B. Pottier comme une « unité fonctionnelle, mémorisée en compétence » (Pottier 1974 : 326). Il s'agit d'une unité autonome, plus souple que le morphème : sa taille peut varier de l'élément simple à un texte entier comme une prière. J. Tournier propose la classification suivante des différents types de lexies : les lexies primaires, les lexies dérivées, les lexies composées, les lexies prépositionnelles et les lexies complexes. Les lexies primaires « comportent un seul élément lexical autonome, sans affixe perçu comme tel. Ce sont des lexies à partir desquelles il n'est pas possible d'induire de mécanismes de formation » (Tournier 2004 : 13). Sont par exemple des lexies primaires « pierre », « rouge », et « armoire » dont la partie « -oire », qui est étymologiquement un suffixe, n'est plus perçue comme tel par les locuteurs. Il s'agit en fait d'unités constituées par un seul morphème autonome. Les lexies dérivées « comportent un seul élément lexical autonome et un ou plusieurs affixes perçus comme tels, qui permettent, par conséquent, d'induire les mécanismes d'affixation » (*ibid.* : 14). Sont par exemple des lexies dérivées « démesure » et « mathématicien ». Les lexies primaires et les lexies dérivées, n'étant composées que d'un seul élément lexical autonome, sont regroupées sous l'appellation « lexies simples ». Les lexies composées « sont formées d'au moins deux éléments lexicaux autonomes et se comportent comme une unité du point de vue de la classe de mots » (*ibid.* : 14). Sont par exemple des lexies composées « arc-en-ciel » et « prendre peur ». Les lexies prépositionnelles « sont constituées d'un syntagme prépositionnel lexicalisé, c'est-à-dire un groupe introduit par une préposition » (*ibid.* : 15). Sont par exemple des lexies prépositionnelles « avec mes sincères salutations » et « sur le feu ». Les lexies complexes sont « toutes les autres lexies, y compris les phrases lexicalisées, comme les proverbes ou certaines citations passées dans l'usage » (*ibid.* : 15). Sont par exemple des lexies complexes « il était une fois » et « que Dieu te bénisse ».

Prenant appui sur cette classification, nous regrouperons dans notre travail sous le terme « mot », qui a l'avantage d'être d'un emploi commun et donc de ne pas freiner la

lecture, les lexies primaires, dérivées et composées. Nous rejoignons ainsi la définition du mot proposée par M. Riegel *et al.* :

*[...] la notion de mot regroupe toutes les unités **préconstruites** (ou **précodées**) que la langue fournit au locuteur pour construire ses énoncés. Ainsi la, et, ovale, petite, billet, table, traîne(r), etc., mais aussi des séquences comme carte grise et chemin de fer sont des unités en quelque sorte préfabriquées, stockées dans notre mémoire lexicale et que nous combinons pour former des phrases selon les besoins de la communication (p.ex. La carte grise et le billet de chemin de fer traînent sur la petite table ovale). (Riegel *et al.* 1998 : 531-532)*

Les notions de morphème et de lexie ont été créées pour pallier la difficulté qu'avaient les lexicologues à donner une définition du mot en termes de segmentation. Une autre notion a été constituée pour tenter de pallier la richesse sémantique du mot : l'« unité lexicale » (Cruse 1986, Fradin 1996). D.A. Cruse la définit ainsi : « A lexical unit is [...] the union of a lexical form and a single sense »⁸ (Cruse 1986 : 77). L'expression « lexical form » est équivalente au lemme, à savoir « the abstract unit of form which is realised in actual sentences as the appropriate member of a set of word forms differing only in respect of inflections »⁹ (*ibid.* : 77), tandis que le critère du « single sense » soulève quelques interrogations. D.A. Cruse admet lui-même qu'on ne peut attribuer une signification et une seule à un mot : « the meaning of any word form is in some sense different in every distinct context in which it occurs »¹⁰ (*ibid.* : 51). Il distingue cependant la « modulation », qui représente l'activation d'un des sens d'une même unité lexicale, de la « sélection contextuelle », qui représente l'activation d'un sens parmi plusieurs unités lexicales en fonction du contexte (p. 52). Les trois tests suivants permettent de décider si l'on a affaire à une ou à plusieurs unités lexicales :

i) s'il existe un synonyme pour une occurrence d'une forme, mais qu'il n'est pas synonyme d'une autre occurrence de la même forme, syntaxiquement identique, dans un autre contexte, alors il s'agit de deux unités lexicales (par ex. : « baie » est synonyme de

⁸ Nous traduisons : « Une unité lexicale est [...] l'union d'une forme lexicale et d'un sens unique ».

⁹ Nous traduisons : « l'unité formelle abstraite qui se réalise effectivement dans des phrases en tant que membre approprié d'un ensemble de formes différant uniquement du point de vue de la flexion ».

¹⁰ Nous traduisons : « le sens de n'importe quelle forme est, d'une certaine manière, différent dans chacun des contextes dans lesquels elle apparaît ».

« vitre » dans « Ne t'appuie pas sur la baie » mais pas dans « Ces baies font des taches difficiles à enlever »).

ii) s'il existe un mot entrant dans une relation d'opposition avec une occurrence d'une forme, mais qu'il n'est pas en relation d'opposition avec une autre occurrence de la même forme, syntaxiquement identique, dans un autre contexte, alors il s'agit de deux unités lexicales (par ex. : « jet » est en relation d'opposition avec « barque » dans « Son nouveau jet lui a coûté une fortune » mais pas dans « On a installé un nouveau jet d'eau dans le jardin »).

iii) s'il existe un mot qui est de la même famille qu'une occurrence d'une forme, mais qu'il n'est pas de la même famille qu'une autre occurrence de la même forme, syntaxiquement identique, dans un autre contexte, alors il s'agit de deux unités lexicales (par ex. : « location » est de la même famille que « louer » dans « Désirez-vous louer ou acheter ? » mais pas dans « Il n'a cessé de louer ta prudence »).

P. Bogaards conclut des exemples suivants (d'après Kleiber 1990) :

- *Ce livre se trouve dans toutes les bonnes librairies.*
- *Ce livre a été tiré à 15 000 exemplaires.*
- *Jean est parti à la campagne pour écrire un livre.*
- *Ce livre a été un fiasco pour son éditeur.*
- *Ce livre est ennuyeux.*
- *Ce livre est sale et déchiré.* (Bogaards 1994 : 21)

que la forme substantive « livre » constitue une seule unité lexicale, parce que « [t]outes ces différences de sens sont à considérer comme de simples variations contextuelles et aucun locuteur ou apprenant du français ayant compris ce qu'est un livre, n'aura de difficulté à comprendre les exemples donnés » (*ibid.* : 21). Or, si l'on peut remplacer « livre » par « conte » dans « Ce livre est ennuyeux », on ne le peut dans « Ce livre est sale et déchiré ». Le critère du « sens unique » paraît ainsi difficile à cerner, et nous continuerons donc d'employer le terme de « mot », tel que nous l'avons défini ci-dessus.

Les mots sont ainsi des unités éventuellement polysémiques. Il n'appartient pas à cette étude de différencier polysémie et homonymie, les critères devant fonder cette distinction n'ayant pas été formellement établis. La tradition veut, depuis M. Bréal, que la polysémie rende compte d'un lien étymologique entre les différentes significations d'un mot :

Le même terme peut s'employer tour à tour au sens propre ou au sens métaphorique, au sens restreint ou au sens étendu, au sens abstrait ou au sens concret... A mesure qu'une signification nouvelle est donnée au mot, il a l'air de se multiplier et de produire des exemplaires nouveaux, semblables de forme, mais différents de valeur.

Nous appellerons ce phénomène de multiplication la polysémie. (Bréal 1976 : 143-144)

L'homonymie, quant à elle, rendrait compte de mots de même forme (homographie et/ou homophonie) mais ayant des étymons différents. Cependant, certains mots de même forme et d'origine commune doivent être nommés homonymes, en raison de leur trop grande distance sémantique (« grève : terrain plat situé au bord de la mer » et « grève : cessation volontaire et collective du travail » par exemple). La distinction synchronique traditionnelle consiste à dire, par exemple selon les termes de B. Pottier, que la polysémie est la « relation entre des substances de signifié se recouvrant partiellement, sous un signifiant unique » (Pottier 1974 : 329) et l'homonymie la « relation entre deux signes ayant un même signifiant, et deux substances de signifié totalement disjointes » (*ibid.* : 325). Selon J. Picoche, la polysémie « consiste en ce que les emplois d'un signifiant donné, tout en reposant sur un certain contenu sémique commun, se ramifient, par le jeu des contextes, en un certain nombre d'acceptions ». Elle ajoute que ces acceptions sont « parfois si diverses que le rapport de base peut devenir pratiquement imperceptible à l'utilisateur dans l'exercice normal de son langage, c'est-à-dire quand il ne porte pas une attention particulière aux mots qu'il emploie » (Picoche 1997 : 73). Il apparaît donc que différencier la polysémie de l'homonymie en synchronie est affaire délicate. Nous renvoyons à la littérature spécialisée pour des études détaillées (Victorri & Fuchs 1996 pour une réflexion linguistique sur la polysémie, utilisée pour l'élaboration d'un modèle mathématique, et une implémentation informatique de ce modèle ; Kleiber 1999, 2008 pour une réflexion sur les propriétés sémantiques des mots polysémiques ; Soutet 2005 pour un ensemble d'études recourant à la notion de polysémie), et réfléchissons dans ce travail à la transparence au niveau du mot, en tant qu'entité éventuellement polysémique, donc possédant plusieurs acceptions, et nous laisserons aux lexicographes de nos dictionnaires de référence le choix de différencier les polysèmes des homonymes selon les critères qui leur semblent appropriés.

La notion de mot étant suffisamment précisée pour l'usage qui en sera fait dans la présente étude, nous pouvons maintenant nous pencher sur les différentes définitions qui ont été données de la transparence lexicale.

1.1.2. Les définitions traditionnelles

La terminologie employée pour désigner la transparence au niveau du mot est variée. Certaines appellations renvoient à des notions légèrement différentes, d'autres se recouvrent. Les principales sont « mots transparents »¹¹, « vrais ou bons amis »¹², « cognats » (version francisée de l'anglais *cognates*)¹³ et « congénères »¹⁴. Les deux premières renvoient à une transparence (dont nous allons voir les définitions ci-après) en synchronie, tandis que « cognats » et « congénères » renvoient à une transparence qui trouve sa raison dans une étymologie commune. Néanmoins, l'emploi de « cognats » s'est généralisé pour désigner également soit des mots qui n'ont pas d'étymon commun mais dont l'un est un emprunt au lexique de l'autre, soit des mots qui n'ont pas nécessairement de relation historique (« cognats » est dans ce dernier cas équivalent de « mots transparents » et « vrais amis »).

Quelle que soit l'appellation choisie pour nommer les paires de mots concernés, la transparence est traditionnellement définie, par les auteurs faisant appel à cette notion pour leur étude, comme une (quasi-)identité de forme et de signification. Nous n'avons pas relevé de définition prenant en compte le troisième pôle de caractérisation du mot selon A. Meillet, à savoir sa catégorie grammaticale. L'auteur avait en effet mis l'accent sur le fait qu'un mot devait être conçu comme une association de trois facteurs : « Un mot résulte de l'association d'un sens donné à un ensemble de sons donnés susceptible d'un emploi grammatical donné » (Meillet¹⁵, cité dans Lehmann & Martin-Berthet 2005 : 1). Nous faisons l'hypothèse que, pour ces auteurs faisant appel à la notion de transparence, l'identité de signification inclut, en la sous-entendant, l'identité de catégorie grammaticale, c'est-à-dire que deux mots ne peuvent avoir la même signification que s'ils appartiennent à la même catégorie.

¹¹ Valli 2001, Castagne 2003, Rivenc 2003

¹² Bertrand & Lévy 1999, Walter 2001b

¹³ Bogaards 1994, Caid 2008

¹⁴ Séguin 1994, Degache & Masperi 1998

¹⁵ A. Meillet, *Linguistique historique et linguistique générale*, Champion, 1921, p. 30.

Les citations suivantes illustrent la définition de la transparence comme une identité ou une quasi-identité de forme et de signification entre plusieurs mots :

*[...] cognates are pairs of tokens of different languages which share "obvious" phonological or orthographic and semantic properties [...].*¹⁶ (Simard et al. 1992 : 71)

*Les cognates sont généralement définis comme des **mots de langues différentes ayant (à peu près) les mêmes formes et (à peu près) les mêmes sens**, comme régulier français et regular anglais.* (Bogaards 1994 : 153)

*In the narrow sense used in historical linguistics, cognates are words in related languages that have developed from the same ancestor word. An example of a cognate pair is French lait and Spanish leche, both of which come from Latin lacte. In other contexts, including this paper, the term is often used more loosely, denoting words in different languages that are similar in form and meaning, without making a distinction between borrowed and genetically related words; for example, English sprint and the Japanese borrowing supurinto are considered cognate, even though these two languages are unrelated.*¹⁷ (Kondrak 2001 : 103)

Nous pouvons [...] définir la transparence comme l'intelligibilité immédiate d'un mot inconnu de la langue L2 en raison d'une identité morpho-sémantique (partielle) avec un mot connu (ou plusieurs mots connus) de la langue L2 ou d'autres langues (L1, L3, L4...). (Hausmann 2002)

La transparence dans le cas de deux langues voisines, repose sur la proximité des signifiants : elle va de la coïncidence morpho-sémantique totale à la coïncidence

¹⁶ Nous traduisons : « Les cognats sont des paires de chaînes appartenant à des langues différentes, qui partagent « manifestement » des propriétés phonologiques ou orthographiques et sémantiques ».

¹⁷ Nous traduisons : « Selon le sens restreint utilisé en linguistique historique, les cognats sont des mots appartenant à des langues apparentées qui ont évolué à partir du même étymon. Un exemple d'une paire de cognats est le français *lait* et l'espagnol *leche*, les deux venant du latin *lacte*. Dans d'autres contextes, et notamment dans ce papier, le terme est souvent utilisé de façon plus souple et renvoie à des mots de langues différentes qui sont similaires des points de vue de la forme et de la signification, sans qu'il y ait de distinction entre les mots empruntés et ceux qui sont reliés génétiquement ; par exemple, l'anglais *sprint* et l'emprunt japonais *supurinto* sont considérés comme des cognats, même si ces deux langues ne sont pas apparentées ».

morpho-sémantique partielle, qui est de loin la plus fréquente [...]. (Hédiard 2003 : 167)

Cognates : similitude entre deux (ou plusieurs) signifiants de langues différentes et similitude partielle au niveau du signifié. (Robert 2004a : 508)

On s'étonnera, relativement à la définition de M. Hédiard, que la proximité des *signifiants* manifeste une coïncidence *morpho-sémantique*. Au vu de ces quelques définitions, la conclusion tirée par J. Murillo Puyal *et al.* semble justifiée :

*[...] la notion de « transparence », souvent invoquée pour rendre compte des phénomènes d'intercompréhension spontanée entre locuteurs de langues romanes distinctes et, d'une manière générale, entre locuteurs de langues distinctes mais typologiquement et/ou génétiquement apparentées, est par trop impressionniste et imprécise, et la référence à des étymons communs, dont le « locuteur moyen » actuel n'a en général ni conscience ni connaissance, n'est pas susceptible d'applications didactiques. (Murillo Puyal *et al.* 2008 : 343)*

En effet, des expressions du type « similitude partielle » ou « (à peu près) les mêmes formes » laissent libre voie à des interprétations diverses et subjectives, et ne permettent pas de tracer la frontière entre les mots transparents et ceux qui ne le sont pas.

1.1.3. Tentatives de caractérisation précise

Certaines études sont expressément consacrées à une caractérisation précise de la transparence lexicale.

L'une d'elles est présentée par M.-C. Tréville, dans son ouvrage *Rôle des congénères interlinguaux dans le développement du vocabulaire réceptif en français langue seconde*, paru en 1993. Elle y expose les travaux de P. Hammer et M.-J. Monod, qui ont rédigé un dictionnaire, qui n'a pas été publié, le *English-French Cognate*

*Dictionary*¹⁸. Pour recenser les congénères français-anglais, ces auteurs ont travaillé à partir du *Dictionnaire moderne français-anglais* des éditions Larousse de 1960, et ont sélectionné les mots d'après les critères morpho-sémantiques suivants :

Sont considérés de sens similaire « items that have the same meaning in both languages in ordinary use » (p.iv). Pour être considérés de forme similaire, les mots doivent remplir une des conditions suivantes : soit leur graphie en anglais et en français diffère d'une seule lettre (oncle/uncle, avantage/advantage), soit leur terminaison entière diffère selon des règles généralisables (French -té = English -ty, French -eux = English -ous, etc.). (Tréville 1993 : 71)

La tentative de caractérisation manque de précision : la définition de la similarité sémantique est tautologique, et les règles définissant la similarité formelle ne sont pas justifiées. Non seulement le choix de la différence possible d'une seule lettre est arbitraire, mais il empêche également, comme le signale M.-C. Tréville, de considérer comme proches de mots français des mots anglais tels que *government* et *campain* (sic) (*ibid.* : 71).

C. Haeusser et D. Hirst décrivent, dans leur court article « Transparence/opacité : contribution à une étude de la contrastivité lexicale français-anglais, dans une perspective didactique », l'état d'une recherche qui a dû être interrompue prématurément. La transparence formelle y est définie par « le nombre, l'ordre et le pourcentage de graphèmes communs à deux unités » (Haeusser & Hirst 1987 : 280). Il n'est malheureusement pas fait référence aux raisons qui ont présidé au choix de la valeur de ce pourcentage, censé déterminer le « seuil de transparence » (*ibid.* : 280). La transparence sémantique est définie par « le nombre de sens symétriques et leur poids relatif dans la configuration sémantique de chacune des paires L1/L2 » (*ibid.* : 281). Manquent ici aussi les informations de détail, sur la détermination des configurations sémantiques et la constitution des paires.

R. LeBlanc et H. Séguin se sont également intéressés au couple de langues français/anglais, et en ont dressé la liste des congénères. Pour ce faire, ils ont dépouillé le *Robert-Collins* de 1978 et le *Larousse-Saturne* de 1981, et en ont extrait les 6 447 mots homographes entre l'anglais et le français, et les 16 713 paragraphes. Ces derniers ont été

¹⁸ Hammer P. & Monod M.-J., *English-French Cognate Dictionary*, Edmonton : The University of Alberta, 1976

répertoriés « au jugé » (Séguin 1994 : 196). Ce n'est qu'une fois que les paires de mots ont été subjectivement sélectionnées que les auteurs ont dressé la liste des règles de correspondance. La similitude formelle n'est donc pas, dans cette étude, définie. C'est l'intuition des auteurs qui est caractérisée après-coup.

La caractérisation des cognats par M. Simard *et al.*, dans leur fameux article de 1992 « Using cognates to align sentences in bilingual corpora », soulève également quelques problèmes. Sont considérés par eux comme cognats, outre les signes de ponctuation identiques et les chaînes identiques contenant au moins un chiffre, les paires de mots composés d'au moins quatre lettres, et dont les quatre premières sont identiques. Les auteurs ne précisent pas pour quelles raisons les quatre mêmes premières lettres suffisent à faire de deux mots des cognats ; de plus, avec cette définition, les mots de moins de quatre lettres, et ceux du type de *government*/« gouvernement » ne peuvent pas être considérés comme cognats. Enfin, seul le versant formel est pris en compte pour l'attribution de la qualité « cognat ».

Ce sont précisément les méthodes évaluant la similarité formelle qui ont été le plus développées. Les principales sont la distance de Levenshtein, le coefficient de Dice, le coefficient de sous-chaîne maximale et les distances de Jaro puis Jaro-Winkler.

La distance de Levenshtein permet de calculer le coût minimal pour passer d'une chaîne de caractères à une autre, en termes d'insertion, de suppression et de substitution de caractères :

Assume afternoon is pronounced as [ˈæftəˌnɪm] in the dialect of Savannah, Georgia, and as [ˌæftərˈhuːn] in the dialect of Lancaster, Pennsylvania. Changing one pronunciation into the other can be done as follows (ignoring suprasegmentals and diacritics for this moment):

<i>æftən#n</i>	<i>delete ə</i>	<i>1</i>
<i>æftən#n</i>	<i>insert r</i>	<i>1</i>
<i>æftər#n</i>	<i>subst. #/u</i>	<i>1</i>
<i>æftərnun</i>		

3

(Gooskens & Heeringa 2004 : 196)¹⁹

Pour éviter que les mots les plus longs aient une distance plus importante que les mots plus courts en raison du nombre probablement plus grand d'opérations à faire, et ce alors qu'ils pourraient sembler plus proches, la somme des opérations est divisée par la longueur du plus long alignement, c'est-à-dire celui qui permet le plus grand nombre de correspondances. Appliqué à l'exemple ci-dessus :

<i>æ</i>	<i>ə</i>	<i>f</i>	<i>t</i>	<i>ə</i>		<i>n</i>	<i>#</i>	<i>n</i>
<i>æ</i>		<i>f</i>	<i>t</i>	<i>ə</i>	<i>r</i>	<i>n</i>	<i>u</i>	<i>n</i>
	<i>1</i>				<i>1</i>		<i>1</i>	

The total cost of 3 (1+1+1) is now divided by the length of 9. This gives a word distance of 0.33 or 33%.²⁰ (ibid. : 196)

Dans d'autres travaux (Kessler 1995, Nerbonne *et al.* 1999), toutes les opérations n'ont pas le même coût (la substitution d'une voyelle à une autre voyelle coûte par exemple moins que la substitution d'une voyelle à une consonne).

Le coefficient de Dice²¹ s'appuie sur le nombre de caractères que deux chaînes ont en commun. Il s'agit de diviser le nombre de caractères en commun multiplié par deux par le nombre de caractères total. Le résultat va de 0 (similarité nulle) à 1 (similarité parfaite). Par exemple, le coefficient de Dice appliqué à la paire *actually/actuellement* est :

¹⁹ Nous traduisons : « Partons du principe que *afternoon* est prononcé [ˈæftə,nʌn] dans le dialecte de Savannah en Géorgie, et [ˌæftərˈhʌn] dans le dialecte de Lancaster en Pennsylvanie. Le fait de passer d'une prononciation à une autre peut être représenté ainsi (en ne tenant pas compte pour le moment des éléments suprasegmentaux et des diacritiques) ».

²⁰ Nous traduisons : « Le coût total de 3 (1+1+1) est maintenant divisé par la longueur de 9. Cela donne une distance de mots de 0,33 ou 33 % ».

²¹ Il fut pour la première fois utilisé pour la comparaison de mots par A.M. McEnery et M.P. Oakes dans leur article de 1995, « Sentence and word alignment in the CRATER project : methods and assessment » (*Proceedings of the EACL-SIGDAT Workshop*, Dublin).

$$\frac{6 \text{ (nombre de caractères en commun)} \times 2}{20 \text{ (nombre de caractères total)}} = 12/20 = 0,6$$

Le coefficient de Dice ne prend toutefois pas en compte le problème posé par les anagrammes. Par exemple, « migraine » et « imaginer » ont un coefficient de $(8 \times 2)/16 = 1$. La similarité parfaite selon ce coefficient n'est donc manifestement pas la même que celle des chercheurs s'intéressant à la transparence lexicale.

Pour pallier cette difficulté, ainsi que celle créée par la méthode consistant à comparer deux mots par le nombre de lettres identiques contiguës, la notion de « sous-chaîne maximale » (SCM) a été introduite : « Par exemple, pour « docteur » et « dottore », la SCM est de longueur 4 : **d-o-t-r** » (Kraif 1999). Il est possible, à partir de cette notion, de calculer le coefficient de sous-chaîne maximale, soit le *Longest Common Subsequence Ratio* (LCSR) présenté par I.D. Melamed dans son article « Bitext maps and alignment via pattern recognition » :

The LCSR of two tokens is the ratio of the length of their longest (not necessarily contiguous) common subsequence (LCS) and the length of the longer token. In symbols,

$$\text{LCSR}(A, B) = \frac{\text{length}[\text{LCS}(A, B)]}{\max[\text{length}(A), \text{length}(B)]}$$

For example, gouvernement, which is 12 characters long, has 10 characters that appear in the same order in government. So, the LCSR for these two words is 10/12. On the other hand, the LCSR for conseil and conservative is only 6/12.²² (Melamed 1999 : 113)

²² Nous traduisons : « Le coefficient de sous-chaîne maximale de deux chaînes est le quotient de la longueur de leur plus longue sous-chaîne maximale (pas nécessairement contiguë) par la longueur de la plus longue chaîne. Sous forme de symboles, [...] Par exemple, *gouvernement*, qui a 12 caractères, possède 10 caractères qui apparaissent dans le même ordre que dans *government*. Ainsi, le coefficient de sous-chaîne maximale de ces deux mots est de 10/12. Cependant, le coefficient de sous-chaîne maximale de *conseil* et *conservative* est seulement de 6/12 ».

Les deux dernières méthodes que nous allons aborder sont les distances de Jaro et de Jaro-Winkler, la dernière étant une variante de la première. Comme le coefficient de Dice, elles permettent de calculer la similarité entre deux chaînes de caractères, en donnant des résultats s'échelonnant de 0 à 1. La distance de Jaro, notée Φ , entre deux chaînes s_1 et s_2 , de longueur respective str_len1 et str_len2 , est :

$$\Phi (s_1,s_2) = 1/3 [\#common/str_len1 + \#common/str_len2 + (\#common - \#transpositions)/\#common] \text{ (adapté de Porter \& Winkler 1999 : 191)}$$

Le nombre de caractères correspondants est noté par $\#common$. Pour que deux caractères identiques soient correspondants, il faut qu'ils apparaissent à la même position dans les deux chaînes, ou que leur éloignement ne dépasse pas la moitié de la longueur de la plus courte chaîne. Le nombre de transpositions est noté par $\#transpositions$. Il y a demi-transposition quand deux caractères identiques correspondants ne sont pas à la même position dans les deux chaînes. La distance de Jaro-Winkler reprend l'équation ci-dessus, mais propose une amélioration si les deux chaînes commencent par les mêmes caractères, au nombre maximum de quatre. Si le nombre de premiers caractères identiques est noté i (égal à 1, 2, 3 ou 4), alors la distance Jaro-Winkler est égale à :

$$\Phi + i \times 0,1 \times (1 - \Phi) \text{ (Winkler 1990 : 356)}$$

Illustrons cette formule par un exemple²³. Soient deux chaînes s_1 *MARTHA* et s_2 *MARHTA*. La table de correspondances est :

	M	A	R	T	H	A
M	1	0	0	0	0	0
A	0	1	0	0	0	0
R	0	0	1	0	0	0
H	0	0	0	0	1	0
T	0	0	0	1	0	0
A	0	0	0	0	0	1

²³ Nous nous inspirons pour cela de la page http://fr.wikipedia.org/wiki/Distance_de_Jaro-Winkler (consultée le 22/06/09).

- $\#common = 6$ (nombre de 1 dans la table)
- $str_len1 = 6$
- $str_len2 = 6$
- Les caractères correspondants sont $\{M,A,R,T,H,A\}$ pour $s1$ et $\{M,A,R,H,T,A\}$ pour $s2$. En considérant ces ensembles ordonnés, on a donc 2 couples (T/H et H/T) de caractères correspondants différents, soit deux demi-transpositions.

La distance de Jaro est :

$$\Phi = 1/3 (6/6 + 6/6 + (6 - 1)/6) = 0,944$$

La distance de Jaro-Winkler avec $i = 3$ est égale à :

$$0,944 + [3 \times 0,1 (1 - 0,944)] = 0,961$$

La distance de Levenshtein, le coefficient de Dice, le coefficient de sous-chaîne maximale et les distances de Jaro puis Jaro-Winkler sont des méthodes qui rendent toutes compte du degré de similarité formelle entre deux chaînes de caractères (symboles API ou lettres). Elles ne nous paraissent pas exploitables pour la caractérisation des mots transparents, d'une part parce qu'elles ne s'attachent qu'à la forme des mots (le mot allemand *Baum* « arbre » ne saurait être considéré comme transparent avec le mot français « baume » même si leur coefficient de sous-chaîne maximale est égal à 0,8), et d'autre part parce que, pour précises qu'elles soient, elles ne permettent pas de savoir à partir de quand un mot va pouvoir être reconnu par un lecteur. On ne peut remettre en cause le fait qu'un mot dont une de ces méthodes a mis en évidence un taux de similarité formelle important par rapport à un mot de la langue d'un lecteur va vraisemblablement être facilement identifié par lui. Cependant, si l'on cherche à déterminer le nombre de mots transparents dans une langue pour un lecteur d'une autre langue, on ne peut utiliser aucune de ces méthodes. Pour pouvoir les utiliser dans ce but, il faudrait fixer un seuil à partir duquel un mot serait considéré comme transparent (par exemple, il faudrait qu'il ait une distance de Jaro-Winkler au moins égale à 0,75). Il paraît néanmoins difficile d'établir et de justifier un tel seuil. Il n'a d'ailleurs, à notre connaissance, jamais été publié de recherches sur ce point.

Nous venons de présenter, dans ce chapitre, différentes études portant sur la transparence lexicale. Les unes utilisent cette notion et d'autres cherchent à la définir. Aucune des définitions proposées ne nous paraît exploitable en fonction des objectifs que nous nous fixons, notamment en raison de la subjectivité des critères alloués et du recours

à l'étymologie, dont l'apprenant n'a pas connaissance, pour l'établissement de la similarité sémantique. Nous établirons au début de la deuxième partie de notre travail les critères que nous considérons comme primordiaux pour une caractérisation de la transparence (section 2.1.1. De nouveaux objectifs).

Dans les définitions que nous venons d'examiner, la transparence est considérée comme un phénomène binaire : un mot d'un lexique donné est soit transparent soit opaque vis-à-vis d'un mot appartenant à un autre lexique. Soit ils sont (quasi-)identiques des points de vue formel et sémantique, soit ils sont dissemblables. Des propositions plus fines ont été faites par des équipes travaillant sur l'intercompréhension, dans lesquelles est introduit un niveau intermédiaire, constitué des mots ni suffisamment proches de mots de la langue maternelle du lecteur pour être immédiatement compris, ni suffisamment différents pour être dits opaques. Nous exposerons ces études dans la section 1.2.3. Avant cela, il nous faut consacrer un chapitre à la notion d'intercompréhension, à ses principes, et aux différents programmes qui en ont fait leur objet d'étude.

1.2. Une nouvelle conception de la transparence : les programmes d'intercompréhension

Nous commencerons par situer l'intercompréhension par rapport au concept de multilinguisme, puis nous concentrerons sur les différentes formes que prend la formation à l'intercompréhension. Nous verrons enfin que des définitions plus fines que celles exposées précédemment de la transparence ont été proposées dans le cadre de programmes formant à l'intercompréhension.

1.2.1. Le multilinguisme et l'intercompréhension

Il convient de commencer par exposer ce que l'on entend par « multilinguisme ». Nous le définissons par comparaison au « plurilinguisme ». Bien que ces deux termes soient définis différemment selon les auteurs qui les utilisent ou qu'ils soient parfois employés l'un pour l'autre²⁴, nous choisissons de leur assigner les significations les plus fréquentes, exposées par C. Hagège dans son article « Le plurilinguisme européen », paru en 2004 :

[...] l'usage actuellement répandu est d'appeler « plurilinguisme » la faculté que possède un individu ou une collectivité de parler deux ou plusieurs langues, et d'en avoir, dans le meilleur des cas, une connaissance qui est celle de locuteurs de naissance ; la notion de « multilinguisme », quant à elle, est appliquée, à la situation d'états ou de régions où coexistent deux ou plusieurs communautés usagères de langues différentes. (Hagège 2004 : 35)

Comme le précise l'auteur, « [l]es deux situations peuvent évidemment coïncider, mais ce n'est pas là une nécessité » (*ibid.* : 35). Par exemple, le Québec est une zone multilingue, où sont parlés l'anglais et le français, et ses habitants sont plurilingues : ils maîtrisent les deux langues. Au contraire, la Belgique est un pays multilingue, où sont parlés le flamand et le français (ainsi que l'allemand dans une petite zone de l'est du

²⁴ Voir également Truchot 1994 : 21, Degache 2006 : 10 et Hélot 2007 : 26.

pays) ; la tension linguistique régnant en Belgique entraîne un attachement fort des habitants à leur langue, si bien que la majorité d'entre eux affirment leur monolinguisme (Mackey 1976 : 46-47, 76-78).

Face à une situation de multilinguisme, plusieurs réponses peuvent être apportées pour que les locuteurs puissent communiquer entre eux :

- les locuteurs d'une des langues peuvent apprendre à parler l'autre langue (dans l'hypothèse où seules deux langues sont parlées), de sorte que tous communiquent en faisant usage de cette dernière. Ce fut par exemple le cas lorsque l'Angleterre fut envahie par les Français après la bataille d'Hastings en 1066 : pour appartenir à la classe dirigeante, il fallait parler le français ;

- les locuteurs peuvent tous apprendre une autre langue étrangère, par le biais de laquelle ils communiquent entre eux. C'est actuellement le cas en Europe, puisque la majorité des élèves choisissent l'anglais comme langue vivante (Graddol 2006 : 82-103) – bien qu'il ne s'agisse pas toujours d'un choix : les jeunes Français sont dans l'obligation d'avoir l'anglais parmi leurs langues vivantes ;

- les locuteurs peuvent se créer une nouvelle langue pour les besoins de la communication. Ainsi la *lingua franca* s'est-elle développée au Moyen-Âge. Langue véhiculaire s'appuyant sur plusieurs langues déjà existantes, elle a permis particulièrement aux marchands, ainsi qu'aux autres populations voyageant dans le bassin méditerranéen, d'échanger (Kappler & Méjean-Thiolier 2009 : 20). Sur un autre plan, l'espéranto, langue créée à la fin du XIX^{ème} siècle, est également un représentant de ce type de solutions apportées au multilinguisme²⁵ ;

- enfin, les locuteurs d'une langue peuvent continuer à parler leur propre langue tout en comprenant celle des locuteurs de l'autre langue. Cette dernière situation est désignée par le terme d'« intercompréhension ». Les locuteurs « s'intercomprennent », et n'ont pas besoin d'apprendre à parler une langue étrangère. C'est le cas par exemple en Scandinavie. C. Hagège affirme en effet que « la parenté du danois, du norvégien et du suédois est assez étroite pour permettre une communication presque normale entre habitants des trois pays correspondants, lors même que chaque participant parle et répond dans sa langue » (Hagège 1996a : 250). Les quelque 85 à 90 % des mots qui ont, dans ces trois langues, « une forme et un sens largement semblables » (DGLFLF 2006 : 5) permettent de parler de

²⁵ Voir le site Internet <http://esperanto-france.org/>.

« Scandinavian language community » (Börestam Uhlmann 2002). C'est également le cas par exemple en Inde, où, bien que les 1 600 langues qui y sont parlées n'appartiennent pas toutes à la même famille, « nombre d'Indiens maîtrisent couramment trois ou quatre langues indiennes » et où « des films produits dans une langue sont appréciés par les locuteurs d'une autre » (Kichenassamy 2004 : 45).

Nous allons maintenant nous intéresser à la situation linguistique en Europe, zone multilingue.

L'Union européenne compte actuellement 23 langues officielles, qui sont, classées par ordre alphabétique, les suivantes : l'allemand, l'anglais, le bulgare, le danois, l'espagnol, l'estonien, le finnois, le français, le grec, le néerlandais, le hongrois, l'irlandais, l'italien, le letton, le lituanien, le maltais, le polonais, le portugais, le roumain, le slovaque, le slovène, le suédois et le tchèque. En prenant en compte les « langues de minorités nationales ou régionales, [...] nous relevons en gros 60 à 70 langues en Europe » (Gogolin 2002 : 7).

Loin de choisir la solution du « tout-anglais », ou d'élire quelques langues comme vecteurs principaux de l'information, l'Europe a au contraire adopté une politique résolument axée sur le multilinguisme. Il est noté dans le texte intitulé « Un nouveau cadre stratégique pour le multilinguisme » publié par la Commission des communautés européennes que « le respect de la diversité linguistique constitue une valeur fondamentale de l'Union européenne » (COM 2005 : 4). Ainsi, selon l'article 22 de la Charte des droits fondamentaux de l'Union européenne, l'Union s'engage à respecter la diversité culturelle, religieuse et linguistique, et l'article 21 interdit toute discrimination fondée sur divers motifs, notamment la langue. De même, le traité d'Amsterdam prévoit « que tout citoyen peut s'adresser aux institutions de l'Union dans sa propre langue, et recevoir une réponse dans la même langue » (Herbillon 2003). La Commission précise à quel point il lui paraît nécessaire de recourir à de multiples langues :

La capacité de comprendre et d'utiliser plusieurs langues, qui est déjà une réalité quotidienne pour une majorité de personnes dans le monde, est une aptitude souhaitable pour tous les citoyens européens. Elle nous encourage à plus d'ouverture envers les autres, leurs cultures et leurs valeurs, améliore les capacités cognitives et renforce les compétences des apprenants dans leur langue maternelle [...] (COM 2005 : 4)

Le gain économique et politique est également noté :

Une politique du multilinguisme réussie peut élargir les perspectives offertes aux citoyens : elle peut augmenter leur employabilité, faciliter leur accès aux services et l'exercice de leurs droits et contribuer à la solidarité par un renforcement du dialogue interculturel et de la cohésion sociale. (COM 2008 : 3)²⁶

Pour concrétiser son engagement en faveur du multilinguisme et du plurilinguisme, l'Union a notamment créé les instances que sont la Division des politiques linguistiques du Conseil de l'Europe, la Commission européenne chargée du multilinguisme, le Centre européen pour les langues vivantes et l'Observatoire européen du plurilinguisme. Chacune d'entre elles veille au respect des langues vivantes d'Europe. Des dispositions ont également été adoptées dans ce sens, telles que la décision prise en 2002 par les chefs d'États et de gouvernements réunis à Barcelone en faveur de l'enseignement d'au moins deux langues étrangères dès le plus jeune âge pour les élèves de l'UE, ou bien la *Charte européenne des langues régionales ou minoritaires*, signée en 1992, visant à protéger et promouvoir ces langues.

Deux des principales entreprises du Conseil de l'Europe pour la construction d'une Europe unie dans la diversité de ses langues furent la création du Cadre Européen Commun de Référence pour les langues (dorénavant noté CECR) et du Portfolio européen des langues (dorénavant noté PEL).

Le CECR, publié en France en 2001, fournit un référentiel commun pour l'évaluation du niveau en langue étrangère. Il « offre une base commune pour l'élaboration de programmes de langues vivantes, de référentiels, d'examens, de manuels, etc. en Europe » (Conseil de l'Europe 2001 : 9). Il propose trois niveaux (A, B, C), chacun étant subdivisé en deux (A1, A2, B1, B2, C1, C2). Le niveau A est celui de l'utilisateur élémentaire, le niveau B celui de l'utilisateur indépendant et le niveau C celui de l'utilisateur expérimenté. Pour chaque niveau, un descriptif détaille ce que l'apprenant doit savoir faire. La grande spécificité du CECR est qu'il donne une description des connaissances pour chaque compétence langagière : comprendre (écouter et lire), parler (prendre part à une conversation et s'exprimer oralement en continu) et écrire. Ainsi, les apprenants peuvent se situer sur une échelle allant de A1 (niveau introductif ou

²⁶ Notons que, dans cette citation, les auteurs ont donné au mot « multilinguisme » le sens que nous avons donné à « plurilinguisme », c'est-à-dire la capacité pour un même individu de parler plusieurs langues.

découverte) à C2 (maîtrise) pour chacune de leurs compétences, et reconnaître par exemple qu'ils maîtrisent bien la lecture, même si l'expression en continu leur pose encore problème :

De ce point de vue, le but de l'enseignement des langues se trouve profondément modifié. Il ne s'agit plus simplement d'acquérir la « maîtrise » d'une, deux, voire même trois langues, chacune de son côté, avec le « locuteur natif idéal » comme ultime modèle. Le but est de développer un répertoire langagier dans lequel toutes les capacités linguistiques trouvent leur place. (ibid. : 11)

Le PEL, lancé en 2001, est un document individuel, que chacun peut se constituer. Il permet de faire le bilan de ses connaissances en langues, grâce aux niveaux du CECR. Il est composé de trois parties : un passeport des langues, une biographie langagière et un dossier. Le passeport donne une vue d'ensemble des capacités de l'apprenant dans différentes langues, de ses certifications et diplômes, le tout ayant été acquis en contexte scolaire ou extra-scolaire. La biographie permet d'impliquer l'apprenant dans son apprentissage. Il y dresse lui-même ses compétences, ses expériences et ses prochains buts. Enfin, le dossier permet à l'apprenant de fournir des documents, manuscrits ou électroniques, attestant de ses capacités.

L'Europe est, nous venons de le voir, pleinement reconnue en tant que territoire multilingue. Pour que ses habitants puissent communiquer entre eux, c'est-à-dire pour qu'ils développent une compétence plurilingue, nous avons évoqué plus haut les solutions que constituent le recours à une langue unique (qu'elle soit une des langues du territoire ou une langue tierce) d'une part et la mise à profit des compétences de compréhension de l'autre. Nous allons maintenant voir que l'intercompréhension se révèle être une réponse parfaitement adaptée aux objectifs d'égalité des peuples et de maintien de la diversité que s'est fixés l'Union européenne.

L'intercompréhension est définie, par les auteurs qui utilisent cette notion, comme la capacité par des locuteurs de communiquer entre eux en *comprenant* la langue de l'autre et non en l'*utilisant*. Voici quelques définitions qui ont été proposées :

Par le terme intercompréhension, on désigne la capacité de comprendre une langue étrangère sur la base d'une autre langue sans l'avoir apprise. (Meissner et al. 2004 : 20)

L'intercompréhension est une forme de communication dans laquelle chaque personne s'exprime dans sa propre langue et comprend celle de l'autre.

Cette définition présente l'avantage de clarifier deux points : l'intercompréhension inclut la communication tant orale qu'écrite, et exclut l'utilisation active de la langue étrangère. (Doyé 2005 : 7)

En somme, on pourrait ainsi résumer l'intercompréhension entre langues : « Je comprends la langue des autres, sans être en mesure de la parler. C'est pourquoi, quand j'échange avec eux, je leur parle ma langue et je comprends la leur. ». (DGLFLF 2006 : 3)

Nous avons dit que l'intercompréhension fonctionnelle entre langues était le fait de comprendre une ou plusieurs langues sans la ou les parler. La performance d'une telle communication plurilingue est optimale dans les situations où les personnes maîtrisent déjà les deux langues utilisées dans la communication (par exemple, un francophone connaissant l'anglais et un anglophone connaissant le français). Cette pratique est également très fonctionnelle dans les situations où les personnes ne maîtrisent que les compétences de compréhension(s) écrite et/ou orale des autres langues. (Castagne 2007a : 156)

Toutes ces définitions mettent l'accent sur la capacité de compréhension d'une langue étrangère. Selon les auteurs en revanche, le fait qu'un locuteur ne s'exprime pas dans la langue de l'autre peut tenir d'une incapacité à utiliser cette langue, ou bien d'un choix de pas l'utiliser alors qu'il en serait capable, et ce pour des raisons de facilité d'expression par exemple.

Les principes de l'intercompréhension sont dignes des valeurs prônées par l'Union. En faisant l'effort de comprendre la langue de l'autre, un locuteur laisse son interlocuteur dans une position de confort, parce que ce dernier utilise sa langue maternelle, et ne le place donc pas dans une position d'infériorité. Le fait de s'exprimer dans sa langue permet

en effet de mettre à profit toutes les subtilités dont on dispose – et dont on ne peut disposer que dans la langue qui nous est la plus familière – et de produire des énoncés clairs et construits. Nous en sommes régulièrement les témoins pendant les colloques scientifiques internationaux : à la fin d'une intervention, il n'est pas rare qu'un membre de l'auditoire demande dans sa langue des précisions au conférencier, parce que c'est comme cela qu'il peut le mieux faire sentir quelles sont ses interrogations, et que le conférencier réponde à son tour dans sa langue, parce que c'est également pour lui le meilleur moyen d'explicitier sa pensée.

Il est extrêmement important de pouvoir accéder à l'information formulée dans sa langue originale, parce que « la formulation de toute entité culturelle (littératures, traditions, etc.) dans sa langue d'origine fait intimement partie de son contenu » (Castagne 2004a : 10). Les langues n'étant pas de simples enveloppes différentes que l'on poserait sur une pensée universelle, mais les véritables systèmes dans lesquels les populations se sont construit leurs cultures²⁷, il convient que chacun s'exprime dans sa langue pour donner la vue la plus fine et la plus subtile possible du message à transmettre :

Dans un contexte didactique, on rappellera qu'une vue sur le monde particulière ancrée dans chaque idiome représente un motif fort pour apprendre plusieurs langues. Les langues n'enrichissent pas seulement nos cultures, leur connaissance augmente aussi notre capacité d'interagir avec celles-ci. Cette vue, qui montre le monde dans ses facettes culturelles diverses, passe bien sûr (sauf pour les langues purement orales) par une compétence de lecture [...]. (Meissner et al. 2004 : 11)

Faire cas de son interlocuteur, c'est non seulement lui proposer un discours cohérent, mais aussi faire en sorte que son processus de compréhension soit le plus facile possible. É. Castagne, dans son article « Intercompréhension européenne et plurilinguisme : propositions pour quelques aménagements linguistiques favorisant la communication plurilingue » publié en 2004, recense huit aménagements linguistiques généraux et deux aménagements rythmiques et prosodiques permettant d'adapter son discours à son interlocuteur et de faciliter la compréhension de celui-ci. Dans les aménagements linguistiques généraux, l'auteur préconise de « développer une syntaxe avec des verbes tensés plutôt que des nominalisations », d'« éviter les phrases jugées trop

²⁷ Voir à ce propos dans ce travail le passage sur l'hypothèse Sapir-Whorf p. 119.

longues », d'« éviter les concentrations lexicales », d'« éviter les compléments en tête d'énoncé, les sujets postposés, les constructions apposées, les épithètes antéposées, les verbes à la voix passive, les dislocations », d'« éviter les enchâssements de propositions », d'« éviter l'accumulation de difficultés simultanées d'ordres différents », d'« utiliser la technique de l'affinage sémantique » et d'« exploiter la richesse lexicale et syntaxique d'une langue ». Dans les aménagements rythmiques et prosodiques, il conseille d'« adapter son débit de parole » et de « mettre en évidence les groupes syntactico-sémantiques » (Castagne 2004c). Le nombre des aménagements à utiliser est bien entendu à moduler en fonction des capacités de compréhension de son interlocuteur. Il ne s'agit pas de laisser de côté la subtilité d'expression rendue possible par l'emploi de la langue maternelle, mais de faire preuve d'adaptabilité, qui n'est précisément pleinement possible qu'en expression en langue maternelle.

L'intercompréhension, par son respect des langues et des locuteurs, mérite donc d'être développée. Elle se pratique évidemment depuis des siècles, mais des chercheurs s'y intéressent depuis seulement une vingtaine d'années. Des équipes de chercheurs, en didactique notamment, se sont donné pour but de la théoriser et de créer des méthodes formant à l'intercompréhension. Pour la première fois, des apprenants en langue sont formés à développer leur seule compétence de compréhension. L'Europe, ayant misé sur une politique du multilinguisme, a montré son intérêt pour ces nouvelles méthodes en soutenant les projets EuroCom, EuRom4, Euromania et Galatea. Nous allons, dans la section suivante, passer en revue les principaux programmes européens d'intercompréhension, et voir quels sont les grands principes qui ont été tirés de ces expériences.

1.2.2. La formation à l'intercompréhension

Nous présenterons d'abord les principaux programmes européens visant à former des apprenants à la compréhension de langues étrangères, puis nous attacherons à décrire les points clés de la formation à l'intercompréhension, par le biais de la description du programme InterCompréhension Européenne, développé à l'université de Reims.

1.2.2.1. Les différents programmes européens

Les paragraphes qui suivent donnent une courte description des principaux programmes européens formant à l'intercompréhension, classés par ordre alphabétique, et donnés avec leur page Internet d'informations²⁸.

- EuroCom (EuroCompréhension) : cette équipe de recherche a vu le jour en 1998 à l'université de Giessen, en Allemagne. Il s'agissait au départ de former des adultes germanophones qui maîtrisaient déjà une langue romane à comprendre des textes écrits dans d'autres langues romanes. Cette branche constitue désormais EuroComRom, à laquelle se sont ajoutées les branches EuroComGerm, pour l'intercompréhension entre langues germaniques, et EuroComSlav, pour l'intercompréhension entre langues slaves. Un texte rédigé en langue étrangère peut être compris, selon ce programme, grâce à sept « tamis » qui permettent de le filtrer jusqu'à aboutir au noyau compréhensible. Ces sept tamis sont présentés dans l'ouvrage de Horst G. Klein et Tilbert D. Stegmann, *EuroComRom – Die sieben Siebe : Romanische Sprachen sofort lesen können*, publié en 2000. Le premier tamis comprend le lexique international, le deuxième comprend le lexique pan-roman, le troisième traite des correspondances phonologiques, le quatrième des graphies et des prononciations, le cinquième traite de la syntaxe pan-romane, le sixième filtre les éléments morpho-syntaxiques et le septième traite des préfixes et des suffixes. L'équipe propose en outre une formation en-ligne sur son site.

Site Internet : <http://www.eurocomcenter.com/index2.php?lang=fr>

²⁸ Consultées le 24/07/09.

- EuRom4 : cette méthode a été mise au point dans les années 1990 par les équipes de quatre universités : Lisbonne, Salamanque, Rome et Aix-en-provence. Elle permet de former des adultes qui ont une langue romane pour langue maternelle à la compréhension écrite de quatre langues romanes (portugais, espagnol, italien et français). Les trois langues autres que celle des lecteurs sont vues simultanément, par le biais d'articles de journaux et de magazines, qui sont donnés dans leurs versions écrites et lues par un locuteur natif. Le développement de la méthode a conduit à la production d'un cédérom, utilisable en auto-formation ou en formation guidée. Signalons qu'une réactualisation de la méthode (EuRom5), avec l'ajout du catalan, est en cours de développement.

Site Internet : <http://sites.univ-provence.fr/delic/Eurom4/>

- Euromania : cette méthode a été développée de 2005 à 2008 par l'IUFM Midi-Pyrénées en partenariat avec l'université de Valladolid, l'institut polytechnique de Leiria, le Ciid de Rome et l'éditeur scolaire roumain Humanitas. Elle a la particularité de s'adresser aux enfants de 8 à 11 ans, qui ont une langue romane pour langue maternelle. Par cette méthode d'apprentissage disciplinaire (sciences, mathématiques, histoire, géographie et technologie), les enfants sont formés à la compréhension des six langues romanes que sont l'espagnol, le français, l'italien, l'occitan, le portugais et le roumain.

Site Internet : <http://www.euro-mania.eu/index.php>

- ICE (InterCompréhension Européenne) : ce programme, créé en 2001 à l'Université de Reims Champagne-Ardenne, est inspiré de la méthode EuRom4. Il vise à former des adultes à la compréhension écrite d'articles de presse, dont est également donnée la version lue par un locuteur natif. Les langues sont également vues simultanément, mais le programme ICE innove, notamment, en ce qu'il propose une formation à la lecture en langues du même groupe aussi bien que d'un autre groupe que celle des lecteurs. Ainsi, pour un francophone, les langues romanes abordées sont le portugais, l'espagnol, l'italien et le roumain, et les langues germaniques sont l'anglais, l'allemand et le néerlandais. Une extension aux langues nordiques et aux langues slaves est en projet.

Site Internet : <http://www.logatome.eu/>

- IGLO (Intercomprehension in Germanic Languages Online) : ce projet en-ligne est né en 1999 grâce à un partenariat entre les universités de Tromsø, Reykjavík, Lund, Salzburg, Antwerp, Hagen, et la Copenhagen Business School. Il vise à former des adultes ayant une langue germanique pour langue maternelle à comprendre à l'écrit (puis à l'oral dans un second temps) une autre langue germanique parmi les sept suivantes : l'anglais, l'allemand, le néerlandais, le norvégien, le suédois, le danois et l'islandais. Les apprenants sont guidés par plusieurs modules de présentation des langues germaniques, de leur grammaire, de leur vocabulaire et de leur phonologie.

Site Internet : <http://www.hum.uit.no/a/svenonius/lingua/index.html>

- Itinéraires romans : cette méthode en-ligne, proposée par l'Union Latine en 2003, permet à un public de collégiens qui ont une langue romane pour langue maternelle de découvrir d'autres langues romanes : le catalan, l'espagnol, le français, l'italien, le portugais et le roumain. Une initiation à la compréhension écrite et orale de ces langues leur est proposée par le biais de six modules animés et ludiques.

Site Internet : <http://dpel.unilat.org/DPEL/Creation/IR/index.fr.asp>

- Galanet : cette formation à l'intercompréhension entre langues romanes (portugais, espagnol, italien et français), développée en 2001 à l'université de Grenoble 3 pour un public d'adultes, passe par une plateforme Internet. Des modules d'aide (en grammaire, morphologie, etc.) sont proposés aux apprenants pour les guider dans la rédaction d'un « dossier de presse ». Ce dernier doit être rédigé à plusieurs : on attend donc des apprenants, qui ont pour langues maternelles les langues romanes citées ci-dessus, qu'ils communiquent entre eux en utilisant à souhait les unes ou les autres de ces langues pour la préparation de la documentation, dans les forums et pour la publication du document plurilingue final.

Site Internet : <http://www.galanet.be/>

- Galapro : ce projet a été créé en 2008 par un partenariat entre les universités d'Aveiro, de Grenoble 3, de Madrid, de Barcelone, de Mons-Hainault, de Lyon 2, de Cassino et de Al.I.Cuza. Il a pour but la formation de formateurs à l'intercompréhension entre langues romanes, en les dotant de compétences théoriques, méthodologiques et pratiques.

Site Internet : <http://www.galapro.eu>

- Galatea : cette méthode a été développée dans les années 1990 à l'université de Grenoble 3. Elle a pour but de former des étudiants et des adultes parlant une des langues étudiées à la compréhension écrite et orale d'une des trois autres langues romanes : le français, l'espagnol, l'italien et le portugais. L'équipe a mis au point des cédéroms d'autoformation, comprenant une présentation de la langue romane considérée, des modules d'aide, et des textes à comprendre. Le parcours de l'apprenant est enregistré, de sorte qu'un tuteur puisse ultérieurement lui donner conseil.

Site Internet : <http://w3.u-grenoble3.fr/galatea/>

Pour former des apprenants à la compréhension de textes produits en langues étrangères, les méthodes que nous venons de citer ont dû théoriser les processus en œuvre lors de cette compréhension, et dégager les principes facilitant l'accès aux textes. Nous allons maintenant exposer quels sont ces principes et quels sont les concepts clés auxquels la formation à l'intercompréhension familiarise ses apprenants. Pour ce faire, nous présenterons en détail le programme ICE, que nous connaissons le plus précisément pour en faire partie depuis cinq ans ; certains des points sont, cela s'entend, également valables pour d'autres méthodes.

1.2.2.2. Le programme ICE

Le programme ICE (InterCompréhension Européenne) prévoit de former des apprenants à comprendre, à partir de leur connaissance de deux ou trois langues, des textes rédigés en sept langues étrangères apparentées et/ou voisines. Sont considérées, par les membres de ce programme, comme apparentées des langues « proches par la parenté linguistique » et comme voisines des langues « proches par le voisinage géographique et les échanges que ce voisinage a autorisés et continue à autoriser » (Tyvaert 2008 : 256)²⁹. On propose ainsi à un apprenant francophone des textes rédigés dans les langues voisines et apparentées que sont le portugais, l'espagnol et l'italien, dans la langue apparentée mais non voisine qu'est le roumain, et dans les langues voisines mais non apparentées que sont

²⁹ Pour des réflexions sur les notions de « langues proches », « langues voisines » et « langues apparentées », voir Robert 2004b, Degache 2006 : 11, et le n°134 de la revue *Ela*, « Accès aux langues proches et aux langues voisines », paru en 2004.

l'anglais, l'allemand et le néerlandais. Pendant les premières séances, les apprenants cherchent à comprendre des textes rédigés dans des langues appartenant au même groupe que la sienne (romane ou germanique), puis, une fois que les processus d'accès au sens ont été identifiés et assimilés, dans des langues appartenant à l'autre groupe. Chaque séance, à raison d'une ou deux par semaine, dure deux heures, et toutes les langues de chaque groupe sont abordées lors de chaque séance. Le paragraphe suivant donne un bref descriptif du déroulement d'une séance.

Les apprenants sont regroupés par groupes de dix à quinze personnes. Le premier texte, un article de presse, rédigé dans une langue la plupart du temps inconnue des apprenants, est projeté, et le modérateur, celui qui anime la séance, diffuse la version lue par un locuteur natif pendant que les apprenants découvrent le texte. Une fois que celui-ci a été parcouru dans son ensemble, le modérateur demande au groupe si le thème général a été compris, et quelles sont les informations les plus importantes. Chose faite, le modérateur demande à un premier apprenant (ou un apprenant se propose) de prendre en charge l'interprétation du titre. Sa version lue est rediffusée, et l'apprenant propose une première interprétation. Si elle paraît acceptable au modérateur, un autre apprenant prend en charge la première phrase du texte, et si non, le premier apprenant relit le titre et donne de nouvelles interprétations, de plus en plus fines, jusqu'à ce qu'une d'entre elles soit considérée comme satisfaisante. Une difficulté peut également être mise de côté dans un premier temps pour être abordée à nouveau plus tard, à la lumière de ce qui aura été vu plus loin dans le texte. Pour la progression, le modérateur et d'autres personnes du groupe peuvent donner des indications, mais en aucun cas une réponse arrêtée face au problème rencontré. Ensuite, chaque phrase du texte est étudiée de la sorte par une personne différente. Une fois le texte terminé, le modérateur présente les versions écrite et lue du texte suivant, dans une autre langue du même groupe, qui peut éventuellement entretenir des liens avec le premier texte, que ces liens soient d'ordres thématique ou linguistique.

Cette description, si brève soit-elle, énonce successivement les points clés de la méthode, que nous allons maintenant présenter plus longuement. Ces points sont les suivants :

- la formation à la seule compréhension,
- des langues apparentées et/ou voisines,
- les langues sont vues simultanément,
- comprendre des textes,

- le choix des articles de presse,
- la mise en commun préalable des informations disponibles,
- l'interprétation,
- l'approximation,
- un groupe d'apprenants,
- le profil du modérateur,
- une méthode valorisante.

La formation à la seule compréhension

Comme dans tous les programmes d'intercompréhension, les apprenants sont uniquement formés à comprendre des langues étrangères, et ce pour plusieurs raisons. D'abord parce que la compréhension demande des efforts (ne serait-ce que des efforts de mémoire) moindres que la maîtrise de l'expression :

[...] la compréhension doit être distinguée de l'expression, la première ne requérant en fait, comme nous l'avons vu, aucun savoir pré-requis sur les langues alors que la seconde exige des savoirs linguistiques fins, ne serait-ce que du point de vue de la bonne maîtrise de la morphologie indispensable à la formulation correcte des représentations. (Tyvaert 2008 : 266)

Les apprenants n'ont pas à retenir quoi que ce soit en vue d'une réutilisation active ultérieure ; ils sont formés à une méthodologie, une pratique de lecture. Une fois que cette méthodologie a été acquise, elle fait partie du bagage de compétences des lecteurs. Selon une formule qui est chère à É. Castagne, fondateur et actuel directeur du programme, « on ne donne pas de poissons aux apprenants, on leur apprend à pêcher ». Deuxièmement, le fait d'avoir une approche réceptive d'une nouvelle langue suit naturellement l'ordre de développement des compétences lors de l'acquisition de la langue maternelle : « chez l'enfant, il semblerait que la phase de compréhension précède d'environ un an la phase de production » (Germain 1993 : 285). Dans leur article « La perception et la production de la parole avant deux ans », J. Bertoncini et B. De Boysson-Bardies décrivent les différentes étapes de l'acquisition de la langue maternelle. Les nourrissons mettent en œuvre des mécanismes perceptifs dès la naissance, et produisent leurs premiers mots vers l'âge de 12 mois. Les auteurs indiquent d'ailleurs : « Que la perception précède et soit nécessaire à la production est une évidence » (Bertoncini & De Boysson-Bardies 2000 : 126).

Puisqu'une période d'exposition à une langue est préalablement nécessaire à l'expression dans cette langue, il semble étonnant que, dans les méthodes d'apprentissage des langues, l'on demande aux apprenants de produire des textes construits si peu de temps après les séances d'initiation. Enfin, l'entraînement à la seule compréhension donne aux apprenants une aisance et une indépendance non négligeables, puisque, ne devant intervenir que dans leur langue maternelle pour interpréter les textes, la prise de parole est facilitée.

Des langues apparentées et/ou voisines

Le programme ICE propose à ses apprentis lecteurs de comprendre des textes rédigés dans des langues appartenant à deux groupes différents. Le choix des langues romanes, pour un francophone, a été effectué par la majorité des autres programmes cités plus haut. C'est la proximité génétique qui est alors exploitée, toutes ces langues descendant du latin. Cette origine commune leur confère des similitudes lexicales, morphologiques et syntaxiques telles que l'intercompréhension entre les langues néo-latines demande un nombre réduit d'heures de formation (48 heures de formation pour les langues romanes avec le programme ICE, 36 heures avec EuRom4). Le choix des langues germaniques, pour un francophone, est en revanche inédit. C'est la proximité géographique qui est alors exploitée, le français, l'anglais, l'allemand et le néerlandais étant parlés dans la même zone nord-occidentale de l'Europe (est alors également exploitée une proximité génétique plus éloignée, les langues romanes et les langues germaniques appartenant à la même famille des langues indo-européennes). Les liens entre langues qui sont parlées dans des zones proches les unes des autres sont étudiés par la linguistique aréale, qui « privilégie les contacts géographiques et conduit à raisonner en termes d'extension et de recouvrement d'aires, en maillons de chaîne » (Szulmajster-Celnikier 1998 : 76). Ce type d'études a vu le jour avec la publication en 1872 par J. Schmidt de l'ouvrage *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. L'auteur y développe la « théorie des ondes » (*Wellentheorie*), selon laquelle les innovations linguistiques ont une expansion concentrique à partir d'un foyer, telles les ondes à la surface de l'eau à partir du point d'impact d'une pierre. Ainsi, les ressemblances entre langues ne s'expliquent plus exclusivement par des relations de filiation. A. Crépin précise :

La linguistique aréale s'attache à repérer les traits partagés par des langues en contact. Elle cherche l'origine des phénomènes : conséquences de bilinguisme, emprunts, invasion ou prestige économique-culturel. (Crépin 1994 : 20)

Il mentionne, à propos du français, de l'anglais, de l'allemand et du néerlandais :

On note un certain nombre de caractères partagés par ces langues. Nos quatre langues expriment obligatoirement le sujet grammatical (sauf pour un verbe coordonné de même sujet) : je jeterai le filet, I will let down the nets, alors que d'autres langues, comme le latin, peuvent ne pas l'exprimer : laxābo rēte. Nos quatre langues emploient des temps composés avec des auxiliaires [...]. Nos quatre langues possèdent des articles, le latin n'a pas d'article : le filet, the nets, mais en latin simplement rēte ; une grande quantité de poissons, a great shoal of fish, mais LT piscium multitudinem cōpiōsam. (ibid. : 20)

On pourrait ajouter d'autres caractères partagés, tels que l'utilisation d'un même alphabet, l'accord du substantif, etc.

Le programme ICE tire ainsi parti des deux types de proximité qui relient le français à l'anglais, à l'allemand et au néerlandais : leur proximité génétique relative (96 heures de formation sont prévues pour les langues germaniques contre 48 pour les autres langues romanes) et leur proximité géographique. C'est ce que précisent É. Castagne *et al.* dans l'article « Les langues européennes et l'intercompréhension », publié en 2007 :

[...] même si les liens de parenté jouent un rôle évident dans l'intercompréhension, les grandes migrations historiques ou les échanges de voisinage plus modernes ne sont pas à négliger, notamment dans les zones frontalières. C'est avec cette approche hybride et les concepts de langues apparentées et langues voisines qu'est développé le programme ICE. (Castagne et al. 2007 : 22)

Les langues sont vues simultanément

Les similitudes entre langues de même groupe sont exploitées par la méthode. En effet, plutôt que d'être formés à la compréhension d'une langue particulière, les apprenants acquièrent une compétence de compréhension en sept langues étrangères. Lors de la première phase de formation, les apprenants francophones sont formés à la compréhension de quatre langues romanes (portugais, espagnol, italien, roumain), et, lors de la seconde phase, à la compréhension de trois langues germaniques (anglais, allemand, néerlandais). Les langues d'un même groupe sont toutes vues au cours d'une même séance (c'est ce qui

est signifié par formation *simultanée* à la compréhension). Ainsi, les données reconnues dans une langue qui sont également valables pour les autres langues du même groupe pourront être réexploitées par les apprenants lors de l'étude des textes rédigés dans les autres langues :

C'est pour cette raison qu'est préconisée la présentation systématique de textes dans un bouquet de [...] langues apparentées [...], car se trouvent ainsi multipliées les occasions d'identification sémiotique concernant l'ensemble des langues [...], ce qui attire l'attention sur des propriétés générales valables au niveau de la fratrie de langues. (Tyvaert 2008 : 263)

Le modérateur encourage également les apprenants à tirer parti de leurs connaissances dans une ou plusieurs autres langues étrangères qu'ils maîtrisent pour comprendre un texte donné. Les allées et venues constantes entre langues facilite en effet la mise en réseau des informations.

Comprendre des textes

Dès la première séance, le modérateur demande aux apprenants d'essayer de comprendre des *textes*. Jamais il ne leur est présenté de points de morphologie, de syntaxe ou de phonologie des langues : les apprenants peuvent être totalement débutants. La compréhension des textes ne passe pas par l'apprentissage préalable, ni consécutif d'ailleurs, de listes de vocabulaire, ni par la présentation, comme c'est le cas dans d'autres méthodes, de points de conjugaison par exemple, mais par les seules ressources des lecteurs. Comme le note J.-E. Tyvaert :

On ne travaille pas sur des illustrations de propriétés linguistiques à assimiler mais sur des textes (en langues diverses) présentés comme supports d'informations à appréhender. (Tyvaert 2008 : 259)

Comme nous le notions dans le point « La formation à la seule compréhension », le jeune enfant, lors de l'acquisition de sa langue maternelle, passe par une phase d'écoute et de compréhension avant de passer lui-même à l'expression. Il convient maintenant de noter que c'est à la production de *textes* qu'il est soumis, de la part de son entourage. Nous donnons au texte la définition que lui donne F. Rastier : « *Un texte est une suite linguis-*

tique empirique attestée, produite dans une pratique sociale déterminée, et fixée sur un support quelconque » (Rastier 2002a). C'est en entendant des textes que l'enfant se construit le système de sa langue, qu'il maîtrise déjà en arrivant à l'école, et donc qu'il maîtrise avant de recevoir un savoir théorique sur sa langue.

Ainsi, le programme ICE propose uniquement de former à la compréhension de textes, écrits en l'occurrence. La méthode n'étant pas une formation à l'expression en langue étrangère, on ne demande pas aux apprenants de retenir précisément tous les éléments vus lors d'une séance. Il leur suffit de savoir découvrir (et pas obligatoirement de savoir les produire à leur tour) les indices suffisant à la construction du sens.

Le choix des articles de presse

Les textes ne sont pas choisis au hasard. Il s'agit d'articles de presse, et plus spécifiquement d'articles de presse informatifs. Ces textes présentent l'avantage, pour la méthode, de respecter une présentation-type, plus ou moins équivalente dans toutes les langues concernées. Ils appartiennent à un genre très codifié, dont les règles d'écriture sont par exemple enseignées dans les écoles de journalisme. Ainsi, les apprenants ont déjà un savoir sur la structure du texte : ils s'attendent à trouver un titre, éventuellement un chapeau, et, dans le corps de l'article, le récit d'un événement, par les réponses aux « questions de référence » que sont « Qui, quand, quoi, où, comment, pourquoi ? » (Agnès 2002 : 97)³⁰. Leur lecture est donc déjà guidée par la recherche d'un certain type d'informations attendues :

C'est que les auditeurs possèdent une représentation de ce que devrait être tel type de texte et peuvent mettre à profit cette connaissance dans le décodage d'un message. Même les tout jeunes enfants ont une perception, par exemple, de la structure du conte de fée et réagissent très négativement à toute tentative du locuteur de s'éloigner de cette norme [...]. (Germain 1993 : 289)

L'article de presse n'est pas le seul genre codifié ; tous les genres le sont. Celui-ci répond néanmoins particulièrement bien aux besoins de la méthode pour deux raisons. D'abord parce que ces textes sont des textes courts, c'est-à-dire qui peuvent être vus en une séance ou deux. Les apprenants peuvent donc prendre connaissance du texte dans son entier, lors de la première lecture-écoute, avant de commencer à proposer des

³⁰ Voir à ce propos le passage sur l'horizon d'attente du lecteur p. 170.

interprétations phrase par phrase. Les articles de presse informatifs sont des textes indépendants les uns des autres, dont la lecture permet d'obtenir une information principale formant un tout. L'autre raison ayant présidé au choix des articles est que ces derniers relatent des événements actuels, soit internationaux, soit locaux mais, dans ce dernier cas, relatant un fait d'actualité comme tous les journaux en rapportent. Les apprenants ont donc déjà un savoir encyclopédique minimal sur le sujet abordé, et ils ont éventuellement déjà lu un article sur le même thème dans leur langue. La compréhension est ainsi aidée par les connaissances qu'ils possèdent sur les personnes, les lieux ou les questions mentionnés. J.-E. Tyvaert résume :

Ce n'est pas un hasard non plus si le genre textuel exploité permet de mobiliser avant même le début de la tâche de compréhension et tout au long de son accomplissement des informations pragmatiques en nombre, qui focaliseront la recherche de la signification des formes linguistiques observées. Ainsi est fixé le cadre général de la représentation à élaborer sur la base d'indications qu'on sait être de type sémiotique, la compréhension s'ordonnant finalement à l'affinement d'un sens qui est, à un niveau de généralité extrême, déjà là. On ne construit pas le sens à partir d'indications significatives élémentaires, on le détermine en le spécifiant de manière de plus en plus précise à partir d'une appréhension générale pré-existante. (Tyvaert 2008 : 265 ; nous soulignons)

La mise en commun préalable des informations disponibles

Cette étape constitue également une aide à la compréhension du texte étudié. Après la première lecture-écoute, le modérateur demande aux apprenants s'ils ont compris le sujet abordé dans l'article, et si oui, quels sont les éléments qui leur ont permis de le comprendre. Les apprenants citent alors éventuellement les noms propres, les dates et les mots qu'ils ont reconnus. Une fois que le sujet de l'article a été cerné par tous, le modérateur demande aux apprenants, en faisant abstraction du texte quelques minutes, de dire ce qu'ils savent sur ce sujet, jusqu'à ce qu'une mise au point générale soit faite. É. Castagne précise l'importance de cette phase :

Nous sommes donc partis du principe qu'une partie importante du travail de compréhension se jouait avant le début du texte : à moins de choisir un texte dont le sujet nous est totalement inconnu, on possède toujours beaucoup d'informations

sur un thème développé, plus que l'on ne pensait, enfouie dans notre mémoire, dont une partie émergera du texte. (Castagne 2004b : 94)

Cette mise en commun préalable à l'interprétation phrase par phrase permet aux apprenants d'avoir à l'esprit des informations qu'ils seront susceptibles de trouver dans le texte, c'est-à-dire d'avoir des mots déjà disponibles à l'esprit, qu'ils pourront produire plus facilement au moment de l'interprétation³¹.

L'interprétation

Pour comprendre un texte qu'on leur présente, les apprenants n'ont parfois aucune connaissance sur la langue en question. Il s'agit alors pour eux de trouver par eux-mêmes dans le texte les signes qui leur permettront de faire progresser leur représentation. Ces signes sont de natures diverses : des nombres, des noms propres, des mots transparents (dont nous nous attacherons à donner une définition dans la deuxième partie de ce travail), des marques morphologiques, des structures syntaxiques identifiables. De plus, comme nous l'avons indiqué dans les paragraphes ci-dessus, les apprenants ont déjà un savoir sur le texte (type d'informations, syntagmes attendus, etc.). Dans les faits, lorsqu'un apprenant cherche à comprendre une phrase, il commence par repérer tous les signes (identifiables par lui, car toutes les marques ne sont pas également significatives pour tous les apprenants) qui constitueront la base pour sa première interprétation de la phrase. On conseille de remplacer les séquences qui n'ont pas été reconnues par des « mots vides » (Valli 2001 : 9), en l'occurrence « machin » ou « machiner », « qui peuvent supporter toutes les informations grammaticales reconnues (comme la catégorie, le genre et le nombre du N ou la personne et le temps du V), et qui évitent de laisser vides certaines positions et de bloquer la progression dans le texte » (Castagne 2004b : 98). É. Castagne illustre ce procédé par un exemple tiré de l'expérience EuRom4 :

Par exemple, si l'on reprend l'énoncé du texte italien consacré à une découverte archéologique :

*13a) la loro costruzione risale a 1000 anni prima delle piramidi egiziane
ils peuvent bloquer au premier passage sur risale qui n'est pas transparent alors
même qu'ils ont repéré qu'il s'agissait visiblement d'un verbe. Dans un premier
passage, ils peuvent proposer :*

³¹ Voir à ce propos le passage sur l'amorçage sémantique p. 165.

13b) la construction « machine » à mille ans avant les pyramides égyptiennes

L'usage du mot vide permet ainsi de prendre connaissance, dans une structure construite, du complément temporel avec lequel le stéréotype le plus courant impose de traduire risale par « remonte à » :

13c) la construction remonte à mille ans avant les pyramides égyptiennes
(Castagne 2002 : 11-12)

Les mots vides sont importants en ce qu'ils permettent de produire un énoncé syntaxiquement complet. Ils suscitent donc une première interprétation d'ensemble, avec des lacunes certes, mais qui portent déjà des informations de catégories, ce qui, combiné aux autres séquences identifiées de la phrase, commence à leur apporter un sens. Les séquences opaques sont ainsi comprises au fur et à mesure des relectures d'une même phrase. À chaque relecture, l'apprenant affine son interprétation, notamment en faisant des inférences. Nous empruntons à P. Bougé et S. Caillies leur définition de l'inférence :

Effectuer une inférence est un processus par lequel le lecteur établit lors de la compréhension de textes une relation entre les propositions sémantiques en cours de traitement et les propositions antérieurement traitées et/ou les connaissances stockées en mémoire à long terme [...]. (Bougé & Caillies 2004 : 80)

Sont évoqués dans cette définition les deux principaux types d'inférences pratiquées par les apprenants ICE : l'inférence de liaison et l'inférence d'élaboration. « Une inférence de liaison consiste à lier des éléments dans un texte alors qu'une inférence d'élaboration consiste à récupérer des informations en mémoire à long terme, liées à la situation décrite par le texte » (*ibid.* : 80). Les inférences de liaison peuvent notamment être phraséologiques : par exemple, c'est grâce à une collocation que l'apprenant a trouvé « la construction *remonte* à mille ans avant les pyramides égyptiennes » (exemple cité ci-dessus). Pour les inférences d'élaboration, les informations que l'apprenant a sur le texte avant même sa première lecture sont particulièrement utiles. Soit le début d'un article du *Times* du 19/07/09 :

Ah, the olive groves of balmy England

SUBTROPICAL crops such as dates, figs and rice could become staples of British agriculture within 20 years, according to government forecasts.

The assessment, produced by officials at the Department for Environment, Food and Rural Affairs (Defra), outlines future possibilities for British food production based on recent climate data.³²

Les mots *crops* et *staples*, de la première phrase, ne sont pas transparents pour un francophone. Dans un premier temps, l'interprétation de cette première phrase par un lecteur qui aurait quelques bases en anglais pourrait être « Les machins subtropicaux tels que les dattes, les figues et le riz pourraient devenir des machins de l'agriculture britannique d'ici 20 ans, selon le gouvernement ». Grâce à la lecture du titre, du reste de l'article et de ses connaissances, un lecteur pourrait mettre en relation le fait qu'on parle d'olives en Angleterre, de données récentes sur le climat, et que le réchauffement climatique fait dire aux scientifiques que les zones chaudes remonteront vers le nord. *Crops* serait probablement interprété par inférence de liaison : les dattes, les figues et le riz sont des « produits », des « aliments » subtropicaux. *Staples* pourrait être interprété par inférence d'élaboration : la température risquant d'augmenter en Angleterre, on y cultivera des produits qu'on cultive actuellement dans les pays chauds. Alors qu'ils sont aujourd'hui absents du paysage anglais, les produits subtropicaux pourraient devenir « communs », des « produits courants » d'ici une vingtaine d'années.

L'inférence est une des procédures les plus importantes pour la méthode ICE. Le modérateur incite les apprenants à y recourir dès qu'une séquence reste opaque à leurs yeux, c'est-à-dire s'ils n'identifient pas de marques morphologique, lexicale ou syntaxique supplémentaires qui leur permettraient d'affiner leur interprétation. Si une séquence peut être découverte par une inférence de liaison, le modérateur met l'accent sur les éléments linguistiques du contexte qui peuvent déclencher une interprétation ; et si une séquence peut être découverte par une inférence d'élaboration, il suscite chez les apprenants des représentations de la scène ou de l'événement décrits (par le biais de questions du type « Et vous, si vous étiez dans cette situation, que feriez-vous ? », « Quelle réaction pourrait vraisemblablement entraîner cette annonce ? »). Étant donné que « [l]es inférences ne sont pas les mêmes pour tous, car l'expérience et la connaissance du monde varient selon les

³² <http://www.timesonline.co.uk/tol/news/environment/article6719157.ece> (consultée le 26/07/09)

individus » (Lerot 1993 : 55), plusieurs inférences peuvent être valables pour tenter de comprendre un mot d'une phrase. Il revient donc au modérateur de confirmer ou d'infirmier une inférence produite.

La version oralisée des textes constitue également une aide pour les apprenants, non seulement pour la reconnaissance de certains mots dont la forme orale est plus facilement identifiable que la forme écrite, mais surtout pour le découpage en syntagmes principaux :

On vérifie que le mode oral est très utile pour découper les phrases en syntagmes principaux puisque les langues étudiées sont assez proches de ce point de vue en termes prosodiques (les divers syntagmes principaux se trouvant en quelque sorte unifiés du point de vue fonctionnel en un fragment prosodiquement identifiable).
(Tyvaert 2008 : 260)

Si nous utilisons le mot « interprétation » et pas celui de « traduction » pour décrire le texte que produisent les apprenants quand ils cherchent à comprendre une phrase, c'est parce qu'on n'attend pas de ces derniers qu'ils proposent un texte en français qui rendrait toutes les nuances du texte de départ. Contrairement à la traduction qui consiste à produire un texte qui peut se substituer au texte source, l'interprétation de la méthode ICE a pour seul but de montrer que l'apprenant a compris le sens global du texte. Phrase après phrase, le modérateur vérifie que les apprenants comprennent les informations principales, mais peut délibérément décider de passer sous silence une nuance, ou même une information qui ne seraient pas capitales. La méthode a donc pour but de former les apprenants à savoir retirer les informations principales d'un texte, grâce à la reconnaissance de mots transparents et de structures syntaxiques identifiables et grâce à leurs capacités inférentielles.

L'approximation

L'approximation est un principe fondamental de la méthode. Elle est non seulement acceptée, mais aussi conseillée. Il s'avère en effet qu'elle constitue un moyen de comprendre plus facilement les textes, grâce à la possibilité qu'elle laisse aux apprenants de ne pas produire des interprétations figées, et donc non évolutives et non perfectibles. Par exemple, l'usage du mot « machin », que nous avons cité dans la section précédente, permet aux apprenants de proposer une interprétation valable, en le remplaçant par des

mots de signification d'abord très générale, puis de moins en moins générale, jusqu'à ce que le modérateur estime que le mot proposé en français convient. Ce dernier peut d'ailleurs tout à fait décider de garder un mot de signification plus générale que celui qui serait donné dans une traduction s'il suffit à comprendre l'information véhiculée. Le modérateur peut également estimer qu'un mot ou un syntagme n'ont pas besoin d'être rendus pour que l'interprétation soit satisfaisante : c'était le cas dans l'exemple du *Times* donné ci-dessus, où on avait vu que « according to government forecasts » pourrait être interprété par « selon le gouvernement ». *Forecasts* a été laissé de côté, et seul *government* a été interprété, mais ces approximations ne gênent en rien la compréhension de l'énoncé. Les apprenants n'étant pas des spécialistes des langues envisagées, et ne cherchant pas à le devenir, il est naturel qu'ils ne produisent pas des traductions et qu'on ne leur demande pas d'en produire. J.-E. Tyvaert déclare à ce propos qu'il faut savoir trouver une « souplesse de lecture » (Tyvaert 2008 : 262). Cette souplesse intervient tant au niveau de l'interprétation produite, comme nous venons de le voir, qu'à celui du processus de compréhension. Les apprenants doivent apprendre à ne pas se focaliser sur chaque mot et à accepter de passer sur les passages difficiles, quitte à y revenir plus tard. Si une phrase paraît trop longue, ils peuvent ne pas prendre en compte dans un premier temps certains syntagmes non nécessaires, s'ils parviennent à les repérer eux-mêmes, ou sur conseil du modérateur : c'est ce que C. Blanche-Benveniste appelle « aménagements progressifs de la syntaxe » (Blanche-Benveniste 2004) :

Le droit à l'approximation se manifeste aussi par d'autres choix, qui consistent à ne garder que certaines parties du texte et à en abandonner d'autres, dont l'importance est jugée secondaire, du moins dans une première étape. Le cas typique est celui des passages mis entre parenthèses ou entre tirets : comme un débutant juge toujours difficile de les intégrer dans l'ensemble de la phrase qu'il déchiffre, il a tendance à les éliminer. La méthode EuRom4 l'y encourage. (Blanche-Benveniste 2007 : 174)

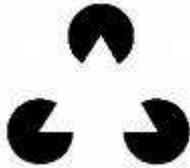
L'auteur cite en exemple, outre les passages entre parenthèses ou entre tirets, le cas de la mention de l'auteur, du lieu, du temps qui suivent des paroles rapportées et le cas des appositions.

Le modérateur peut décider de laisser en suspens un mot qui pose problème et de passer à la phrase suivante. Un mot peut en effet se trouver éclairé rétroactivement par un passage plus loin dans le texte. Que ce soit au niveau d'une phrase ou au niveau du texte, il faut savoir lire, non pas uniquement de gauche à droite et de haut en bas, mais en faisant des allées et venues. C'est ce qu'É. Castagne appelle la « progression en couches successives alternées » :

En observant que les lecteurs débutant dans une langue étrangère adoptaient très souvent une approche mot à mot, même quand ils utilisaient une approche globale dans leur langue maternelle, nous avons décidé qu'il fallait les convaincre d'utiliser une approche qui leur permettait d'accéder au sens du texte, non pas au premier passage mais après plusieurs passages, et non pas seulement dans un seul sens (de gauche à droite pour nous) mais aussi après des allers et retours. Cette approche souple compte plusieurs avantages : les candidats ne sont pas sous tension puisqu'ils savent qu'ils ont droit à plusieurs essais avant de réussir ; ils ont la possibilité de sauter les difficultés, de prendre connaissance de l'ensemble des structures transparentes, et, à partir de là, de recueillir des informations permettant souvent de procéder à davantage d'inférences, souvent meilleures. (Castagne 2004b : 97-98)

La capacité d'approximation passe également par la capacité à ne pas s'arrêter sur la forme exacte des mots et sur le sens qu'ils suggèrent. Les apprenants sont en effet invités à pratiquer des « ajustements formels » et des « ajustements sémantiques » (notions que nous développerons en détail dans les sections 2.3.1. et 2.3.2.). Ces deux méthodes, qu'ils découvrent par eux-mêmes, sont peut-être les meilleurs exemples de « souplesse de lecture » : il s'agit de savoir reconnaître une forme proche d'une forme connue de sa langue maternelle ou d'une autre langue maîtrisée sous une forme apparemment non identifiable de la langue étrangère, et de savoir trouver un sens adéquat au contexte sous un mot qui, de par sa forme, suggère un autre sens, inadéquat celui-là. Par exemple, pour un traducteur, le substantif anglais *umbrella* n'est pas transparent puisqu'il ne ressemble pas à son équivalent « parapluie ». Dans le cadre du programme ICE, *umbrella* est considéré comme transparent puisque sa forme est proche de celle d'un mot français, « ombrelle », et que la signification de ce mot français est proche de celle du mot que l'on aurait tendance à utiliser en français dans le contexte considéré, « parapluie ».

B. Laks, dans son article « De l'approximation dans la relation phonétique/phonologie », montre avec force précisions que, cognitivement, il n'est pas nécessaire de percevoir tous les détails d'un objet (quel qu'il soit) pour pouvoir se construire une représentation. À propos de certaines illusions d'optique, dont les triangles de Kanizsa,



il déclare :

C'est dire que l'approximation, ou la sous détermination informationnelle de base, est suppléée par des inférences sémantiques ou conceptuelles de haut niveau dans une perception qui est d'emblée holistique. (Laks 2007 : 185)

En processus de lecture, le bruitage des caractères n'est pas non plus un frein à la représentation, comme le prouve ce texte :

Selon une étude de l'Université de Cambridge, l'ordre des lettres dans un mot n'a pas d'importance, le seule chose importante est que la première et la dernière soit à la bonne place. Le reste peut être dans n'importe quel ordre et vous pouvez toujours lire sans problème. C'est parce que le cerveau humain ne lit pas chaque lettre elle-même, mais le mot comme un tout. (ibid. : 189)

L'inévitable recherche de sens conduit les lecteurs à construire ce sens grâce aux éléments disponibles, en recréant le reste des informations du message :

Cette dynamique systémique met en place des phénomènes d'auto-complétion qui permettent de pallier globalement à [sic] l'absence ou la dégradation d'informations pertinentes de bas niveau. Le bruitage ou l'absence de certaines de ces informations est ainsi rarement destructeur et la cohérence d'ensemble y supplée le plus souvent. (ibid. : 189)

R. Jakobson, dans ses *Essais de linguistique générale*, avait déjà noté la capacité d'un auditeur à reconstruire un message grâce aux informations fragmentaires dont il dispose :

D'habitude, [...] le contexte et la situation nous permettent de négliger un pourcentage élevé des traits, phonèmes et séquences du message reçu sans en compromettre la compréhension. La probabilité d'occurrence dans la chaîne parlée varie suivant les traits et de même pour chaque trait suivant les contextes. Pour cette raison, il est possible, d'un certain point de la chaîne, de prévoir avec une précision plus ou moins grande quels seront les traits suivants, de reconstruire les précédents, et, finalement, d'inférer, à partir de certains traits d'un faisceau, les autres traits simultanés. (Jakobson 1963 : 106)

Tout comme, en langue maternelle, on n'a pas besoin de prêter attention à chaque lettre ou à chaque phonème d'un mot (voir les différents types de variations phonétiques dans l'article de B. Laks), un lecteur débutant en langue étrangère n'a pas besoin de se focaliser sur tous les signes du texte qu'il cherche à comprendre.

Cette approximation, dont il faut savoir faire preuve, s'explique par la notion de « redondance », introduite par la théorie de l'information dans les années 1950. A. J. Greimas la définit ainsi :

*[...] la **redondance** désigne, pour une quantité d'information donnée, l'écart entre le nombre minimal de signaux (ou d'opérations d'encodage et de décodage) nécessaires à sa transmission, et celui – généralement de beaucoup supérieur – de signaux (ou d'opérations) effectivement utilisés. Sont considérés comme **redondants** les signaux superflus parce que répétés. (Greimas & Courtés 1979 : 309)*

Comme le signalent J. Dubois *et al.*, au point de vue linguistique :

Le rôle joué par la redondance est double :

- a) *la redondance conserve l'information que des « bruits » peuvent supprimer ;*
- b) *la redondance fonctionne comme facteur de cohésion syntagmatique. (Dubois et al. 2002 : 402)*

Ils précisent également :

Dans les langues naturelles, le taux de redondance [...] est très élevé : 50 p. 100 en moyenne en anglais et en français. Les manifestations de la redondance y sont diverses et se retrouvent à tous les niveaux de la langue : niveau phonétique, morphématique, syntaxique, sémantique. (ibid. : 401)

Les apprenants n'ont donc pas à se focaliser activement sur tous les signes du texte qu'ils ont sous les yeux : le taux de redondance implique que la même information se manifeste sur plusieurs éléments du texte, et donc qu'il suffit parfois d'accéder au sens d'un seul de ces éléments pour obtenir l'information.

La perception approximative se révèle donc être une pratique non seulement acceptable, mais encore naturelle. Elle est à l'œuvre dans les différents types de perceptions sensorielles, et en particulier en lecture en langue maternelle. La méthode ICE, à la suite de ces constatations, préconise également l'approximation pour la lecture en langue étrangère, en tant que principe essentiel dans le processus de compréhension.

Un groupe d'apprenants

Contrairement aux méthodes disponibles dans le commerce, avec lesquelles les apprenants se forment seuls, ou à l'enseignement des langues dans le secondaire et à l'université, où les apprenants, bien qu'appartenant à une classe, sont en relation binaire avec l'enseignant, la méthode ICE tire profit de la présence de plusieurs apprenants pendant les séances. Ces derniers sont tous interrogés tour à tour pour interpréter une phrase, mais l'interaction est un élément important pour la réussite du processus de compréhension. Lorsqu'un apprenant ne parvient pas à progresser dans la compréhension d'une phrase, les autres peuvent proposer leur aide, en signalant une transparence qui serait passée inaperçue, ou en attirant l'attention sur la structure de la phrase par exemple. « Les interactions entre participants sont toujours encouragées » (Castagne 2002) ; elles sont en effet profitables tant pour la personne qui reçoit les informations que pour celles qui les donnent : les apprenants sont ainsi toujours actifs, pendant toute la séance, et non seulement au moment de comprendre leur phrase, et, puisqu'il est bien connu qu'on ne comprend vraiment que ce que l'on a à verbaliser, ils progressent au moment où ils font progresser leurs camarades. Enfin, il est intéressant de tirer parti de la diversité des profils :

ceux qui ont des connaissances en langues romanes peuvent aider ceux qui ont des connaissances en langues germaniques pendant la première phase de la formation, et inversement pendant la seconde phase.

Le profil du modérateur

Comme nous l'avons déjà signalé, la méthode ICE ne demande aucun pré-requis en langues étrangères. Et, contrairement à d'autres méthodes d'intercompréhension, aucun savoir linguistique ou métalinguistique n'est enseigné aux apprenants durant les séances. Le modérateur a donc un rôle fort différent de celui d'un professeur de langue. Il n'est pas indispensable que le modérateur soit un spécialiste des langues étudiées ; il doit en revanche maîtriser la méthodologie de recherche du sens par essais successifs. Son rôle consiste par exemple à attirer l'attention des apprenants, s'ils ne parviennent pas à comprendre une phrase, sur telle marque linguistique du texte qui leur aurait échappé ou à leur rappeler que tel mot avait été rencontré sous une forme presque similaire dans un texte rédigé dans une autre langue. Il ne délivre pas de nouvelles connaissances, il laisse aux apprenants la possibilité de trouver les informations par eux-mêmes en les guidant si besoin est. En faisant prendre conscience aux apprenants qu'ils ont les capacités de trouver les indices nécessaires à la construction du sens, et en leur rappelant le contexte dans lequel se passe le fait rapporté par l'article pour l'élaboration d'inférences, le modérateur joue le rôle de « mémoire collective des apprenants »³³. De plus, comme le montrent É. Castagne et J.-P. Chartier dans leur article « Modélisation de la formation d'éducateurs à l'intercompréhension de plusieurs langues : réflexions et pistes », le modérateur doit accepter d'une part de *ne pas être considéré* comme un « puits de savoir », et d'autre part de *ne pas être* un « puits de savoir ». Il faut savoir dire « Je ne sais pas ; l'un d'entre vous aurait-il une piste à me proposer ? » etc. Les formateurs doivent donc « accepter les lacunes aussi bien chez les apprenants que chez eux-mêmes » (Castagne & Chartier 2007).

Une méthode valorisante

Cette méthode est, pendant les premières séances, relativement déconcertante pour les apprenants. On leur demande directement de comprendre des textes alors qu'ils ne connaissent pas les langues dans lesquelles sont rédigés ces textes ; ils voudraient connaître des règles grammaticales mais on ne les leur présente pas ; ils ne parviennent pas à

³³ Thèse de T. Chazal, en cours de parution au CIRLEP à l'Université de Reims Champagne-Ardenne.

comprendre une phrase et on leur dit que ce n'est pas grave, etc. La différence avec l'enseignement en langues qu'ils ont reçu jusque là est telle qu'ils ont parfois du mal, au début, à accepter cette nouvelle pratique. Cette phase de transition n'est cependant guère longue, et le flottement méthodologique cède vite la place à l'enthousiasme. Puisqu'ils parviennent à comprendre des textes grâce à leurs seules capacités interprétatives, cette méthode est rapidement considérée comme « gratifiante »³⁴. Jusqu'à présent, ils avaient dû apprendre par cœur des tableaux de conjugaison, de déclinaisons, des listes de vocabulaire, pour un résultat qui n'était parfois pas à la hauteur de leurs espérances. Grâce au programme, ils découvrent qu'ils ont le moyen, s'ils se font confiance, de comprendre des articles entiers après quelques heures seulement de formation. C'est en effet une question de confiance, envers soi-même et envers le texte. Le texte est cohérent et il a été retenu parce qu'il donne suffisamment d'indices pour permettre une compréhension globale réussie.

La méthode est également considérée comme valorisante grâce au droit à l'erreur et à l'approximation (Castagne 2007b). Le recours à des mots de signification plus générale que celle du mot étranger considéré est encouragé ; on ne demande pas aux apprenants de connaître quelque mot que ce soit pour comprendre une phrase ; et s'ils se trompent ou hésitent, ils peuvent relire, demander de nouvelles écoutes et proposer de nouvelles interprétations de la phrase autant que nécessaire.

Nous venons de développer les points qui nous paraissent essentiels pour décrire la méthode d'acquisition de langues telle qu'elle est mise en œuvre dans le cadre du programme ICE. Il nous semble intéressant de la comparer à la méthode « lecture-traduction » qui est une méthode d'apprentissage de langues (Besse 1985 : 28-31). Ces deux méthodes présentent en effet un certain nombre de points communs. La méthode lecture-traduction, appliquée au XVIII^e siècle à l'enseignement du latin et des langues vivantes, préconise en effet pour les apprenants une première phase de contact avec la langue étrangère. Ce contact prend la forme de lecture de textes, « des 'morceaux choisis' de prose ou des 'documents authentiques' dont les références et le contenu ne les déroutent pas trop », qui « peuvent être ou non oralisés par l'enseignant et les enseignés » (Besse 1985 : 28). Les réflexions d'ordre grammatical ne doivent apparaître

³⁴ Remarque entendue lors d'une séance du séminaire « Dialogues multilingues », organisé à Reims en 2008-2009 par É. Castagne.

que bien plus tard dans l'apprentissage, une fois que les apprenants se sont familiarisés avec la langue étrangère :

La sensation, la pratique doit nécessairement précéder la réflexion. [...] l'important est que, dans une première étape, la compréhension en L2 l'emporte sur l'expression dans cette langue, qu'elle soit écrite ou orale. (ibid.)

On retrouve là les préoccupations de la méthode ICE quant à l'importance des textes et au fait que leur lecture doit précéder toute réflexion métalinguistique. Les textes sont considérés, dans les deux méthodes, comme des matériaux suffisants pour l'accès à une langue étrangère, en ce qu'ils contiennent les éléments qui permettront aux apprenants de les comprendre, et sans que ces derniers aient préalablement eu besoin d'apprendre des règles de grammaire.

La méthode lecture-traduction présente néanmoins une différence capitale avec la méthode ICE : les textes y sont présentés avec leur traduction dans la langue maternelle des apprenants. Il ne s'agit donc pas pour les apprenants de chercher à comprendre le sens du texte par eux-mêmes, mais de repérer et de retenir des régularités entre les deux langues. Par conséquent, cette méthode incite « les étudiants à percevoir la L2 essentiellement comme un stock d'étiquettes qu'on substitue aux signifiants de la L1, puisqu'on fait comme si le passage d'une langue à l'autre ne modifiait pas les concepts dénotés » (*ibid.* : 30). La méthode ICE, quant à elle, ne s'appuie que sur les capacités des apprenants à construire le sens à partir des indices qu'ils auront pu trouver dans le texte qu'ils cherchent à comprendre.

Nous retenons du développement que nous venons de consacrer aux buts visés par les programmes d'intercompréhension et plus particulièrement au programme ICE que la formation à l'intercompréhension donne une « souplesse de lecture » aux apprenants, qu'ils investissent à plusieurs niveaux lors du processus de compréhension par une approche non rigide de la phonologie, des éléments lexicaux et des structures syntaxiques. Ainsi, les mots étrangers qui ont une forme et une signification très proches du mot équivalent dans la langue du lecteur ne sont plus les seuls à être considérés comme transparents, puisque d'autres mots sont également mis à profit. Certaines équipes ont formalisé cette nouvelle approche de la transparence lexicale, qui rend compte du nombre plus important d'items couverts par la notion.

1.2.3. Aménagements de la définition traditionnelle de la transparence

Dans toutes les définitions de la transparence lexicale que nous avons citées dans les sections 1.1.2. et 1.1.3., les mots sont envisagés comme des unités transparentes ou non, sans prise en compte d'un éventuel moyen terme. Avec l'approximation qui caractérise l'intercompréhension, les mots peuvent être plus ou moins transparents, selon que l'on est habitué aux principes de lecture imprécise ou non, donc selon la capacité que l'on a à jouer sur la forme et la signification des mots pour tenter de reconnaître un mot familier. Nous nous intéresserons dans cette section aux approches proposées par les membres des équipes EuroCom (en l'occurrence EuroComRom), Galatea et ICE, dans lesquelles on trouve une modélisation des différents moyens de comprendre des mots, qui, à première vue, ne sembleraient pas transparents à des lecteurs débutants.

Selon le programme EuroComRom, un texte rédigé en langue étrangère peut être « filtré » au moyen de sept tamis, qui permettent d'avoir pour point de départ tous les éléments exploitables par un lecteur au fait de la méthode (nous ne nous intéresserons pas ici aux cinquième et sixième tamis, qui traitent du niveau syntaxique). Les deux premiers tamis concernent les mots les plus accessibles, à savoir le vocabulaire international et le vocabulaire pan-roman. Le vocabulaire international (Meissner *et al.* 2004 : 31-32) est constitué des « cultismes », c'est-à-dire des « mots d'origine latine, grecque et quelque fois hébreux appartenant à la république polyglotte européenne des lettres » (tels que « les descendants de l'étymon grec *ἐνθουσιασμός* : lat. *enthusiasmus*, angl. *enthusiasm*, fr. *enthousiasme*, it./esp. *entusiasmo*, all. *Enthusiasmus* ou du latin *Genius* : fr. *génie*, prov. *genh*, all. *Genius/Genie*, angl. *genius*, esp./it. *genio...* »), des « scientismes », qui sont des « créations savantes internationales et monosémiques formées sur une base grecque ou latine » (ex. : oxydation, atomiser), des « semi-scientismes », de création relativement récente, qui sont des « néo-latinismes, composés souvent d'éléments gréco, [et qui] n'ont pas existé dans la langue de Rome » (ex. : lubrification, libéralisme, succès), des « exotismes de signification », qui « désignent des phénomènes typiques pour une culture donnée » (ex. : cosaque, vodka) et des « créations de la vie moderne », du type « *anorak, cigarette, latex, PC [...]* ». Le vocabulaire pan-roman est constitué des mêmes types de mots, mais qui ne sont partagés que par les langues romanes.

Les troisième, quatrième et septième tamis donnent des règles d'équivalences entre langues qui permettent d'« identifier la parenté lexicale » (*ibid.* : 69) là où un lecteur

totallement débutant ne verrait qu'un mot opaque (un mot « de profil » selon la terminologie de l'équipe). Pour les mots concernés par le troisième tamis, les auteurs de *EuroComRom – Die sieben Siebe* déclarent (mais cela s'appliquerait tout autant, selon nous, aux mots concernés par les quatrième et septième tamis) :

*Die Ähnlichkeit der Wörter ist evident, dennoch ist ihre Kostümierung, die sie im Laufe ihrer sprachlichen Entwicklung erfahren haben, bisweilen so weit gegangen, daß man einige Zeit benötigt, um ein verkleidetes Wort wiederzuerkennen.*³⁵ (Klein & Stegmann 2000 : 62)

Le troisième tamis traite des correspondances phonologiques. Les auteurs donnent les (groupes de) phonèmes qui, dans les autres langues, correspondent à certains (groupes de) phonèmes dans chaque langue considérée. L'apprenant doit ainsi se familiariser avec les régularités du type « it. *notte*, fr. *nuit*, esp. *noche*, pg. *noite* d'une part et it. *latte*, fr. *lait*, cat. *llet*, esp. *leche*, pg. *leite* de l'autre » (Meissner *et al.* 2004 : 68). Une fois qu'il a pris connaissance de ces règles, il est censé être capable de faire des modifications sur le signifiant des mots pour aboutir au signifiant de mots qu'il connaît dans une langue qu'il maîtrise. Le quatrième tamis traite des graphies et prononciations, c'est-à-dire des différences qui existent entre les graphies des langues considérées. Les auteurs mettent pour cela à la disposition des apprenants un tableau récapitulatif des différentes prononciations possibles, pour chaque langue, des principales graphies. Enfin, le septième tamis donne les équivalences interlinguistiques des préfixes et suffixes.

Les troisième, quatrième et septième tamis fournissent donc aux apprenants de la méthode EuroCom (*EuroComRom* en l'occurrence) des outils, sous forme de règles systématisées, pour jouer sur la forme des mots non immédiatement identifiables qu'ils rencontrent, de façon à ce qu'ils puissent leur appliquer les équivalences interlinguistiques qui feront apparaître un mot susceptible d'être reconnu. Les mots identifiables par les apprenants ayant suivi la méthode sont donc d'une part les mots immédiatement transparents (appartenant aux vocabulaires international et pan-roman) et d'autre part les mots qui, sous réserve de quelques modifications sur le signifiant, suggèrent un mot connu. Les sept tamis ne proposent, au niveau lexical, que des règles de modification du

³⁵ Nous traduisons : « La ressemblance entre les mots est évidente. Cependant, les déguisements dont ils se sont parés au cours de leur développement linguistique ont parfois connu une évolution telle qu'on a besoin d'un peu de temps pour reconnaître un mot déguisé. »

signifiant. Il est en effet signalé que les modifications du signifié ne sont pas étudiées, pour la raison exposée avant un tableau présentant le vocabulaire pan-roman :

Die Bedeutungen der Wörter einer Zeile ist [sic] nicht immer in allen Sprachen exakt gleich geblieben. (Dies gilt auch für alle nachfolgenden Tabellen.) Mit leichten Bedeutungsnuancierungen muß öfter gerechnet werden. Für unser Ziel des annäherungsweisen Erschließens wird dies aber nur selten zum Stolperstein.³⁶
(Klein & Stegmann 2000 : 41)

C. Degache, membre de l'équipe Galatea, présente quant à lui, dans son mémoire d'HDR, les différents types d'« analogies » existant entre deux congénères (non obligatoirement reliés étymologiquement), en prenant en compte les dimensions formelle et sémantique. Dans le tableau que nous reproduisons ci-après, les analogies sont classées en quatre catégories (Degache 2006 : 63-65) : une « analogie totale » représente une « correspondance morpho-sémantique totale », une « analogie partielle » représente une « correspondance morpho-sémantique partielle », une « analogie marginale » représente une « correspondance morphologique totale ou partielle entre items sémantiquement éloignés et difficilement assimilables » et une « analogie trompeuse » représente une « correspondance morphologique totale ou partielle entre items sémantiquement divergents mais facilement assimilables ».

Légende³⁷ :

LS = langue source

LC = *langue cible*, en italiques

≠: ne signifie pas, est différent de

(...) : traduction de l'unité en langue source, (...||...) : deux acceptions différentes

> : à rapprocher de... (par différents moyens : étymologie, glissement sémantique)

³⁶ Nous traduisons : « Les significations des mots d'une ligne ne sont pas toujours restées exactement identiques dans toutes les langues. (Cela vaut également pour tous les tableaux suivants.) Il faut assez souvent compter avec de légères nuances de signification. Pour notre objectif de compréhension approximative, cela ne constitue toutefois que rarement une difficulté. »

³⁷ Présentée p. 63 (nous ne la mettons pas en retrait ni en italiques comme les autres citations car il importe ici que la mise en forme soit exactement celle d'origine).

Types d'analogie	Catégories de cognats (ou congénères)	Propriétés	Exemples pour l'espagnol langue cible (LC) pour francophones (français langue source LS)	
Analogies totales ou quasi-totales (accent...)	Emprunts intégraux	Forme graphique (quasi-) originale	<i>boutique, menu, puré, tisé, carné, ciprés, barón, capó</i>	
	Congénères homographes	Formes graphiques identiques	Mots en <i>-ante (arrogante)</i> , en <i>-al (ancestral)</i> , verbes en <i>-ir (venir, dormir, sentir...)</i>	
Analogies partielles	Emprunts intégrés	Acceptions perdues lors de l'emprunt	<i>televisión, avión, patata, cacao, tabaco, tomate, hamaca</i>	
	Congénères paragrammes	- voie philologique	<i>bonbon</i> (chocolat), <i>chance</i> (occasion), <i>michelin</i> (pli adipeux), <i>culotte</i> (cuissard cycliste)	
		- médiation allomorphique	<i>diente</i> (dent), <i>llama</i> (flamme), <i>campo</i> (champ), <i>señor</i> (seigneur), <i>oreja</i> (oreille), <i>nilo</i> (fil), <i>leche</i> (lait), <i>milagro</i> (miracle)	
		- médiation archaïque,	<i>estrella</i> (stellaire>étoile), <i>agua</i> (aquatique>eau), <i>peso</i> (peser> poids),	
		- recomposition dérivationnelle	<i>casa</i> (casanier>maison), <i>propiciar</i> (propice > fournir), <i>poder</i> (potentiel>pouvoir)	
			<i>dinero</i> (denier>argent), <i>ira</i> (ire>colère)	
			<i>increíble</i> (incroyable), <i>sugerencia</i> (suggestion), <i>provisional</i> (provisoire), <i>coronación</i> (couronnement)	
		Congénères homonymes	Homonymie du congénère en LS	<i>farsa</i> (farse : comédie hachis), <i>quitar</i> (quitter : ôter, enlever laisser, abandonner)
		Para-congénères	Traduction littérale + ajustement sémantique dans la famille de mots et le champ analogique	<i>noticia</i> (notice>document>information>une nouvelle)
		Congénères improbables	Même catégorie grammaticale, contextes différenciés	<i>receta</i> (recette de cuisine>recette de médecine>une ordonnance) regalo (régal>plaisir...d'offrir>un cadeau) <i>dinero</i> (argent)>dîner, <i>partición</i> (partage)>partition, retrato (portrait)>retraite, <i>mancha</i> (tache)>manche, <i>asomar</i> (se pencher)>assomer, <i>dato</i> (donnée)>date, <i>compás</i> (cadence, tempo)>compas, <i>sable</i> (sabre)>sable, <i>plátano</i> (banane)>platane, <i>gafas</i> (lunettes)>gaffes
Analogies marginales (peu probables et/ou peu conséquentes)	Congénères divergents	Acception LC très éloignée du sens de base en LS, ou homonymes en LC	<i>escribano</i> (noiaire)>écrivain, <i>matricula</i> (inscription univ)>matricule, <i>mesa</i> (table)>messe,	
	Congénères polysémiques en LC	Même catégorie grammaticale, mais glissement sémantique possible	<i>exprimir</i> (exprimer presser), <i>casco</i> (casque quartier, cf. <i>casco viejo</i>), <i>equipo</i> (équipement équipement), <i>esperar</i> (espérer attendre), <i>probar</i> (prouver essayer, goûter), <i>curso</i> (cours année scolaire), <i>facilitar</i> (faciliter, rendre facile fournir, remettre)	
	(Quasi)-homographes fortuits	Catégories grammaticales distinctes ou identiques	<i>cañe</i> (lit) > quatre, <i>roto</i> (cassé) > rôti, <i>plata</i> (argent)>plate, <i>planchar</i> (repasser)>plancher, <i>caracol</i> (escargot)>caracole(r)	
			<i>col</i> (chou)>col, <i>timbre</i> (sonnette)>timbre, <i>gato</i> (chat)>gâteau, <i>salir</i> (sortir)>salir, <i>subir</i> (monter)>subir	

Analogies trompeuses	Emprunts divergents en LS ou LC	Commutable dans même champ sémantique	<i>chalé</i> (pavillon, villa), <i>disparate</i> (adj.=hétérogène ≠ nom=abus, excès, divagation), <i>chifón</i> (tissu d'aspect chiffonné), <i>chandal</i> (survêtement) <i>pasaje</i> (billet)≠passage>voyage, <i>quitar</i> (enlever interdire: "el médico le ha quitado la sal") <i>depresión</i> , <i>película</i> (film)≠pellicule, bobine (<i>rollo</i>), clase (classe) leçon: "no quiero perder esta clase" <i>por lo tanto</i> (par conséquent pourant), <i>con respecto a</i> (par rapport à≠avec respect), <i>después</i> (après≠depuis), <i>al principio</i> (au début≠en principe)
	Congénères périphériques	Analogie sur acception périphérique en LS, trompeuse si contexte ambigu	<i>embestir</i> (emboutir)≠embêter, <i>biscocho</i> (biscuit)≠biscotte, <i>tabla</i> (planche)≠table, <i>asombrar</i> (étonner)≠assombrir, <i>plancha</i> (fer à repasser)≠planche, <i>divisar</i> (apercevoir)≠diviser, <i>concurrentia</i> (affluence)≠ <i>competencia</i> (concurrence), <i>consumar</i> (consommer)≠ <i>consumir</i> (consumer), <i>constipado</i> (enrhumé)≠constipé, <i>débil</i> (faible)≠débile
	Congénères discursifs	Articulateurs du discours, relateurs	<i>competencia</i> (compétence concurrence), <i>embarazada</i> (embarassée enceinte), <i>inversión</i> (inversion investissement), <i>partir</i> (partir partager, diviser), <i>contestar</i> (contester répondre), <i>oposición</i> (opposition concours), <i>compromiso</i> (compromis engagement), <i>doblar</i> (doubler les revenus, un cap plier, tourner à gauche ou à droite),
	Congénères sémiques	Mêmes catégorie grammaticale et champ sémantique + traits sémiques – acception commune	
	Congénères sémantiques Polysémie du congénère en LS ou LC	Mêmes catégorie grammaticale et champ sémantique + au moins une acception commune	

Ce classement présente l'avantage de distinguer ce qui relève des différences morphologiques et ce qui relève des différences sémantiques. Le souci de finesse dans la description des phénomènes entraîne cependant parfois des découpages peu évidents. Le cas de « *retrato* (portrait) ≠ retraite » (analogies marginales/congénères improbables/même catégorie grammaticale, contextes différenciés) paraît par exemple similaire à celui de « *mesa* (table) ≠ messe » (analogies marginales/congénères divergents/acception LC très éloignée du sens de base en LS). De plus, il est fait appel au critère étymologique pour différencier certains cas de figure par ailleurs identiques (par exemple pour différencier « *boutique* » (analogies totales ou quasi-totales/emprunts intégraux/forme graphique (quasi) originale) de « *cacao* » (analogies totales ou quasi-totales/congénères homographes/formes graphiques identiques)), alors que ce critère n'implique rien pour l'analogie actuelle et qu'il n'est pas connu des apprenants.

Retenons de ce classement que certains mots étrangers sont parfaitement transparents, que d'autres demandent des manipulations formelles (« voie philologique », « recomposition dérivationnelle ») pour être identifiés (par ex. *campo* (champ)), et que d'autres encore demandent un passage par un autre mot en LS (« médiation allomorphique », « médiation archaïque », « ajustement sémantique dans la famille de mots et le champ analogique ») pour être correctement compris (par ex. *dinero* (denier>argent)).

É. Castagne, membre de l'équipe ICE, dans son article « Transparences lexicales entre langues voisines », propose une typologie des différents types de transparence. Il distingue dans un premier temps les mots « transparents » des mots « non transparents » :

[...] nous appellerons « mot transparent » tout terme lexical, présent dans un texte en une LE donnée qui d'une part possèdera une forme qui suggère un terme lexical en LM de forme (écrite ou orale) quasi identique, et qui d'autre part se trouve posséder une signification quasi-identique dans les deux langues. [...]

Nous appellerons « mot non transparent » tout terme lexical présent dans un texte en LE qui peut être jugé illisible par un lecteur « débutant » dans cette même LE à partir des connaissances extraites dans sa LM. (Castagne 2007a : 157-158)

Sont par exemple transparents les substantifs PT *universidade*, ES *universidad*, IT *università*, FR *université*, GB *university*, NL *universiteit* et DE *Universität*. Les mots non transparents se subdivisent en deux classes distinctes, les mots « opaques » et les mots « semi-transparentes » :

Les « mots opaques » [...] sont des termes lexicaux (présents dans un texte en une LE donnée) qui possèdent une forme qui, soit ne suggère aucun terme lexical en LM, soit suggère un terme lexical mais qui possède une signification autre. (ibid. : 158)

Par exemple, le mot NL *zolder* (FR « grenier ») ne suggère aucun terme lexical en LM (en français en l'occurrence), et le mot IT *alcune* (FR « quelques ») suggère un terme lexical en LM mais qui possède une signification autre (on parle alors traditionnellement de faux amis).

Les « mots semi-transparentes » sont les mots qui ne sont pas jugés immédiatement lisibles par un lecteur « débutant » dans une LE à partir des connaissances extraites dans sa LM, mais ils comportent un (des) indice(s) permettant d'en comprendre la signification au moyen d'un « mot-relais ». (ibid. : 159)

La catégorie des mots semi-transparentes se subdivise également en deux sous-types, selon le type de mot-relais. Selon le premier sous-type :

Le terme de la LE suggère formellement un terme de la LM dont la signification, même si elle est autre, est tout de même accessible en tenant compte des autres significations présentes dans le contexte : le terme fonctionne alors comme une sorte d'approximation appelant un dépassement par substitution avec un terme ajusté au contexte. Le « mot-relais » est dans ce cas un mot réel de la LM. (ibid. : 160)

Par exemple, le mot GB *umbrella* peut faire penser au mot FR « parapluie » par l'intermédiaire du mot-relais FR « ombrelle ».

Selon le deuxième sous-type :

Le terme de la LE ne suggère, de par sa forme stricte, aucun terme de la LM, mais peut en suggérer un à condition de faire varier sa forme (ce qui nous éloigne du mot originel) en affaiblissant ici ou là la précision de la réalisation de tel ou tel de ses graphèmes et/ou de tel ou tel de ses phonèmes jusqu'à faire apparaître un terme imaginaire de la LE qui fonctionne alors comme un mot transparent. Le mot-relais est dans ce cas un terme imaginaire de la LE. (ibid. : 160)

Par exemple, le mot NL *miljoen* peut faire penser au mot FR « million » par l'intermédiaire du mot-relais *milion.

Pour ces deux sous-types de mots semi-transparents, É. Castagne parle de « transparence indirecte », puisque dans un cas, le mot en LE suggère formellement un mot en LM, qui, à son tour, suggère un autre mot en LM par la prise en compte du contexte, et dans l'autre cas, le mot en LE, après une médiation formelle, suggère un mot en LM. Les différents types de transparences sont présentés dans le tableau récapitulatif suivant (p. 161) :

mots transparents	mots non-transparents			
	mots semi-transparents		mots opaques	
<i>transparence directe</i>	<i>transparence indirecte</i>		<i>opacité</i>	
	Sous-type 1 le relais effectue une médiation sémantique au sein de la LM (faux-amis solubles)	Sous-type 2 le relais effectue une médiation formelle au sein de la LE	Sous-type 1 ne suggère aucun terme lexical en LM	Sous-type 2 suggère un terme lexical en LM mais qui possède une signification autre (faux-amis non solubles hors contexte)
universidade (PT) universidad (ES) università (IT) university (GB) universiteit (NL) Universität (DE)	umbrella (GB) ombrelle (FR) parapluie (FR)	miljoen (NL) *million million (FR)	zolder (NL) grenier (FR)	alcune (IT) ?aucune quelques (FR)

Grâce à la pratique de l'intercompréhension, les mécanismes de transparence indirecte sont de plus en plus facilement mis en œuvre, si bien que les mots de LE requérant un passage par un mot-relais, pour *semi-transparents* qu'ils soient, sont tout de même transparents. Le tableau devient donc le suivant :

<i>transparence</i>		<i>opacité</i>		
mots transparents	mots semi-transparentes		mots opaques	
	Sous-type 1 le relais effectue une médiation sémantique au sein de la LM (faux-amis solubles)	Sous-type 2 le relais effectue une médiation formelle au sein de la LE	Sous-type 1 ne suggère aucun terme lexical en LM	Sous-type 2 suggère un terme lexical en LM mais qui possède une signification autre (faux-amis non solubles hors contexte)
universidade (PT) universidad (ES) università (IT) university (GB) universiteit (NL) Universität (DE)	umbrella (GB) ombrelle (FR) parapluie (FR)	miljoen (NL) *million million (FR)	zolder (NL) grenier (FR)	alcune (IT) ?aucune quelques (FR)

Le nombre de mots transparents, identifiables par un apprenant ICE, augmente grâce à la maîtrise de certains procédés. Il y a donc des mots qui sont transparents immédiatement, et, après un temps de formation à l'intercompréhension, d'autres qui le deviennent :

La pratique de la forme de plurilinguisme que nous préconisons, que l'on qualifie souvent d'asymétrique, (chacun parle sa langue maternelle et comprend plusieurs langues apparentées ou voisines), nourrit et dynamise continuellement le système de reconnaissance des liens de parenté et de voisinage. Plus on pratique et plus on découvre des transparences là où dans un premier temps on ne les avait pas repérées. (ibid. : 164)

Les trois nouvelles approches de la transparence lexicale que nous venons de voir, à savoir celles proposées par les équipes d'EuroCom, Galatea et ICE proposent toutes, malgré des différences importantes, une vision ternaire de la transparence : il y a les mots qui sont assurément transparents, ceux qui assurément ne le sont pas, et entre ces deux pôles, les mots qui peuvent devenir transparents grâce à quelques ajustements. Ces ajustements peuvent prendre la forme de correspondances formelles préalablement apprises, d'analogies morphologiques et sémantiques ou de mots-relais.

Cette première partie a été pour nous l'occasion de présenter différents travaux ayant trait à la notion de transparence lexicale. Les premiers font appel à cette notion en

tant que qualité attribuable aux mots immédiatement compréhensibles par des lecteurs débutants, c'est-à-dire, selon les termes utilisés dans ces travaux, aux mots possédant une identité formelle et sémantique ou une quasi-identité formelle et sémantique avec un mot de la langue maternelle du lecteur. Les travaux que nous avons présentés à la fin de cette partie prennent en compte plusieurs degrés de transparence. Ces travaux sont le fruit de réflexions menées par des équipes travaillant sur le concept d'intercompréhension et sur la formation à l'intercompréhension. Ces équipes ont observé au cours de séances de formation que les apprenants pouvaient s'aider non seulement des mots identiques ou quasi-identiques aux mots de leur langue, mais aussi des mots dont la proximité avec des mots de leur langue était moins évidente. En tirant bénéfice de ces dernières approches, nous allons, dans la partie suivante, proposer une nouvelle caractérisation de la transparence en ayant auparavant précisé quels sont, selon nous, les critères à respecter pour l'établissement de cette caractérisation.

2. La transparence lexicale

Nous allons, dans cette deuxième partie, nous attacher à proposer une caractérisation précise de la transparence lexicale. Nous procéderons pour cela en trois temps. Le premier chapitre (2.1. Pour une révision de la notion de transparence) constitue le chapitre introductif : nous y développerons les points dont l'éclaircissement nous paraît nécessaire à l'exposé de ce que nous entendons par « transparence lexicale ». Nous présenterons dans le deuxième chapitre (2.2. La transparence directe) notre caractérisation de la transparence conçue comme qualité des mots immédiatement compréhensibles et verrons dans le troisième chapitre (2.3. La transparence indirecte) que cette première caractérisation peut être « étendue », de sorte à concevoir la transparence comme qualité de tous les mots compréhensibles, y compris ceux qui demandent un ou plusieurs ajustements pour être compris.

2.1. Pour une révision de la notion de transparence

Nous commencerons ce chapitre en dressant le bilan des différentes définitions de la transparence lexicale que nous avons présentées dans la première partie. Nous expliquerons pourquoi, selon nous, certains points peuvent être retenus tels quels et d'autres complétés ou laissés de côté. Nous présenterons ensuite la modélisation théorique au sein de laquelle s'organise la transparence telle que nous la concevons, et exposerons enfin pourquoi le versant formel et le versant sémantique de la transparence ne constituent pas deux aspects indépendants du phénomène.

2.1.1. De nouveaux objectifs

Nous avons vu que la transparence lexicale est traditionnellement caractérisée par une proximité formelle *et* sémantique entre deux mots. Définir cette proximité par « deux formes identiques ou quasi-identiques » et « deux significations identiques ou quasi-identiques » manque toutefois de rigueur. Il convient donc d'établir des critères objectifs qui permettent de fixer une limite entre des formes qui sont suffisamment proches l'une de l'autre pour être considérées comme quasi-identiques et celles qui sont trop éloignées pour pouvoir l'être, et entre des significations qui sont suffisamment proches l'une de l'autre

pour être considérées comme quasi-identiques et celles qui sont trop éloignées pour pouvoir l'être.

Comme nous l'avons vu dans la section 1.2.3., certains programmes d'intercompréhension ont proposé une approche plus fine de la transparence que cette approche binaire (transparence ou opacité), en intégrant une zone intermédiaire constituée des mots qui ne sont pas suffisamment proches pour être identifiés de prime abord mais pas suffisamment éloignés pour être considérés comme opaques. Cette approche permet de donner une définition intéressante et originale de la transparence car elle rend compte de tous les mots effectivement compréhensibles par des lecteurs. Nous choisissons ainsi de caractériser la transparence en deux temps : par les critères objectifs qui définissent les proximités formelle et sémantique les plus évidentes dans un premier temps, et par les critères objectifs qui définissent les proximités formelle et sémantique un peu plus éloignées dans un second.

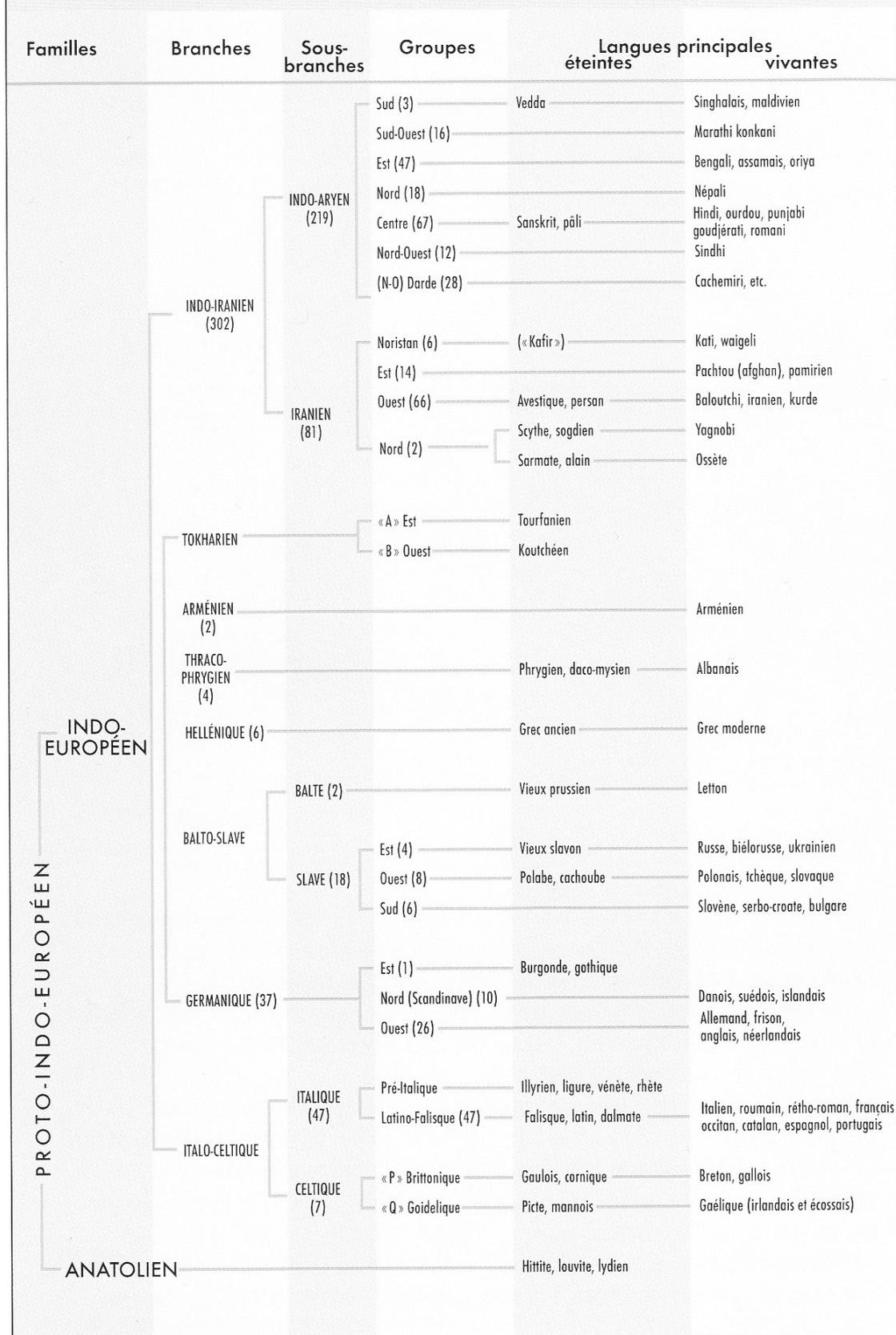
Un premier critère nous semble capital : nous définirons la transparence en synchronie, et non par référence à des données étymologiques. Comme le signale Saussure, les études diachroniques aident certes à justifier un état de fait, mais ne sont d'aucune utilité pour le décrire :

La première chose qui frappe quand on étudie les faits de langue, c'est que pour le sujet parlant leur succession dans le temps est inexistante : il est devant un état. Aussi le linguiste qui veut comprendre cet état doit-il faire table rase de tout ce qui l'a produit et ignorer la diachronie. Il ne peut entrer dans la conscience des sujets parlants qu'en supprimant le passé. (Saussure 1995 : 117)

Certes, pour les langues qui nous intéressent (le français, l'anglais, l'allemand et le néerlandais), nous savons que les mots qui sont les plus facilement identifiés par les apprenants tirent leur proximité d'un étymon commun ou d'une relation d'emprunt. Ces langues appartiennent en effet toutes à la famille des langues indo-européennes, présentée ci-après sous forme arborescente³⁸ :

³⁸ Nous choisissons pour sa lisibilité l'arbre proposé dans l'*Atlas des langues du monde* de R. Breton comme exemple d'arbre de parenté des langues indo-européennes.

EXEMPLE D'ARBRE DE PARENTÉ : LA FAMILLE INDO-EUROPÉENNE



(Breton 2003 : 12)

De ce fait, les mots FR « tonnerre », EN *thunder*, DE *Donner* et NL *donder* par exemple présentent des similitudes de forme et de signification par leur racine indo-européenne **tn-*, **ton-* (Onions 1996 : 921).

De surcroît, le français, l'anglais, l'allemand et le néerlandais possèdent des éléments lexicaux en commun grâce aux nombreux liens qu'ont entretenus au fil du temps la langue romane et les langues germaniques au cours de leur évolution (Raynaud 1993 ; Crépin 1994 ; König & Van Der Auwera 1994 ; Walter 1994), liens dont nous allons rappeler brièvement, car nous tenons expressément dans cette étude à ne pas expliquer les liens interlinguistiques par la diachronie, la chronologie.

Le latin, puis ce qui allait devenir le français, a connu de nombreux contacts avec les langues germaniques :

- aux alentours du III^e siècle, des peuples de langue germanique arrivent en Gaule, où le latin est déjà bien installé ;
- au V^e siècle, Clovis se convertit au catholicisme, puis, progressivement, tous les Francs, entraînant ainsi un bilinguisme germanique/latin ;
- au XIII^e siècle, le français emprunte abondamment au néerlandais, en raison d'échanges importants entre la France et les Pays-Bas ;
- enfin, les emprunts du français à l'anglais, qui ont commencé au XVIII^e siècle, sont extrêmement nombreux depuis la deuxième moitié du XX^e siècle.

Les langues germaniques ont également été sous l'influence du latin puis du français :

- au I^{er} siècle av. J-C., les Romains conquièrent l'Europe occidentale, amenant avec eux quantité de mots d'origine latine ;
- parallèlement, les Germains avancent jusqu'à l'Empire romain, et empruntent alors les premiers mots au latin parlé ;
- les Romains restent en Angleterre jusqu'au début du V^e siècle, mais le latin n'a alors pas pénétré jusque dans les couches populaires, où on parle toujours celte ;
- au milieu du V^e siècle, les premiers Anglo-Saxons débarquent sur l'île, introduisant ainsi dans le peuple les premiers éléments du lexique latin qu'ils avaient eux-mêmes empruntés au contact des Romains ;
- un nouvel afflux de termes latins arrive en Angleterre à la fin du VI^e siècle, puis au VII^e siècle dans la partie nord-ouest du territoire européen où on parle des langues germaniques, par l'expansion de la chrétienté ;

- en 1066, la victoire de Guillaume le Normand à Hastings donne au français, sous la forme du normand, un rôle de première importance en Angleterre, qui est seul parlé à la cour et qui reste la langue du pouvoir, y compris économique, jusqu'à la fin du XIII^e siècle ;
- au XIII^e siècle, les relations politiques et de commerce entre les Pays-Bas et la France vont en s'intensifiant, si bien que le néerlandais emprunte au français non seulement des mots mais aussi des constructions, des morphèmes, donnant ainsi à des mots d'origine germanique des allures de langue romane ;
- après la révocation de l'Édit de Nantes, en 1685, des dizaines de milliers de Huguenots français se réfugient aux Pays-Bas, diffusant leur vocabulaire dans le peuple ;
- au XVIII^e siècle, le français, du fait de son prestige culturel, exporte une quantité très importante de mots en Angleterre, en Allemagne et aux Pays-Bas notamment, où ce siècle est appelé *pruikentijd* (« époque des perruques ») parce qu'il était dominé par tout ce qui venait de France, et où de nouvelles formes « à la française » sont créées ;
- enfin, le latin est resté la langue de l'écrit dans toute l'Europe intellectuelle au moins jusqu'à la fin de la Renaissance.

Cette origine lointaine commune et ces emprunts que se sont faits les langues les unes aux autres plus ou moins récemment sont bien entendu la cause d'un nombre certain de transparences exploitables par les apprenants dans leur processus de compréhension. H. Walter note d'ailleurs que le français s'est construit tout autant par ses différents échanges que par filiation avec le latin :

Si l'on veut caractériser d'une phrase la langue française, on peut dire que c'est la plus germanique des langues romanes. [...]

En fait, l'histoire du français est à la fois celle de l'évolution du latin parlé en Gaule et celle d'un enrichissement constant au contact de langues voisines. (Walter 1994 : 225)

Elle rappelle, dans la conclusion de l'ouvrage cité, les liens particuliers qui unissent le français et l'anglais³⁹ :

³⁹ Voir aussi à ce propos son ouvrage *Honni soit qui mal y pense : l'incroyable histoire d'amour entre le français et l'anglais*, paru en 2001.

Le français et l'anglais, par exemple, ont même connu des relations si intimes et si fréquentes qu'on peut suivre dans leur vocabulaire comme une longue histoire d'amour entre la plus latine des langues germaniques – l'anglais – et la plus germanique des langues romanes – le français. Le mouvement s'est longtemps exercé à sens unique, avec un afflux par milliers de mots français en anglais, mais il se trouve aujourd'hui complètement inversé, avec des apports constants de mots anglais en français. (ibid. : 419)

L'anglais est si fortement teinté de français et de racines latines que la compréhension de textes rédigés dans cette langue est grandement facilitée par la présence de mots transparents pour des francophones. De plus, puisque l'anglais est la langue à laquelle les autres langues empruntent le plus aujourd'hui (*ibid.* : 269, 411-412), bon nombre de mots sont également transparents pour des francophones dans des textes rédigés dans les deux autres langues qui nous intéressent, à savoir l'allemand et le néerlandais.

Il convient donc d'avoir à l'esprit cette histoire mêlée de constants échanges entre le français, l'anglais, l'allemand et le néerlandais, mais de savoir aussi l'oublier pour caractériser la transparence, qui doit, selon nos objectifs, être définie en synchronie.

Nous allons maintenant revenir sur les propositions de définitions de la transparence exposées dans la première partie de ce travail, et tenter de préciser en quoi certains points nous paraissent intéressants et d'autres non recevables, pour les objectifs que nous nous sommes fixés.

Nous allons d'abord nous concentrer sur le versant formel de la transparence. Dès lors que l'on veut prendre en compte les formes qui ne sont pas exactement identiques, il faut pouvoir déterminer une limite entre les modifications (et/ou le nombre de modifications) qui n'altèrent pas la reconnaissance et celles (et/ou leur nombre) qui causent un trop grand éloignement pour que deux formes soient considérées comme proches. Les méthodes de comparaison de chaînes que nous avons exposées, telles que le coefficient de Dice ou la distance de Levenshtein (voir section 1.1.3.), sont précises et objectives. Elles permettent en outre d'accorder une valeur, un « taux de transparence » aux mots considérés et ainsi de les classer les uns par rapport aux autres. Néanmoins, elles empêchent toute étude quantitative par l'absence d'un seuil qui déterminerait la valeur au-dessus de laquelle un mot serait considéré comme transparent. D'autres études proposent précisément des

critères qui permettent de parler ou non de transparence : une différence d'une seule lettre est acceptée entre les deux mots, ils doivent commencer par les quatre mêmes premières lettres, etc. Ces critères, même s'ils sont pertinents – là n'est pas la question – ont été édictés par les auteurs de ces études sans justification aucune. On peut donc leur reprocher leur subjectivité. La transparence, dans son versant formel, doit donc, selon nous, être définie par des règles justifiées, en synchronie, et qui permettent de fixer le « seuil de proximité » (des mots donnés seraient alors considérés comme transparents si et seulement si ce seuil est atteint). Pour l'approche plus souple, telle qu'elle est préconisée dans le cadre de l'intercompréhension, il convient que la détermination des ajustements à faire sur la forme des mots suive les mêmes règles, c'est-à-dire qu'elle soit justifiée, qu'elle s'appuie sur la synchronie et qu'elle permette de fixer un seuil d'une part entre les mots qui sont très proches formellement et ceux qui ont une ressemblance moins évidente, et d'autre part entre ces mots qui ont une ressemblance non immédiatement perceptible et ceux qui sont opaques. Les équivalences des troisième, quatrième et septième tamis d'EuroCom sont fournies en l'état, sans être justifiées ; elles ne répondent donc pas aux objectifs que nous nous sommes fixés. Nous ne nous appuyerons pas non plus sur des équivalences entre consonnes qui pourraient être données entre le français, l'anglais, l'allemand et le néerlandais, et qui tireraient leur justification des mutations consonantiques (la première fondant des équivalences entre le français d'une part et les langues germaniques de l'autre, la seconde entre l'allemand d'une part et les deux autres langues germaniques de l'autre), car les équivalences doivent pouvoir être justifiées en synchronie pour rendre compte de la capacité des apprenants à les découvrir sans aucune connaissance préalable en linguistique historique.

Le versant sémantique de la transparence est peut-être moins aisé à caractériser, car les éléments à comparer ne sont pas là deux observables. La plupart du temps, on juge de la proximité de signification entre deux mots sur la base de liens étymologiques. Nous avons déjà indiqué ci-dessus notre choix délibéré de ne pas prendre en compte la diachronie. Il s'agit donc de trouver un moyen de définir la proximité sémantique en synchronie et sans jugement subjectif, et pour la version « étendue » de la transparence, de définir précisément une proximité de signification plus étendue, c'est-à-dire ce que C. Degache entend par « ajustement sémantique dans la famille de mots et le champ analogique », « glissement sémantique » (Degache 2006 : 63), et ce qu'É. Castagne entend par « terme de la LM dont la signification, même si elle est autre, est tout de même

accessible en tenant compte des autres significations présentes dans le contexte » (Castagne 2007a : 160).

Nous allons, dans les chapitres suivants, faire des propositions qui tenteront de répondre aux problèmes soulevés ici. Nous définirons ainsi dans un premier temps la transparence la plus immédiate (2.2. La transparence directe) et dans un second temps la transparence étendue, qui demande des habitudes de lecture telles qu'elles sont données aux apprenants en intercompréhension (2.3. La transparence indirecte). Nous élaborerons nos définitions en prenant garde à respecter les objectifs que nous venons de présenter. Les deux types de transparence seront définis du point de vue formel et du point de vue sémantique. Les critères que nous développerons seront, pour la transparence directe, la proximité formelle directe et la proximité sémantique directe, et, pour la transparence indirecte, la proximité formelle indirecte et la proximité sémantique indirecte (le qualificatif « indirect » indiquant que des ajustements sont nécessaires pour repérer ces proximités).

Ainsi, pour que l'on puisse parler de transparence directe, il faudra que soient respectés les critères de proximité formelle directe *et* ceux de proximité sémantique directe, et pour que l'on puisse parler de transparence indirecte, il faudra qu'il y ait i) proximité formelle directe et proximité sémantique indirecte, ou bien ii) proximité formelle indirecte et proximité sémantique directe, ou bien iii) proximité formelle indirecte et proximité sémantique indirecte.

Avant de proposer nos définitions de ces différents types de proximités, nous allons maintenant présenter le cadre théorique dans lequel s'inscrit notre caractérisation de la transparence.

2.1.2. De l'intersystème à l'hypersystème

Nous allons dans cette section consacrer un développement au passage entre les deux types de transparence, la transparence directe et la transparence indirecte. Nous dirons ainsi que les relations qu'entretiennent les mots directement transparents s'intègrent dans l'« intersystème » tandis que l'« hypersystème » rend compte des relations

qu'entretiennent les mots transparents directement et indirectement. C'est dans son sens d'expression de la « relation ou [de] la réciprocité » (Rey *et al.* 2006 : 1 858) qu'il faut comprendre ici le préfixe « inter- ». Il s'agit de s'intéresser, non pas au système linguistique d'*une seule* langue, mais à celui qui régit les relations entre *plusieurs* langues. L'intersystème est ainsi le système dans lequel s'organisent les différents éléments sinon identiques, du moins très proches, entre langues. Le préfixe « hyper- » est utilisé avec la signification qu'il avait en grec, à savoir « au-dessus, au-delà » (*ibid.* : 1 763). Dans l'hypersystème, il s'agit toujours de comparer des langues entre elles, mais en considérant pour la comparaison non seulement les éléments similaires ou très proches, mais aussi les éléments dont le rapprochement est possible grâce à des ajustements. Un plus grand nombre d'éléments est donc concerné. Au niveau lexical, la comparaison au niveau de l'intersystème porte donc sur la transparence directe, et la comparaison au niveau de l'hypersystème, en allant « au-delà », porte également sur la transparence indirecte.

Nous abordons l'intersystème à la lumière de la définition de l'« interlexème » formulée par l'équipe d'EuroCom :

L'interlexème européen est une unité de base du lexique plurilingue qui représente les (allo-)lexèmes de différentes langues qui, du point de vue du concept, sont équivalents et, du point de vue de la forme, se présentent comme congruents (très proches). (Klein & Rutke 1997)

L'interlexème n'est pas un mot transparent d'une langue étrangère, mais l'unité théorique supérieure qui « représente » les mots effectifs des langues, qui possède les traits communs aux différentes réalisations. Un interlexème ne rend donc pas compte obligatoirement de la transparence entre deux mots seulement, mais entre autant de mots, de langues différentes, compréhensibles par des débutants. Par exemple, on peut évoquer l'interlexème représentant les (allo-)lexèmes PT *comunicação*, ES *comunicación*, IT *comunicazione*, FR « communication », EN *communication*, DE *Kommunikation* et NL *communicatie*. L'ensemble des interlexèmes pour des langues données constitue leur « interlexique » (Meissner *et al.* 2004 : 39).

Nous proposons de développer, à partir des notions introduites par l'équipe d'EuroCom, un système théorique où les différents niveaux de comparaison lexicale

peuvent être décrits. Cet intersystème, qui décrit l'organisation des points de comparaison entre les mots directement transparents entre plusieurs langues, sera ensuite étendu en un hypersystème, qui décrit l'organisation des points de comparaison entre les mots directement et indirectement transparents entre plusieurs langues. Nous ne présentons ici que les cadres théoriques ; les définitions précises que nous formulons de la transparence directe et indirecte seront exposées en détail dans la suite de cette partie.

Notons que l'équipe d'EuroCom a déjà exploité le terme d'« intersystème » :

L'analyse [des « grammaires d'hypothèses plurilingues »] permet de localiser trois zones de transfert : (1) langue-cible, (2) langues activées et (3) systématique identifiable « entre » les langues activées. De par la systématique qui caractérise toute architecture de langue, les sujets identifient, dans chacune de ces zones, de nombreuses analogies « interconnectables ». Pour la zone inter-langues, nous parlons d'un « intersystème » (qui ne connaît, bien entendu, pas seulement des correspondances positives mais aussi des irrégularités : ex. fr. rails > esp. raíl ou riel (peu fréquent), ou carril (fréquent) dans l'espagnol d'Amérique. Dans une première tentative, les transferts s'appuient sur le lexique pour s'étendre ensuite à la morphosyntaxe et à la syntaxe. (Meissner 2008 : 19)

Présentons l'intersystème tel que nous l'entendons par comparaison à celui décrit dans la citation ci-dessus. L'intersystème tel que nous le concevons se distingue de celui décrit par F-J. Meissner en ce qu'il est un *modèle théorique* permettant de mettre en relation plusieurs aspects de langues différentes, et non l'ensemble des régularités entre langues empiriquement identifiées par des apprenants. Il s'agit d'un système descriptif, organisé et purement théorique des relations qui unissent les unités comparables entre langues, non une construction qui évolue à chaque instant ou une description des processus cognitifs des apprenants. Par conséquent, l'intersystème que nous allons développer rend compte de « correspondances » entre langues, mais pas d'« irrégularités ».

Chacune des notions que nous allons présenter sera nommée et développée, mais nous choisissons de ne pas proposer de système de notation pour les concepts introduits. Par exemple, pour l'interlexème dont les réalisations vues ci-dessus sont les lexèmes PT *comunicação*, ES *comunicación*, IT *comunicazione*, FR « communication », EN *communication*, DE *Kommunikation* et NL *communicatie*, nous ne donnons pas de symbole ou de suite de lettres communes pour le représenter. Il en ira de même pour toutes

les notions vues dans cette section, tant pour l'intersystème que pour l'hypersystème. Cette solution nous paraît préférable pour ne pas rendre notre texte d'une lecture difficile, et parce que l'élaboration d'un système de notation est inutile pour notre propos.

Un interlexème est une unité théorique supérieure à des lexèmes donnés de plusieurs langues. Il est la somme des points communs à ces différents lexèmes. Les notions d'« interforme » et d'« intersémème » sont utiles pour le décrire plus précisément.

Une interforme est une unité théorique supérieure à des formes données attestées dans des langues différentes. Elle rend compte du versant formel de la transparence directe entre plusieurs mots, sans prise en compte aucune du versant sémantique. Ainsi les mots FR « sein », DE *sein* (« être », « son ») et NL *sein* (« signal »), qui n'ont pas de lien sémantique, sont-ils les représentants d'une même interforme au même titre que les mots FR « bombe », EN *bomb* et DE *Bombe*, qui, eux, entretiennent un rapport sémantique. La notion d'interforme correspond donc au niveau interlinguistique aux notions intralinguistiques d'homonymie et de paronymie. Nous verrons dans la section 2.2.1. comment savoir si des mots sont suffisamment proches du point de vue de la forme pour pouvoir être mis en rapport avec une même interforme.

Un intersémème⁴⁰ est une unité théorique supérieure aux significations de mots donnés de langues différentes. Il rend compte du versant sémantique de la transparence directe entre plusieurs mots, sans prise en compte aucune du versant formel. Ainsi les mots FR « lunettes », EN *glasses*, et DE *Brille*, qui ont des formes différentes les unes des autres, sont-ils les représentants d'un même intersémème au même titre que les mots FR « train », EN *train*, et NL *trein*, qui, eux, sont proches du point de vue formel. La notion d'intersémème correspond donc au niveau interlinguistique aux notions intralinguistiques de synonymie et de parasynonymie. Nous verrons dans la section 2.2.2. comment savoir si des mots sont suffisamment proches du point de vue de la signification pour pouvoir être mis en rapport avec un même intersémème.

⁴⁰ P. Braun forge dans son article « Internationalismen », paru en 2002, le terme *Intersememe*. Il désigne par là l'acception commune à des mots de langues différentes, à des internationalismes en l'occurrence, qui ont plusieurs acceptions (Braun 2002 : 1381). Nous ne reprenons pas cette signification, mais intégrons l'intersémème en tant qu'unité de description théorique dans l'intersystème.

Nous posons ainsi que :

Interforme + intersémème = interlexème

En d'autres termes, si des mots de langues différentes ont des formes suffisamment proches les unes des autres pour que celles-ci soient rattachées à une même interforme, et des significations suffisamment proches les unes des autres pour que celles-ci soient rattachées à un même intersémème, alors ces mots peuvent être considérés comme diverses réalisations d'un même mot théorique, l'interlexème. L'ensemble des interlexèmes pour un nombre donné de langues constitue l'interlexique de ces langues.

On pourrait imaginer qu'une étude syntaxique fasse état de la notion d'« interphrasème » (par exemple), qui permettrait de mettre en relation des structures de syntagmes et de phrases communes à plusieurs langues ; nous laissons toutefois cet aspect de côté pour nous concentrer sur les phénomènes lexicaux.

Toutes les notions que nous venons de citer sont englobées dans celle d'« interlangue », composante la plus large de l'intersystème. Là encore, l'interlangue telle que nous la concevons est nettement différente de celle des didacticiens des langues étrangères. Cette dernière est le système interlinguistique que se construit l'apprenant au cours de son apprentissage, et est constituée de trois sous-systèmes :

- *une partie du système de la L1 ;*
 - *une partie du système de la langue-cible ;*
 - *un système de règles n'appartenant ni à l'un ni à l'autre de ces deux systèmes, donc spécifique du dialecte idiosyncrasique constitué à un moment donné.*
- (Gaonac'h 1991 : 125)

Dans notre système, au contraire, une interlangue est une langue théorique d'un niveau supérieur à des langues réelles données. L'interlangue en relation à des langues données est constituée des éléments qui sont communs, de tous les points de vue linguistiques, à ces langues. C'est cette interlangue qui rend possible la lecture en langue étrangère par des apprenants débutants, puisqu'elle s'illustre dans le texte à lire par des

éléments qui existent de façon identique ou quasi-identique dans la langue maternelle du lecteur.

L'intersystème que nous venons de décrire – constitué des notions d'interforme, d'intersémème, d'interlexème et d'interlangue – peut être étendu en un hypersystème, qui décrit les relations qu'entretiennent tous les éléments comparables entre langues. En passant du niveau inter- au niveau hyper-, on reste dans une comparaison interlinguistique mais en prenant en compte non seulement les éléments identiques ou quasi-identiques entre langues mais aussi ceux qui sont, pour en donner une caractérisation encore très imprécise, un peu plus éloignés. L'hypersystème rend compte *et* des éléments transparents directement *et* des éléments transparents indirectement. Il contient donc l'intersystème.

En prolongeant ce qui a été observé au sujet de l'interlexème, on peut concevoir un hyperlexème comme une unité théorique supérieure, à un plus haut degré, à des lexèmes donnés de plusieurs langues. Il est la somme d'une hyperforme et d'un hypersémème.

Pour que des formes attestées dans des langues puissent être mises en relation par le biais d'une hyperforme, il faut soit qu'elles soient très proches les unes des autres (auquel cas elles sont également les réalisations d'une interforme), soit qu'elles puissent, après quelques ajustements, être considérées comme proches. Nous verrons dans la section 2.3.1. quelles sont les règles qui définissent la proximité formelle indirecte. Les formes qui sont les réalisations d'une hyperforme mais non d'une interforme sont caractérisées par une ou plusieurs différences entre elles, surmontables par le recours à un « hyperphonème ». Sans entrer dans les détails que nous exposerons dans la section 2.3.1., nous proposons que, pour passer d'une forme à une autre, qui en est éloignée par un phonème différent, il faille passer au niveau de l'hyperphonème, unité phonologique théorique supérieure à des phonèmes donnés de langues différentes, pour retrouver une forme connue ou une forme proche d'une forme connue. L'hyperphonème est ainsi la notion interlinguistique correspondant à la notion intralinguistique d'archiphonème. N. S. Troubetzkoy définit ainsi l'archiphonème :

Dans les positions où une opposition neutralisable est effectivement neutralisée, les marques spécifiques d'un des termes de l'opposition perdent leur valeur phonologique et les traits que les deux termes ont en commun (c'est-à-dire la base

de comparaison de cette opposition) restent seuls pertinents. Dans la position de neutralisation, un des termes de l'opposition devient donc le représentant de l'« archiphonème » de cette opposition : par « archiphonème » nous entendons l'ensemble des particularités distinctives qui sont communes aux deux phonèmes. (Troubetzkoy 1970 : 81)

De même, un hyperphonème est l'ensemble des particularités distinctives qui sont communes à plusieurs phonèmes donnés. Par exemple, pour passer du phonème /b/ au phonème /v/, il convient de passer par l'hyperphonème (auquel, rappelons-le, nous ne donnons pas de dénomination) {'consonne', 'labial', 'sonore'}. Tout comme un phonème est la somme des traits pertinents de plusieurs sons effectivement produits et considérés comme des réalisations de ce phonème, et l'archiphonème la somme des traits pertinents de plusieurs phonèmes au sein d'une langue, l'hyperphonème permet de passer à un niveau de généralité supérieur en étant la somme des traits pertinents de phonèmes différents entre plusieurs langues.

Une hyperforme peut donc se réaliser dans des mots très proches les uns des autres, tels que FR « lit » et EN *lit* (« allumé »), ou dans des mots qui diffèrent par un phonème au moins si ces phonèmes sont les réalisations d'un même hyperphonème, par exemple les mots FR « tour » et EN *sour* (« acide »).

L'hypersémème est, par certains aspects, à l'archisémème ce que l'hyperphonème est à l'archiphonème. B. Pottier définit ainsi l'archisémème :

Lorsqu'on compare les sèmes d'un ensemble de formes, on s'aperçoit qu'il y a des éléments communs (ce sont ces sèmes communs qui ont permis de constituer cet ensemble). Faisant l'intersection de ces ensembles, on obtient l'archisémème. (Pottier 1964 : 122)

De même, un hypersémème est la somme des éléments de signification communs à plusieurs mots, mais sans que ces mots entretiennent obligatoirement une relation sémantique aussi forte que dans le cas de l'archisémème. Alors que les traits articulatoires sont communs aux phonèmes de langues différentes, les sèmes sont des unités théoriques descriptives pour une langue donnée. Un hypersémème n'est donc pas la somme des sèmes communs à plusieurs mots de langues différentes mais la somme des sèmes communs à

plusieurs mots d'une seule langue, somme qui permet de passer à un niveau de généralité sémantique supérieur⁴¹.

Sur le modèle précédent, nous posons que :

Hyperforme + hypersémème = hyperlexème

Signalons que, puisque les réalisations d'un hyperlexème n'ont pas à être strictement équivalentes du point de vue sémantique, elles n'appartiennent pas obligatoirement à la même catégorie grammaticale (ex. : FR « ponctuel » et DE *Punkt*). L'ensemble des hyperlexèmes pour un nombre donné de langues constitue l'hyperlexique de ces langues.

Enfin, l'hyperlangue est la notion qui permet d'expliquer que la lecture en langue étrangère soit possible par des apprenants en intercompréhension qui n'ont jamais appris la langue considérée. Elle correspond en effet à l'ensemble des éléments communs à plusieurs langues, que ces éléments soient directement identifiables ou qu'ils demandent une certaine souplesse de lecture pour être identifiés.

L'hyperlangue s'intégrant dans l'hypersystème que nous avons développé est évidemment à différencier nettement de la notion d'hyperlangue telle qu'elle a été développée par S. Auroux :

La langue empirique n'a pas d'existence autonome [...], elle existe dans des manifestations sonores ou scripturaires sans pouvoir s'y réduire, puisque la caractéristique d'un énoncé linguistique est justement de ne pas être simplement une vibration à l'air ambiant. Mais n'existent, dans certaines portions de l'espace-temps, que des sujets, dotés de certaines capacités linguistiques ou encore de « grammaires » (pas nécessairement identiques), entourés d'un monde et d'artefacts techniques, parmi lesquels figurent (parfois) des grammaires et des dictionnaires. Autrement dit l'espace-temps, par rapport à l'intercommunication humaine, n'est pas vide, il dispose d'une certaine structure que lui confèrent les objets et les sujets qui l'occupent. Appelons hyperlangue cet espace-temps ainsi

⁴¹ Nous renvoyons à la section 2.3.2. pour un développement sur la proximité sémantique indirecte et ainsi sur ce qu'implique la sélection de sèmes évoquée ici.

structuré. [...] En tout état de cause elle est cette réalité ultime qui englobe et situe toute réalisation linguistique et limite concrètement toute innovation. (Auroux 1998 : 115)

L'hyperlangue telle que nous la concevons est une entité purement théorique, appartenant à un système descriptif, et qui représente la somme de tous les éléments qui peuvent être mis en relation entre des langues données.

Les différentes notions présentées ci-dessus peuvent être résumées dans le schéma explicatif qui suit⁴².

L'intersystème, dans le cadre duquel s'inscrit la transparence directe :

Interlangue	Interlexique	Interlexème	Interforme (se réalise dans des mots de formes très proches) + Intersémème (se réalise dans des mots de significations très proches)
		Interlexème	
		Interlexème	
		Interlexème	
	...		

⁴² Nous n'y intégrons que les notions que nous avons traitées, c'est-à-dire celles ayant rapport au lexique, mais nous signalons par les points de suspension que d'autres aspects de l'interlangue et de l'hyperlangue pourraient être intégrés dans les tableaux.

est subsumé par l'hypersystème, dans le cadre duquel s'inscrit de surcroît la transparence indirecte :

Hyperlangue	Hyperlexique	Hyperlexème	Hyperforme (se réalise dans des mots de formes très proches ou différant par la réalisation d'un hyperphonème) + Hypersémème (se réalise dans des mots de significations très proches ou partageant un certain nombre de sèmes)
		Hyperlexème	
		Hyperlexème	
		Hyperlexème	
	...		

Les systèmes théoriques présentés ci-dessus ont pour but de mettre en relation les différents points de comparaison entre langues, et de rendre compte des moyens d'accéder à la compréhension de mots qui appartiennent à des langues inconnues. Nous allons, dans la suite du travail, développer les différents aspects de la transparence, en nous tenant aux objectifs définis dans la section 2.1.1. Nous serons ainsi en possession d'outils permettant de décider si des mots sont transparents ou non, et nous appliquerons ces outils, dans la troisième partie de cette étude, à la détermination de l'interlexique puis de l'hyperlexique des langues germaniques qui nous intéressent et du français.

2.1.3. Du statut inégal de la forme et de la signification

Nous proposerons dans les deux chapitres suivants des critères et des méthodes qui permettent de décider si un mot d'un lexique étranger doit être considéré par nous comme suffisamment proche d'un mot français pour être dit « transparent ». Un mot étranger sera considéré comme transparent s'il est proche d'un mot français *et* du point de vue de la forme *et* du point de vue de la signification. Nous parlerons de « proximité formelle » lorsque le mot étudié aura une forme similaire à celle d'un mot français, et de « proximité sémantique » lorsqu'il aura une signification similaire. Nous distinguerons également la

proximité directe de la proximité indirecte, selon le degré de similitude entre les mots étudiés. Chacune de ces notions sera précisée dans un chapitre qui lui a été dévolu.

Ce plan pourrait suggérer – et nous avons effectivement cherché à construire un système symétrique – que les deux approches, formelle et sémantique, sont parallèles. Néanmoins, même si nous traiterons, d’abord dans la transparence directe puis dans la transparence indirecte, de la proximité formelle puis de la proximité sémantique, ces deux types de proximité ne sont pas de nature comparable, et ne peuvent être tous deux examinés séparément. En effet, si nous pouvons observer des formes pour elles-mêmes, sans prise en compte d’une quelconque signification, nous ne pouvons pas observer des significations sans prise en compte de la forme.

Contrairement à une forme que nous pouvons observer objectivement et à laquelle nous pouvons faire subir diverses manipulations sans prendre en compte la face significative (rien ne nous empêche de jouer sur la forme, qui se trouve être celle d’un verbe néerlandais, *ritselen*, et de construire les formes *litsalen*, *rodsenen*, etc.), une signification ne peut être appréhendée seule. Nous ne pouvons poser l’existence de pensées, analysables en et pour elles-mêmes, auxquelles nous donnerions dans un second temps une face signifiante pour les fixer en langue. Saussure, dans ses textes occultés pendant une longue période et publiés en 2002 dans les *Écrits de linguistique générale*, avait déjà senti la différence de nature entre la forme et la signification (il utilise une terminologie qui n’est pas familière au lecteur du *Cours*, et nous donnons entre crochets des équivalents pour rendre le passage plus clair) :

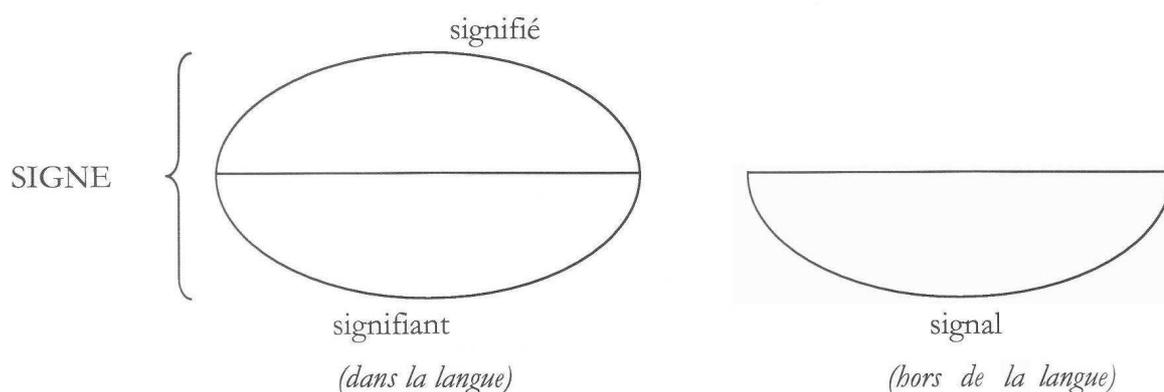
Item. Ce qu’on appelle la signification est ce que nous appelons le parasôme et, à la différence du sôme [≈ forme], ne peut jamais être dégagée de manière à devenir pour elle-même un objet de recherche ou d’observation. Entendons-nous bien : elle peut devenir dans une certaine mesure un tel objet de recherche et d’observation à la condition qu’on en revienne sans cesse au sème [≈ signe], aux différents sèmes qui unissent ce parasôme à quelque chose de matériel, c’est-à-dire au sôme, mais ceci ne constitue rien de semblable à l’étude des sômes, que nous avons reconnue indépendante. (Saussure 2002 : 115)

Les significations doivent être obligatoirement considérées en ce qu’elles sont la face significative d’un signe. De plus, dès que nous parlons de signification, la forme à

laquelle elle est attachée accède au statut de *signifiant*, alors lui-même indissociablement attaché à un *signifié*.

Si nous insistons sur ce point, c'est parce que, pour les lecteurs débutants, un mot étranger ne sera jamais proche d'un mot français uniquement du point de vue de la signification. Ce n'est que parce qu'une forme aura été identifiée que les capacités sémiologiques du lecteur pourront se mettre en branle. Comme le rappelle P. Bogaards, « [c]e à quoi les apprenants sont donc confrontés au premier abord, ce ne sont pas des sens, mais des formes, des formes qui ne sont, dans un premier temps, que des suites de lettres ou de sons » (Bogaards 1994 : 166).

Cette primauté de la forme sur la signification est tout aussi valable en lecture en langue maternelle qu'en lecture en langue étrangère. J.-E. Tyvaert revient, dans son article « La résolution du 'problème de Saussure' et l'élaboration des connaissances par la pratique des langues », sur la pensée de Saussure vulgarisée par Ch. Bally et A. Sechehaye dans le *Cours de linguistique générale* et sur sa pensée « insatisfaite » telle qu'elle est visible dans les *Écrits de linguistique générale*. Après avoir rendu compte des différentes hésitations de Saussure quant au statut des instances constituant le signe linguistique, J.-E. Tyvaert propose, pour rendre compte de ce qu'il appelle « l'asymétrie du signe linguistique », le schéma suivant, qui expose l'antériorité du signal sonore par rapport à l'identification mentale du signe dans son ensemble :



(Tyvaert 2009 : 256)

C'est par l'intermédiaire d'un signal physique (oral ou écrit) qu'un interlocuteur ou un lecteur accède à un signe, et ce signal est analysable pour lui-même, indépendamment de toute considération sémantique.

Ainsi, même si nous parlons successivement de la proximité formelle et de la proximité sémantique, cela ne doit pas faire croire qu'il s'agisse de deux étapes indépendantes l'une de l'autre. Il faut garder à l'esprit que lorsque nous considérons une signification, il s'agit d'une signification identifiée comme étant la signification d'un mot, donc inévitablement liée à un signifiant. Par conséquent, les méthodes de l'ajustement formel et de l'ajustement sémantique, que nous développerons dans le chapitre dévolu à la transparence indirecte, ne peuvent pas non plus être considérées comme deux méthodes que l'on pratiquerait de façon identique mais sur des objets différents. L'ajustement formel se pratique sur une forme (et non sur un signifiant), libre de toute relation à un signifié, tandis que l'ajustement sémantique se pratique sur un signifié, c'est-à-dire sur le contenu sémantique d'un *signe* qui a déjà été identifié formellement.

2.2. La transparence directe

Nous allons, dans ce chapitre, définir la transparence directe, c'est-à-dire préciser quels sont les critères qui doivent être respectés pour que deux mots puissent être dits identiques ou quasi-identiques. Nous nous concentrerons dans un premier temps sur la proximité formelle et dans un second sur la proximité sémantique.

2.2.1. La proximité formelle directe

Nous commencerons par déterminer quels sont les critères qui définissent la proximité formelle directe, puis consacrerons un développement à chacun des critères établis.

2.2.1.1. Détermination des critères définitionnels

Comme nous l'avons vu, la similarité formelle entre un mot d'une langue étrangère et un mot de la langue maternelle de l'apprenant a été maintes fois observée et citée comme aide à la reconnaissance du mot, mais cette ressemblance n'a jamais été clairement définie. Nous avons donc décidé d'établir quels sont les critères objectifs qui permettent de décider si un mot d'une langue étrangère doit être considéré comme ayant une forme suffisamment proche de celle d'un mot en langue maternelle pour qu'elle permette à un lecteur de penser à ce mot en langue maternelle. Nous appelons la situation dans laquelle des mots se ressemblent formellement les uns les autres la « proximité formelle directe », où « directe » signifie que nous décrivons la proximité la plus immédiate, celle à laquelle il est fait référence habituellement, mais sans qu'elle ait jamais été caractérisée précisément. Pour la définir, nous avons analysé un dictionnaire de faux amis. En effet, pour avoir une liste de mots réputés proches formellement, un dictionnaire de ce type nous a paru être l'outil idéal, les faux amis étant définis comme des « unités lexicales qui ont, en deux ou plusieurs langues, la même forme signifiante ou une forme quasi-identique, si bien que l'on pourrait en conclure, à tort, qu'ils ont le même signifié » (Goffin 2000 : 73). Nous

trouvons dans les introductions aux dictionnaires de faux amis anglais-français d'A. Petton et de J. Van Roey *et al.* une précision supplémentaire quant à l'identité ou la quasi-identité des deux formes : il s'agit d'un phénomène d'homographie. Voici leur définition respective :

Les « faux-amis » [...] sont donc des homographes (ou quasi-homographes) de deux langues différentes. Cette similitude formelle ne s'accompagne, malheureusement, que rarement d'une similitude sémantique. (Petton 1995 : 9)

[...] des mots qui, dans leur forme écrite tout au moins, frappent par leur ressemblance ou sont même identiques. (Van Roey *et al.* 1998 : XVI)

Le fait que les faux amis soient, au point de vue définitoire, considérés comme des exemples types de mots proches formellement mais non sémantiquement nous a incitée à tirer de leur observation des règles permettant de définir la proximité formelle. Il ne s'agit pas, comme dans l'étude de R. LeBlanc et H. Séguin que nous avons exposée p. 26, de justifier après-coup notre intuition. Nous ne tirons pas de règles de l'observation de mots que nous aurons jugés transparents, ou congénères, mais de l'observation de mots qui, précisément, ont pour définition d'être proches formellement. Le sémantisme est relégué au second plan, et ce sont ainsi des formes qui ont pour seule caractéristique d'être similaires que nous allons analyser (le fait que les mots aient parfois un lien de signification étant dû au hasard).

Exposons maintenant comment nous avons choisi le dictionnaire à partir duquel pratiquer l'analyse. Dans le but de rendre compte de la capacité d'identification de formes par un apprenant francophone, nous avons lancé nos recherches parmi les dictionnaires de faux amis de langues étrangères par rapport au français. Il n'existe pas, à notre connaissance, de dictionnaire de faux amis néerlandais-français. Nous en avons trouvé un entre l'allemand et le français et plusieurs étaient à notre disposition pour l'anglais et le français. Il fallait que le dictionnaire mette en regard une forme d'une langue étrangère et une forme en français pour avoir une base objective de comparaison. Il fallait en outre que le choix des mots apparaissant dans le dictionnaire n'ait rien à devoir à une étymologie commune ou à une quelconque similitude de signification. En effet, nous tenions à travailler uniquement sur des ressemblances formelles, et le fait qu'elles soient fortuites ne constitue pas, à notre avis, une raison pour ne pas faire apparaître certains mots dans un

dictionnaire de faux amis, les apprenants ne connaissant pas obligatoirement l'étymologie des mots qu'ils essayent de comprendre ou de traduire. Nous avons donc rejeté le *Dictionnaire des faux amis Deutsch-français / Wörterbuch der Faux amis français-Deutsch* de F. Vanderperren puisque les entrées y « sont constituées de l'emprunt tel que les Allemands l'utilisent, suivi du mot français qui en est à l'origine » (Vanderperren 2001 : 7), celui de J. Van Roey *et al.*, puisqu'ils ont exclu de leur ouvrage « les cas de pure coïncidence formelle, tels que *pain/pain* ou *chat/chat*, sauf si, comme pour *mâcher/mash*, la proximité sémantique justifie de les inclure. » (Van Roey *et al.* 1998 : XVI), ainsi que *Les faux amis de l'anglais* de F. Allinne puisque dans ce dictionnaire « [n]e sont pas recensés les mots dont le sens est supposé parfaitement connu (comme l'anglais *chair* qui ne signifie pas *chair* mais *chaise*), non plus que les simples similitudes orthographiques (la note française *do* et le verbe anglais *do*) ». Qui plus est, apparaissent dans ce dictionnaire des expressions qui ne respectent pas la définition d'un faux ami, telle que *I'm afraid* comme faux ami de « j'ai peur » (car elle peut aussi signifier « j'ai bien peur que », « je crois bien que » etc.), qui n'est pas proche formellement mais qui a d'autres traductions possibles que celle habituellement sue. Or, nous le répétons, ne voulant travailler justement que sur des cas de ressemblance formelle, et que celle-ci soit une coïncidence ou pas n'important pas, nous n'avons pas retenu ces trois dictionnaires. Nous n'avons également pas sélectionné *Les faux-amis anglais en contexte* d'A. Petton, ce dictionnaire ne donnant pas le mot français évoqué par la forme du faux ami anglais ni *Un air de famille, les faux amis français-anglais* de C. Rivière, ce dernier ne donnant pas systématiquement le mot français. Notre choix s'est ainsi porté sur l'ouvrage *Les faux amis* de M. Ballard, publié en 1999, dont la liste de faux amis répond à tous nos critères de sélection. L'auteur précise en effet :

[...] nous préférons [...] prendre des critères plus larges que ceux de la communauté étymologique pour définir les faux amis. On ne saurait bien entendu dresser une liste exhaustive de ces ressemblances mais nous avons introduit dans notre glossaire quelques « faux amis fortuits » classiques, ou sur lesquels nous avons pu constater des erreurs. (Ballard 1999 : 12)

Ne sont pas uniquement recensés les faux amis qui auraient un lien, même ténu, du point de vue sémantique, mais aussi ceux qui n'en ont aucun. La ressemblance formelle avec un mot français y est donc considérée comme seul critère de sélection.

Nous allons maintenant présenter la méthodologie que nous avons suivie pour analyser cette liste. Rappelons que les faux amis sont des faux amis pour les élèves en cours d'apprentissage d'une langue ainsi que, dans le cas de l'intercompréhension, pour les personnes qui cherchent à lire un texte en langue étrangère, sans forcément être expertes dans cette langue. C'est la forme écrite du mot étranger (voir les définitions ci-dessus sur le caractère homographique des deux signes) qui peut les mettre sur une mauvaise piste lors d'une tentative de compréhension. Nous avons donc relevé ce qu'il y a, dans la forme écrite du faux ami, de commun et de différent par rapport à la forme du mot évoqué en français, sans invoquer aucune connaissance d'ordres étymologique, morphologique ou catégoriel que nous ne pouvons attribuer par défaut à un lecteur éventuellement débutant.

Nos observations nous font aboutir à deux conclusions de nature différente sur la façon dont on peut définir la proximité formelle d'après l'analyse d'un dictionnaire de faux amis. Nous allons présenter ces deux voies successivement.

1) Dans le but de définir la proximité formelle directe, et ainsi savoir quand on peut déclarer deux mots comme étant proches formellement ou non, nous avons relevé, grâce à une analyse méthodique, toutes les différences entre les paires de mots (les faux amis et les mots proches formellement en français) pour savoir jusqu'où précisément peut aller la quasi-homographie. Il s'agit de savoir ce qui peut être différent entre deux formes, pour ensuite dégager les critères essentiels à une proximité formelle. Nous avons donc procédé à une analyse intentionnellement naïve, avec des remarques telles que « remplacement d'une voyelle par une consonne », même si nous savions que ce changement s'expliquait autrement par la prise en compte de connaissances en phonétique diachronique. Ainsi, c'est à dessein que nous n'avons pas utilisé de terminologie technique (tel que « commutation » ou « groupe consonantique »). La seule connaissance que nous pouvons attribuer à un lecteur francophone débutant en langue étrangère en matière de reconnaissance graphique (rappelons qu'il s'agit d'homographie et de quasi-homographie), c'est l'identification des lettres comme étant des consonnes et des voyelles d'après la connaissance de son alphabet. Nous allons présenter les différences graphiques observées sur les formes anglaises par rapport aux formes françaises en deux temps : d'abord les différences initiales et médianes, puis les différences finales, parce que nous avons relevé les mêmes types de différences dans les débuts et milieux de mots. Nous exposons les différences en termes d'ajout, de suppression ou de remplacement de consonnes et de voyelles, sans prendre en compte le phénomène, qui manifestement ne pose pas de problèmes, de la suppression de signes

diacritiques et celui, très marginal, de l'ajout d'espaces (seul exemple : *quid pro quo*/« quiproquo »). Nous faisons le choix d'appeler « consonne initiale » la première consonne du mot, y compris si elle est précédée d'une ou plusieurs voyelles, et non la première consonne uniquement si elle est la première lettre du mot. Il en va de même pour la « voyelle initiale », qui est la première voyelle du mot (même si elle est précédée d'une ou plusieurs consonnes), de la « consonne finale », qui est la dernière consonne du mot (même si elle est suivie d'une ou plusieurs voyelles) et de la « voyelle finale », qui est la dernière voyelle du mot (même si elle est suivie d'une ou plusieurs consonnes). Nous avons fait ce choix car il importe de pouvoir parler de la dernière consonne en tant que dernière des consonnes.

Nous avons donné pour chaque cas un exemple où, dans la mesure du possible, seule la différence concernée était visible (pour chaque exemple, la forme anglaise est donnée la première, car c'est celle que le lecteur a sous les yeux, et vient ensuite la forme française, qui est la forme de référence pour ce lecteur, et donc celle à partir de laquelle il constate une différence ; « ajout d'une voyelle » signifie donc qu'il y a une voyelle de plus en anglais qu'en français).

Différences initiales et médianes observées :

- Ajout d'une consonne au contact d'une autre consonne (*advance*/« avance »)
- Suppression d'une consonne au contact d'une autre consonne (*deliver*/« délivrer »)
- Remplacement d'une consonne par une autre consonne au contact d'une autre consonne (*accomplishment*/« accomplissement »)
- Ajout d'une voyelle (*filibuster*/« flibustier »)
- Ajout d'une voyelle au contact d'une autre voyelle (*achievement*/« achèvement »)
- Suppression d'une voyelle (*department*/« département »)
- Suppression d'une voyelle au contact d'une autre voyelle (*furniture*/« fourniture »)
- Remplacement d'une voyelle par une autre voyelle (*marmalade*/« marmelade »)
- Remplacement d'une voyelle par une autre voyelle au contact d'une autre voyelle (*pain*/« peine »)
- Remplacement d'une voyelle par une consonne au contact d'une autre consonne (*stable*/« étable »)

Différences finales observées :

- Ajout d'une consonne (*delicatessen*/« délicatesse »)
- Ajout d'une consonne au contact d'une autre consonne (*ancient*/« ancien »)
- Ajout d'un groupe de consonnes (*benevolent*/« bénévole »)
- Suppression d'une consonne (*bribe*/« bribes »⁴³)
- Remplacement d'une consonne par une autre consonne (*benefit*/« bénéfice »)
- Ajout d'une voyelle (*concourse*/« concours »)
- Ajout d'une voyelle au contact d'une autre voyelle (*appeal*/« appel »)
- Suppression d'une voyelle (*branch*/« branche »)
- Suppression d'une voyelle au contact d'une autre voyelle (*diner*/« dîneur »)
- Remplacement d'une voyelle par une autre voyelle (*drama*/« drame »)
- Remplacement d'une voyelle par une autre voyelle au contact d'une autre voyelle (*actual*/« actuel »)
- Remplacement d'une voyelle par une consonne au contact d'une autre consonne (*abstract*/« abstrait »)

Nous venons de donner la liste des différences graphiques ne remettant pas en cause la proximité formelle. S'il n'y a aucune de ces différences entre deux mots (et c'est d'ailleurs souvent le cas, par exemple entre *report* et « report »), les faux amis sont homographes, et s'il y a au moins une de ces différences, ils sont quasi-homographes.

En comparant ces listes, on observe les mêmes différences dans les différences initiales et médianes d'une part et les différences finales de l'autre, exception faite de certaines différences entre consonnes : dans les différences initiales et médianes, il y a ajout, suppression ou remplacement de consonnes seulement si ces consonnes sont au contact d'autres consonnes, alors que dans les différences finales, il y a ajout, suppression ou remplacement de consonnes même si elles ne sont pas au contact d'autres consonnes. On constate donc que la seule manipulation interdite, c'est-à-dire qui rend les deux mots trop dissemblables et ainsi qui remet en cause la proximité formelle, est celle d'ajouter, de supprimer ou de changer une consonne initiale (à savoir, rappelons-le, la première consonne d'un mot, y compris si elle est précédée d'une ou plusieurs voyelles) ou une consonne médiane sans qu'elle soit au contact d'une autre consonne. De plus, nous

⁴³ Bien que le mot « bribes » existe au singulier en français, nous avons gardé la forme plurielle puisque nous travaillons sur les formes que M. Ballard a choisi de mettre en regard.

n'avons relevé de différences entre les consonnes finales (à savoir les dernières consonnes des mots, y compris si elles sont suivies d'une ou plusieurs voyelles), c'est-à-dire un ajout, une suppression ou un remplacement, que lorsqu'il y avait déjà au moins deux consonnes identiques entre les deux mots.

Nous formulons ainsi la règle suivante : *pour que deux mots soient proches formellement directement, ils doivent avoir la même « trame consonantique », c'est-à-dire les mêmes consonnes apparaissant dans le même ordre, avec éventuellement dans un des mots une consonne de plus ou de moins au contact de ses consonnes initiales ou médianes, et la dernière consonne, si elle est au moins la troisième, éventuellement différente. La trame vocalique constitue bien entendu une aide supplémentaire si elle est identique entre les deux mots, mais il n'est pas nécessaire qu'elle le soit (refresh/« rafraîchir », stuff/« étoffe »).*

Les différences citées sont ainsi cumulables, tant que les deux mots gardent la même trame consonantique (comme dans *gum*/« gomme » : de gauche à droite, et en observation de l'anglais par rapport au français, remplacement d'une voyelle initiale, un <o>, par une autre voyelle, un <u>, suppression d'une consonne finale, un <m>, au contact d'une autre consonne, un <m>, et suppression d'une voyelle finale, un <e> ; ou bien dans *benefit*/« bénéfice » qui constituait un des exemples des différences finales observées : remplacement d'une consonne finale, un <c>, par une autre consonne, un <t>, et suppression d'une voyelle finale, un <e>).

La proximité formelle a été définie ci-dessus du point de vue graphique. Il s'avère en réalité que des critères d'ordre phonologique doivent entrer en ligne de compte. En effet, nous avons observé que certaines paires de mots différaient parfois par une consonne initiale ou médiane. Ces différences ne remettent cependant pas en cause la trame consonantique dans la mesure où ces consonnes graphiques sont unies par un lien phonologique, qui peut être de plusieurs ordres. Les consonnes en question peuvent noter le même phonème (*hazard*/« hasard »), ou bien être en relation parce qu'elles peuvent noter, en français, le même phonème, même si ce n'est pas le cas dans le couple de mots considérés (*gest*/« geste » ; *complacency*/« complaisance »). Nous expliquons de la même façon le fait que *ability* ait été considéré comme faux ami de « habileté » : le <h> au début du mot français n'est pas considéré comme un ajout initial de consonne, donc comme une remise en cause de la trame consonantique, puisqu'il ne se prononce pas (notons en revanche qu'aucun mot anglais commençant par un <h> n'a été considéré comme faux ami

d'un mot français qui n'aurait pas de <h> à l'initiale, puisque le <h> anglais transcrit, lui, un phonème). Une consonne écrite qui ne se prononce pas entre donc dans la trame consonantique graphique, mais n'est pas considérée comme une entrave à la trame consonantique phonologique.

Que des consonnes écrites soient considérées comme équivalentes grâce au lien phonologique qui les unit peut s'expliquer. Nous formons pour cela un système inspiré du signe linguistique conçu comme l'union d'un signifiant et d'un signifié. Ce système unit, par comparaison, une face signifiante, un graphème (un phonogramme plus précisément), et une face signifiée, un phonème. Le graphème est défini par N. Catach comme la « plus petite unité distinctive et/ou significative de la chaîne écrite, composée d'une lettre, d'un groupe de lettres (digramme, trigramme), d'une lettre accentuée ou pourvue d'un signe auxiliaire, ayant une référence phonique et/ou sémique dans la chaîne parlée » (Catach 1986 : 16). Nous considérons ici un type particulier de graphèmes, les phonogrammes, c'est-à-dire les « graphèmes chargés de transcrire les phonèmes » (*ibid.* : 16), qui représentent en français 80 à 85 % des unités graphiques (*ibid.* : 23). L'idée d'associer entre eux le graphème et le phonème, le premier étant comparé au signifiant et l'autre au signifié, est présentée dans l'ouvrage de V. G. Gak et I. Vildé-Lot, *L'orthographe du français : essai de description théorique et pratique*, publié en 1976 : « L'union de la graphie (signifiant) et du phonème qu'elle transcrit (signifié) forme le signe graphique, le graphème, unité minimale constitutive du système graphique » (Gak & Vildé-Lot 1976 : 24). Nous nous éloignons de cette définition en ce que nous réservons le terme « graphème » à la face signifiante, et non à l'union du signifiant et du signifié, que nous appelons « signe graphique ». Si nous appliquons à ce signe graphique ce que Saussure dit du signe linguistique, alors « les termes impliqués [...] sont tous deux psychiques. » Il ajoute : « Le signe linguistique unit non une chose et un nom, mais un concept et une image acoustique. Cette dernière n'est pas le son matériel, chose purement physique, mais l'empreinte psychique de ce son » (Saussure 1995 : 98). De même, le signe graphique unit non une lettre (ou un groupe de lettres) et un son, mais un graphème et un phonème. « p, P, p, ꝑ, etc. » sont des réalisations du graphème <p> et notent toutes de la même façon le phonème /p/, et non tel son physique particulier. N. S. Troubetzkoy déclare, dans l'ouvrage *Principes de phonologie*, que « le phonème coïncide, non pas avec une image phonique concrète, mais seulement avec les particularités phonologiquement pertinentes de cette image. On peut dire que le phonème est la somme des particularités phonologiquement

pertinentes que comporte une image phonique » (Troubetzkoy 1970 : 40), et nous disons après lui que le graphème est la somme des particularités pertinentes que comporte une lettre (dans notre exemple, le dessin d'un jambage descendant à l'immédiate gauche d'un cercle suffit à représenter le <p>, et les autres caractéristiques importent peu pour sa reconnaissance).

Saussure déclare également que « le lien unissant le signifiant au signifié est arbitraire » (Saussure 1995 : 100). En effet, rien ne prédispose le graphème <p> à transcrire la consonne du français occlusive bilabiale sourde, et un même phonème (/ʃ/ par exemple) peut avoir pour signifiant des graphèmes différents dans des systèmes différents (<sch> en allemand et <š> en tchèque). De plus, nous savons depuis l'article « Nature du signe linguistique » d'É. Benveniste que le lien qui unit le signifiant au signifié est non pas arbitraire mais nécessaire : le signifié « est forcément identique dans ma conscience [au] signifiant [...]. Comment en serait-il autrement ? Ensemble les deux ont été imprimés dans mon esprit ; ensemble ils s'évoquent en toute circonstance » (Benveniste 2002 vol.1 : 51). Ainsi du signe graphique : vu la place qu'occupe le système graphique dans notre société, un lecteur a *forcément* le phonème correspondant à tel graphème à l'esprit pendant sa lecture. C'est pourquoi la prise en compte de critères d'ordre phonologique s'avère nécessaire pour définir la proximité formelle.

Ainsi, sont considérés comme équivalents (qu'ils notent ou pas, dans la paire de mot effectivement considérée, le même phonème) :

- <p> et <pp>
- et <bb>
- <t>, <tt> et <th>
- <d> et <dd>
- <c>, <qu>, <cc>, <cqu>, <ch>, <k>, et <q>
- <g>, <gu> et <gg>
- <f>, <ff> et <ph>
- <v> et <w>
- <ou(+V)> et <w>
- <s>, <ss>, <c>, <ç>, <t(+i)> et <sc>
- <s> et <z>
- <ch> et <sch>
- <j>, <g> et <ge>

- <l> et <ll>
- <r>, <rr> et <rh>
- <m> et <mm>
- <n> et <nn>⁴⁴

Regardons de plus près deux exemples (*parakeet*/« perroquet » et *seizure*/« saisie ») tirés du dictionnaire de faux amis de M. Ballard pour clarifier les différentes règles que nous avons énoncées.

parakeet/« perroquet » : remplacement d'une voyelle par une autre voyelle

parakeet/« perroquet » : suppression d'une consonne au contact d'une autre consonne

parakeet/« perroquet » : remplacement d'une voyelle par une autre voyelle

parakeet/« perroquet » : remplacement d'un graphème consonantique par un autre graphème consonantique, neutralisé par le fait que les deux graphèmes transcrivent le même phonème

parakeet/« perroquet » : ajout d'une voyelle au contact d'une autre voyelle

seizure/« saisie » : remplacement d'une voyelle par une autre voyelle au contact d'une autre voyelle

seizure/« saisie » : remplacement d'une consonne par une autre consonne, neutralisé par le fait que ces deux consonnes peuvent, en français, transcrire le même phonème

seizure/« saisie » : remplacement d'une voyelle par une autre voyelle

seizure/« saisie » : ajout d'une consonne finale, précédée de deux consonnes qui sont communes aux deux mots

Malgré les cinq différences relevées entre *parakeet* et « perroquet » et les quatre différences relevées entre *seizure* et « saisie », ces mots ont été considérés comme proches formellement par M. Ballard. Grâce à notre analyse, nous pouvons attribuer cela au fait qu'ils ont la même trame consonantique.

⁴⁴ Cette liste a été constituée grâce à l'ouvrage *L'orthographe française, Traité théorique et pratique avec des travaux d'application et leurs corrigés* et au *Dictionnaire historique de l'orthographe française* de N. Catach.

2) La deuxième conclusion à laquelle nous aboutissons tient au rôle des premières lettres des mots. Nous avons en effet remarqué que des mots n'étant proches que par leurs premières lettres, et qui n'avaient donc pas la même trame consonantique, ont tout de même été considérés comme proches formellement, puisque M. Ballard les cite dans son dictionnaire de faux amis.

Après observation, nous formulons la règle suivante : *si les deux mots comparés sont composés de quatre lettres ou plus, il suffit qu'ils aient au moins les quatre premières en commun pour être considérés comme proches formellement (apply/« appliquer »). Si l'un des deux mots a moins de quatre lettres, il suffit que le mot le plus long commence par toutes les lettres du mot le plus court pour qu'ils soient considérés comme proches formellement (fatuus/« fat »).*

Nous venons de déterminer deux règles qui permettent de définir objectivement la proximité formelle directe entre deux mots : il faut qu'ils aient la même trame consonantique ou bien qu'ils aient leurs quatre premières lettres en commun (auquel cas la suite des mots n'a plus à être identique). Néanmoins, ces règles ont été établies d'après l'analyse d'un dictionnaire de faux amis anglais-français. La suite de nos analyses part donc de l'hypothèse que de telles règles sont également valables entre l'allemand et le français et le néerlandais et le français. Nous ne pensons toutefois pas que cette hypothèse de travail soit déraisonnable, puisque nous avons établi des règles qui permettent de savoir si deux *formes*, deux chaînes de caractères, sont proches. Qu'une forme évoque à un francophone un mot français ne nous semble pas tenir à ce qu'elle appartienne au lexique d'une langue particulière.

Nous allons maintenant développer chacune des deux données qui apparaissent comme les plus importantes pour la reconnaissance d'un mot : les consonnes et le début de ce mot. Nous allons voir ce qu'en disent des auteurs ayant des spécialités et des projets différents, et dont les propos convergent pour confirmer la primauté des consonnes sur les voyelles et du début du mot sur sa fin.

2.2.1.2. Importance des consonnes

Citons pour commencer un extrait de *Alice au pays du langage*. Dans cet ouvrage de vulgarisation scientifique publié en 1981, M. Yaguello présente des théories linguistiques à des lecteurs non initiés, de sorte que nous pouvons y trouver des passages qui ont l'avantage d'illustrer de façon très claire telle ou telle notion. Dans cet extrait, l'auteur fait comprendre à ses lecteurs l'importance des consonnes par le récit d'un jeu :

Savez-vous jouer au zygomar, ce jeu que, dit-on, appréciait particulièrement Saint-Exupéry ? Je choisis un mot, vous en choisissez un autre, du même nombre de lettres ; cinq, par exemple. Je cherche à découvrir le vôtre, vous le mien. Pour cela, chacun propose à tour de rôle un mot de cinq lettres ; l'adversaire précise combien de lettres coïncident, de zéro à cinq. Une fois la première lettre trouvée, il est facile, à condition de procéder par substitutions successives d'une seule lettre à la fois, de trouver les autres. Mais on s'aperçoit assez vite en pratiquant ce jeu (comme, d'ailleurs, les mots croisés) que la valeur d'information des consonnes est supérieure à celle des voyelles ; celles-ci se déduisent facilement du contexte consonantique et sont donc, dans une large mesure, prévisibles, alors que l'inverse n'est pas vrai. La valeur d'information respective des voyelles et des consonnes est proportionnelle à leur nombre dans l'alphabet. Supposons que vous ayez choisi le mot RECIF. Si j'ai trouvé les deux voyelles, il me faudra encore pas mal de questions avant de gagner car le nombre de mots de cinq lettres comportant .E.I. est très grand. Si par contre j'ai trouvé le R et le F, il y a de fortes chances pour que je trouve d'emblée RECIF, en fait une chance sur deux, car RECIF n'est concurrencé que par RETIF. [...]

Ce que le joueur de zygomar découvre de façon empirique, c'est un principe de la théorie de l'information baptisé redondance. Selon ce principe, plus le degré de prédictibilité (en pourcentage) pour que telle ou telle unité se trouve dans tel ou tel contexte est élevé, moins cette unité apporte d'information et donc plus elle est redondante. Ceci explique le procédé de formation des abréviations. En langue écrite, on supprime les voyelles plutôt que les consonnes¹ (sauf les voyelles initiales) : apt, séj, 2 ch, bns, dche, tt cft [...]. En langue orale, il n'est pas possible, évidemment, de supprimer les voyelles, qui sont la base de la prononciation (pas de syllabe sans voyelle). [...] les voyelles sont largement redondantes².

¹ Rappelons que l'hébreu s'écrit sans voyelles.

² On observera de surcroît que les voyelles sont l'élément le plus instable dans l'évolution phonétique des mots.

(Yaguello 1981 : 41-42)

M. Yaguello avance ici trois types d'arguments pour soutenir ce qu'elle avance, tant dans le corps du texte que dans les notes de bas de page :

1) Elle indique que les voyelles « se déduisent facilement du contexte consonantique » et que cela est dû au nombre de voyelles et de consonnes dans l'alphabet. En effet, les voyelles du système écrit étant beaucoup moins nombreuses que les consonnes, elles sont moins discriminantes. Nous retrouvons le même type d'exemple que celui qu'elle donne à propos de « récif » chez P. Léon, qui affirme que les consonnes sont « nécessaires et suffisantes à la compréhension écrite. On retrouve facilement, dans : « Cmm'nt strctrnt ls mts », *Comment se structurent les mots*. Alors que : « o en e uu eo », qui sont les voyelles écrites des mêmes mots, sont impossibles à déchiffrer » (Léon 1996 : 18).

2) Elle rappelle le cas du système graphique de l'hébreu, qui ne note que les consonnes. C'est en effet une particularité des langues sémitiques que d'accorder un rôle précis aux consonnes : la base lexicale est composée d'une suite de consonnes et elle porte le sens plein du mot, tandis que les voyelles, les préfixes et les suffixes apportent la catégorisation grammaticale (Campbell 1998 : 450-451). A. Chabir précise, pour la langue arabe :

La structure du lexique arabe se traduit par son organisation en familles morphologiques ayant pour base la racine dont les lexèmes dérivent. Tout lexème, à l'exception de quelques outils grammaticaux, a pour base lexicale la racine. Celle-ci a une forme purement consonantique constituée par une suite constante et ordonnée de consonnes généralement au nombre de trois, parfois de quatre, très rarement de deux. Mais cette succession de consonnes, représentant le contenu sémantique ou la notion générale commune à l'ensemble des dérivés tirés de cette racine, ne peut constituer une forme linguistique à elle seule. C'est en se combinant

essentiellement aux éléments phoniques a, i, u et à leurs variantes ā, ī, ū qu'elle peut être actualisée. (Chabir 2000 : 117)

P. Léon note la même chose à propos de l'alphabet des Phéniciens : « Les Phéniciens perfectionnèrent les écritures phonographiques en inventant l'écriture alphabétique, dont ils ne notaient que les consonnes. Ils ont senti que les consonnes étaient les éléments porteurs d'intelligibilité pour la transmission du message » (Léon 1996 : 18).

3) Elle signale enfin que les consonnes sont les éléments, phoniques cette fois-ci, les plus stables au cours de l'évolution phonétique des mots. F. Carton, dans la partie de son ouvrage *Introduction à la phonétique du français* consacrée à la phonétique diachronique, déclare de même : « Les consonnes latines et germaniques se sont révélées généralement plus stables que les voyelles » (Carton 1974 : 145).

Ces quelques arguments avancés par différents auteurs, qui, pour l'un retrace l'évolution phonétique des mots d'une langue, pour l'autre veut faire sentir une information à des lecteurs non initiés, etc., montrent le rôle décisif qu'ont les consonnes pour la reconnaissance d'un mot.

Dans le même ordre d'idées, nous nous sommes intéressée aux techniques de sténographie pour savoir si les consonnes étaient des éléments préférentiellement choisis pour noter une prise de parole à un rythme soutenu. C'est en effet le cas, dans beaucoup des systèmes sténographiques, de l'Antiquité à nos jours. Déjà dans les notes tironiennes n'étaient notés que les éléments indispensables pour une relecture ultérieure réussie :

Par ailleurs, plusieurs procédés abrégatifs étaient utilisés : [...] surtout suppression de lettres dont l'absence ne rend pas l'écriture illisible ; en particulier, tout comme dans plusieurs de nos systèmes modernes, les voyelles ne sont généralement pas représentées dans le corps des mots. (Michelot 1959 : 15-16)

Plus tard, la méthode Gabelsberger, la plus utilisée en Allemagne, « se propose de reproduire l'ossature des mots, c'est-à-dire les consonnes » (*ibid.* : 52), et, dans la méthode française Prévost-Delaunay, les signes généraux « représentent une syllabe simple, commençant par une consonne et continuant par un son voyelle quelconque » (*ibid.* : 67).

Par exemple, le signe / peut représenter [da], [di], [dy]... Ces informations confortent l'idée que les consonnes, sinon seules, du moins accompagnées d'un signe indiquant un son vocalique quel qu'il soit, suffisent à faire passer de la signification.

Nous allons maintenant nous attarder sur un article de I. Berent et C.A. Perfetti, publié en 1995 et intitulé « A rose is a REEZ: the two-cycles model of phonology assembly in reading English ». Les auteurs y présentent la thèse selon laquelle la lecture de mots anglais se fait en deux étapes : les lecteurs perçoivent d'abord l'information donnée par les consonnes et ensuite celle donnée par les voyelles. I. Berent et C.A. Perfetti commencent leur article en rappelant que tous les chercheurs s'accordent à dire que le rôle de la phonologie en lecture silencieuse n'est plus à prouver. Il n'y a néanmoins pas consensus sur les modalités d'activation de l'assemblage phonologique : c'est tantôt un processus lent, contrôlé, peu fiable et dont l'effet est limité aux mots de basse fréquence, tantôt un processus rapide, automatique, fiable et dont l'effet est valable pour les mots de toutes fréquences. Les auteurs avancent l'idée que ces différences tiennent aux types d'expériences réalisées pour identifier ce processus, et donc que l'assemblage n'est pas un processus uniforme (et contradictoire au vu de la littérature sur le sujet) mais pluriel, chacune de ses phases étant mise en lumière par tel ou tel type d'expérience. La contradiction se trouve levée si l'assemblage phonologique se fait en deux temps :

*According to the two-cycles model, the derivation of consonants and vowels is accomplished by two distinct mental processes that differ in their speed and automaticity. These processes constitute two cycles of assembly whose output is completed at two consecutive stages. The assembly of consonants is computed by the initial cycle through a relatively automatic process. Vowel information is added to the representation at the second cycle by a slower, controlled process.*⁴⁵ (Berent & Perfetti 1995 : 148)

In summary, the two-cycles model can provide a principled account of the distinct facets of assembly. [...] By assuming that consonants and vowels are computed in

⁴⁵ Nous traduisons : « Selon le modèle à deux cycles, l'activation des consonnes et des voyelles est réalisée lors de deux processus mentaux distincts qui diffèrent par leur rapidité et leur automaticité. Ces processus constituent deux cycles d'assemblage dont la réalisation est achevée lors de deux étapes consécutives. L'assemblage des consonnes est accompli lors du premier cycle par un processus relativement automatique. L'information vocalique est ajoutée à la représentation lors du second cycle par un processus plus lent et contrôlé. »

*different cycles, the model accounts for the contradictory evidence regarding the role of assembled phonology in visual word recognition and provides some methodological insights concerning how experimental methods have determined research conclusions about the nature of the assembly mechanism.*⁴⁶ (*ibid.* : 157)

Nous ne rapporterons ici que la première des sept expériences réalisées par les auteurs pour prouver leurs dires. Ils ont présenté à des sujets 64 mots, qui étaient masqués par quatre types de non-mots : les premiers conservaient les consonnes du mot cible et avaient des voyelles différentes, les deuxièmes conservaient les voyelles et avaient des consonnes différentes, les troisièmes étaient des homophones hétérographes du mot cible et les quatrièmes étaient des masques de contrôle qui ne conservaient ni les consonnes ni les voyelles. Les cibles et les masques étaient présentés selon quatre durées possibles : respectivement 15:30, 30:30, 45:30 et 45:60 ms, et les sujets devaient écrire ce qu'ils avaient perçu. L'hypothèse était la suivante : si l'assemblage pendant le premier cycle ne concerne que les consonnes, alors, avec les plus courtes durées d'exposition, le fait de masquer le mot cible par un mot qui conserve ses consonnes devrait produire un plus grand nombre de réponses exactes que de le masquer par un mot qui conserve ses voyelles. Cet avantage du masque consonantique devrait de plus disparaître avec les durées d'exposition les plus longues, puisque l'information vocalique est alors également activée. Après un exposé chiffré des résultats, les auteurs concluent :

*The results of this experiment demonstrate that the content of the assembled representation varies with processing time, in accord with the predictions of the two-cycles model. At brief durations, evidence for the generation of consonants was obtained in the form of a significant advantage of the consonant-preserving mask compared with the vowel-preserving mask. In contrast, at longer exposure durations, the generation of vowels was reflected in a significant advantage for the homophone.*⁴⁷ (*ibid.* : 163)

⁴⁶ Nous traduisons : « Pour résumer, le modèle à deux cycles rend compte, par ses principes, des différentes facettes de l'assemblage. [...] L'idée selon laquelle les consonnes et les voyelles sont traitées lors de cycles différents permet au modèle d'expliquer les résultats contradictoires quant au rôle de l'assemblage phonologique dans la reconnaissance des mots écrits et de fournir des indications méthodologiques sur le fait que des méthodes expérimentales ont influencé les conclusions sur la nature du mécanisme de l'assemblage. »

⁴⁷ Nous traduisons : « Les résultats de cette expérience montrent que le contenu de la représentation assemblée varie en fonction du temps de traitement, conformément aux prédictions du modèle à deux cycles. Pour de courtes durées, la preuve de l'activation des consonnes a été obtenue sous la forme d'un avantage

Cette idée que les consonnes ont un rôle plus important pour la reconnaissance d'un mot se retrouve également, et c'est là le dernier exemple que nous donnerons pour l'illustrer, en tant que postulat dans le projet SPEEDCOP (Spelling Error Detection/Correction Project), décrit dans l'article « Automatic spelling correction in scientific and scholarly text » de J. J. Pollock et A. Zamora, publié en 1984. Dans le cadre de ce projet a été développé un algorithme destiné à corriger les fautes de saisie dans des textes électroniques. Pour ce faire, chaque mot comportant une faute doit d'abord être mis en relation avec un mot existant (un dictionnaire a été enregistré dans le programme) puis remplacé par le mot sans faute. Nous nous intéressons à la première partie du processus de correction. Les mots erronés ne se trouvant évidemment pas dans le dictionnaire de référence, des « clés de similarité » ont été attribuées à chaque mot du dictionnaire et sont fabriquées pour chaque mot comportant une faute. « A misspelling is then corrected by locating words whose keys collate most closely to the key of the misspelling and selecting the plausible correction(s) from these »⁴⁸ (Pollock & Zamora 1984 : 359). Chaque chaîne (mot erroné ou mot du dictionnaire) se voit attribuer deux clés de similarité. C'est la première qui a retenu notre attention : « the skeleton key ». Y apparaissent les informations suivantes : la première lettre, les consonnes restantes par ordre d'apparition puis les voyelles par ordre d'apparition. Par exemple, la clé de CHEMOGENIC est CHMGNEOI, celle de CHEMOMAGNETIC est CHMGNTEOAI et celle de CHIMICAL est CHMLIA. Les auteurs justifient la forme que prennent les clés par quatre raisons : i) la première lettre tapée est probablement bonne (seulement 7,8 % des premières lettres des mots erronés sont incorrectes, comparé à 11,7 % pour la deuxième lettre et 19,2 % pour la troisième) ; ii) la prééminence des consonnes : « it is generally assumed in the literature that vowels are less important than consonants »⁴⁹ (*ibid.* : 360) ; iii) dans la plupart des mots erronés, l'ordre des lettres n'est que peu altéré ; iv) cette clé est insensible à un nombre significatif de transformations qui créent des fautes, telles que le doublement et le dédoublement de caractère et à la transposition voyelle/consonne. Les auteurs résument cette clé de similarité ainsi : « Since we regard the unique consonants in their original order as the

significatif en faveur du masque qui conservait les consonnes par rapport au masque qui conservait les voyelles. En revanche, pour de plus longues durées d'exposition, l'activation des voyelles a été révélée par un avantage significatif en faveur de l'homophone. »

⁴⁸ Nous traduisons : « Un mot comportant une faute est alors corrigé par le repérage des mots dont les clés correspondent le mieux à la clé du mot erroné puis par la sélection de la ou des corrections plausibles. »

⁴⁹ Nous traduisons : « Il est généralement considéré, dans la littérature, que les voyelles sont moins importantes que les consonnes. »

backbone of a word, we call this key the *skeleton key* »⁵⁰ (*ibid.* : 360). Les consonnes d'un mot sont ici considérées comme sa « colonne vertébrale », et cette expression n'est pas sans rappeler celle d'« ossature » utilisée par M. Michelot lors de la description d'un système sténographique.

Que ce soit un postulat de départ pour une analyse, la conclusion d'une série d'expériences ou un constat au vu de leur rôle respectif pour la communication, la plus grande importance des consonnes vis-à-vis des voyelles semble être établie. Ceci corrobore ce que nous avons observé lors de l'analyse d'un dictionnaire de faux amis ayant pour but la formulation de règles de proximité formelle directe entre deux mots : deux mots sont proches formellement s'ils ont la même trame consonantique.

Nous avons également noté que deux mots étaient considérés comme proches formellement s'ils avaient leurs quatre premières lettres en commun (ou, dans le cas où l'un des deux mots avait moins de quatre lettres, si le mot le plus long commençait par toutes les lettres du mot le plus court). Nous allons procéder de la même manière que ci-dessus, et montrer que l'importance du début du mot est également reconnue et mise à profit.

2.2.1.3. Importance du début du mot

Nous voudrions commencer cette partie par ce avec quoi nous avons commencé la dernière, c'est-à-dire les arguments avancés par M. Yaguello sous forme d'anecdote. Voici à nouveau son texte, dans lequel nous avons réintroduit les éléments qui, maintenant, nous intéressent, et que nous soulignons :

Savez-vous jouer au zygomar, ce jeu que, dit-on, appréciait particulièrement Saint-Exupéry ? Je choisis un mot, vous en choisissez un autre, du même nombre de lettres ; cinq, par exemple. Je cherche à découvrir le vôtre, vous le mien. Pour cela, chacun propose à tour de rôle un mot de cinq lettres ; l'adversaire précise combien

⁵⁰ Nous traduisons : « Puisque nous considérons les seules consonnes dans leur ordre originel comme la colonne vertébrale du mot, nous appelons cette clé la clé *squelette*. »

de lettres coïncident, de zéro à cinq. Une fois la première lettre trouvée, il est facile, à condition de procéder par substitutions successives d'une seule lettre à la fois, de trouver les autres. Mais on s'aperçoit assez vite en pratiquant ce jeu (comme, d'ailleurs, les mots croisés) que la valeur d'information des consonnes est supérieure à celle des voyelles ; celles-ci se déduisent facilement du contexte consonantique et sont donc, dans une large mesure, prévisibles, alors que l'inverse n'est pas vrai. La valeur d'information respective des voyelles et des consonnes est proportionnelle à leur nombre dans l'alphabet. Supposons que vous ayez choisi le mot RECIF. Si j'ai trouvé les deux voyelles, il me faudra encore pas mal de questions avant de gagner car le nombre de mots de cinq lettres comportant .E.I. est très grand. Si par contre j'ai trouvé le R et le F, il y a de fortes chances pour que je trouve d'emblée RECIF, en fait une chance sur deux, car RECIF n'est concurrencé que par RETIF. [...] On constate aussi que le début d'un mot apporte plus d'information que la fin, c'est-à-dire que la fin se déduit plus facilement du début que le début de la fin.

Ce que le joueur de zygomar découvre de façon empirique, c'est un principe de la théorie de l'information baptisé redondance. Selon ce principe, plus le degré de prédictibilité (en pourcentage) pour que telle ou telle unité se trouve dans tel ou tel contexte est élevé, moins cette unité apporte d'information et donc plus elle est redondante. Ceci explique le procédé de formation des abréviations. En langue écrite, on supprime les voyelles plutôt que les consonnes (sauf les voyelles initiales) : apt, séj, 2 ch, bns, dche, tt cft ; la fin des mots plutôt que le début. En langue orale, il n'est pas possible, évidemment, de supprimer les voyelles, qui sont la base de la prononciation (pas de syllabe sans voyelle), mais on supprime volontiers la fin des mots qui se déduit facilement du contexte, d'autant plus que les finales de mots sont souvent des suffixes : compo, interro, bibli, ciné, rétro, Libé, etc. Les finales de mots, comme les voyelles, sont largement redondantes. (Yaguello 1981 : 41-42)

Si, comme le rappelle l'auteur, les procédés abrégatifs en français portent sur la fin des mots, c'est parce que, lorsque nous voulons transmettre un message brièvement, notre choix se porte vers les éléments nécessaires et suffisants au passage de l'information, c'est-à-dire le début des mots (et/ou leurs consonnes comme nous l'avons vu précédemment).

Un des résultats provenant de notre analyse d'un dictionnaire de faux amis, à savoir que si les quatre premières lettres sont communes à deux mots, alors ils sont considérés comme proches formellement, va dans le même sens qu'une information donnée dans l'article de M. Simard, G. F. Foster et P. Isabelle datant de 1992, « Using cognates to align sentences in bilingual corpora ». Ces auteurs, dont nous avons déjà cité les recherches dans la section 1.1.3., travaillent sur des techniques d'alignement, c'est-à-dire sur les moyens de faire correspondre entre eux par ordinateur les items (mots, phrase, paragraphe, etc.) de deux textes différents dont l'un est la traduction de l'autre. Ils avancent, dans l'article cité, l'idée qu'un moyen simple et efficace de s'assurer du bon alignement de deux paragraphes est de s'appuyer sur les cognats, qui sont des éléments fiables. Ils considèrent comme cognats deux candidats dont les quatre premiers caractères sont identiques. Ils ne donnent malheureusement pas de justification pour ce choix des quatre caractères identiques, mais nous trouvons intéressant qu'il rejoigne notre observation.

L. Ferrand consacre une partie de son ouvrage *Psychologie cognitive de la lecture, reconnaissance des mots écrits chez l'adulte*, publié en 2007, à l'exposé d'expériences qui suggèrent que l'effet inhibiteur d'un voisin orthographique (c'est-à-dire d'un mot qui ne diffère d'un autre que par une seule lettre) plus fréquent peut être altéré par la position de la lettre changée pour aboutir à ce voisin. Il synthétise clairement ces expériences :

Grainger et Segui (1990) ont suggéré que le fait qu'un mot comme BILLE est parfois incorrectement reporté comme BIBLE, mais jamais comme VILLE (BIBLE et VILLE sont tous deux des voisins plus fréquents que BILLE) pouvait indiquer que « les voisins différant par les lettres internes sont plus compétitifs que les voisins différant par les lettres initiales. (Ferrand 2007 : 196)

Les facteurs visuels peuvent interagir avec l'effet d'inhibition de la fréquence du voisinage. En effet, Grainger et al. (1992) ont manipulé la visibilité de la lettre qui levait l'ambiguïté d'un mot ayant un voisin orthographique plus fréquent dans la tâche de décision lexicale. Grainger et al. (1992) ont trouvé un effet d'inhibition de la fréquence du voisinage plus faible pour des mots comme CHOPE (ayant un voisin orthographique plus fréquent CHOSE) lorsque les sujets fixaient initialement la 4^e lettre (le P de CHOPE ici, qui lève l'ambiguïté) que lorsqu'ils fixaient la 2^e lettre de mots comme ASTRE (ayant un voisin orthographique plus fréquent AUTRE). Ces résultats suggèrent que les lettres d'un mot ne sont pas toutes

activées également au cours de la lecture. Nous avons vu dans le paragraphe 3.1.5. que la qualité de l'information visuelle échantillonnée par la rétine (l'acuité visuelle) diminue rapidement dès qu'on quitte le point de fixation fovéal (voir O'Regan, 1990). Cela suggère que lorsque nous fixons une suite de lettres, les lettres fixées ou immédiatement autour du point de fixation sont les plus visibles, tandis que les autres lettres sont graduellement moins visibles, sauf les lettres externes qui ne souffrent pas d'inhibition latérale. Par ailleurs, Grainger et al. (1992) ont montré que l'effet d'inhibition de la fréquence du voisinage était plus fort pour les voisins différant par la 4^e lettre (comme CHOPE, voisin de CHOSE) que pour des voisins différant par la 2^e lettre (comme ASTRE, voisin de AUTRE). Cela suggère que les lettres initiales d'un mot entraînent plus d'activation que les lettres finales. De cette manière, l'activation du compétiteur CHOSE va atteindre un seuil d'activation plus élevé durant le traitement du mot CHOPE que le seuil d'activation du compétiteur AUTRE pendant le traitement du mot ASTRE. [...] L'ensemble de ces résultats suggère donc que les effets de compétition entre les représentations lexicales peuvent être modulés par la visibilité des lettres du mot cible. (ibid. : 152)

Nous allons maintenant revenir sur le programme du projet SPEEDCOP dont il a déjà été question précédemment. Une des conclusions de J. J. Pollock et A. Zamora fait écho à ce qu'annonçait M. Yaguello, à savoir que les finales de mots sont redondantes. Ces auteurs indiquent au début de leur article, en s'appuyant sur la seule théorie, que « the most vulnerable aspect of [the skeleton key] is its emphasis on the early consonants. The closer an incorrect consonant is to the start of a word, the greater the collating distance between the keys of the word and misspelling »⁵¹ (Pollock & Zamora 1984 : 360). Cette faiblesse s'est trouvée confirmée, puisque voici ce qu'indiquent les résultats : « 3- and 4-character misspellings are much more likely to be ambiguous than longer ones and therefore to be miscorrected. [...] In fact, although 3-4 character misspellings constitute only 9.24 percent of total misspellings, they generate 42 percent of the miscorrections »⁵² (ibid. : 367). Ceci

⁵¹ Nous traduisons : « L'aspect le plus vulnérable de la clé squelette est le fait qu'elle mette l'accent sur les premières consonnes. Plus une consonne erronée est proche du début du mot, plus la distance entre la clé du mot et celle du mot comportant une faute est grande. »

⁵² Nous traduisons : « Les mots erronés de 3 ou 4 lettres sont beaucoup plus susceptibles d'être ambigus que les mots plus longs et par conséquent d'être mal corrigés. [...] En effet, bien que les mots erronés de 3 ou 4 lettres ne constituent que 9,24 pour cent du nombre total d'erreurs, ils sont la cause de 42 pour cent des mauvaises corrections. »

ne fait que renforcer l'idée que les mots se différencient les uns des autres par leurs premières lettres, qui sont ainsi les plus importantes pour leur identification. C'est en lisant de gauche à droite (en français par exemple) que nous construisons notre représentation, et l'on invoque pour cela la notion de « point d'unicité » :

Pour comprendre cette notion, il faut se placer dans l'hypothèse suivant laquelle le décodage d'un mot se fait linéairement de gauche à droite (du point de vue graphique) ou séquentiellement de phonème en phonème (du point de vue phonique). Le point d'unicité désigne alors le lieu (lettre ou phonème) dans le mot, à partir duquel plus aucun autre mot que celui envisagé n'est possible. Selon qu'on considère la réalité orale ou écrite, on parle de point d'unicité orthographique ou phonique. (Apothéloz 2002 : 121)

Enfin, c'est en se fondant sur les résultats obtenus par J. J. Pollock et A. Zamora que W. E. Winkler a apporté une amélioration à la distance de Jaro, formule qui mesure la similarité entre deux chaînes de caractères. Selon la distance de Jaro-Winkler, que nous avons décrite dans la section 1.1.3., la similarité entre les quatre premiers caractères des deux mots comparés a plus de poids que celle entre les lettres suivantes :

The second enhancement due to Winkler (1990) gives increased value to agreement on the beginning characters of a string. It was based on ideas from a very large empirical study by Pollock and Zamora (1984) for the Chemical Abstracts Service. The study showed that the fewest errors typically occur at the beginning of a string and the error rates by character position increase monotonically as the position moves to the right. The enhancement basically consisted of adjusting the string comparator value upward by a fixed amount if the first four characters agreed; by lesser amounts if the first three, two, or one characters agreed.⁵³ (Porter & Winkler 1999 : 192)

⁵³ Nous traduisons : « La seconde amélioration due à Winkler (1990) donne une plus grande valeur à la concordance des premiers caractères d'une chaîne. Cette idée est fondée sur une importante étude empirique menée par Pollock et Zamora (1984) pour le *Chemical Abstracts Service*. L'étude montrait que, généralement, les erreurs se produisent le moins souvent au début d'une chaîne et que les taux d'erreur en fonction de la position du caractère augmentent de manière linéaire plus cette position se déplace vers la droite. Fondamentalement, l'amélioration consistait à ajuster la valeur du comparateur de chaînes en lui ajoutant une valeur donnée si les quatre premiers caractères concordaient, et une valeur moindre si les premiers caractères concordants n'étaient qu'au nombre de trois, deux ou un. »

Pour résumer ce qui vient d'être énoncé dans les derniers paragraphes, il semble que, pour l'identification de mots, les consonnes aient plus de poids que les voyelles et que le début des mots en ait plus que leur fin. Cela confirme ce que nous avons observé lors de l'analyse d'un dictionnaire de faux amis : c'est la trame consonantique qui prime pour décider que deux mots sont proches formellement, mais la dernière consonne peut être différente. De plus, si les quatre premières lettres sont identiques, alors la suite des deux mots n'a plus à être similaire.

2.2.2. La proximité sémantique directe

Nous venons, dans la section précédente, de définir la proximité formelle directe. Nous avons désormais en notre possession un outil objectif qui nous permet de décider si un mot appartenant à un lexique étranger est suffisamment proche formellement d'un mot de la langue maternelle d'un lecteur pour être considéré comme potentiellement identifiable par lui au premier abord.

Nous allons maintenant définir ce que nous entendons par « proximité sémantique directe », et, chose faite, nous serons en mesure de déclarer que deux mots appartenant à deux lexiques différents sont ou non transparents directement (rappelons que deux mots seront considérés comme transparents directement s'ils sont proches formellement *et* proches sémantiquement directement).

Il peut sembler logique de poser que si deux mots sont proches formellement lorsqu'ils ont (quasiment) la même forme, alors deux mots sont proches sémantiquement lorsqu'ils ont (quasiment) la même signification. Cette affirmation n'est cependant en rien une conclusion, et soulève au contraire bon nombre de questions, notamment parce que la possibilité d'identité de signification entre deux mots n'est rien moins qu'évidente. Pouvons-nous dire que deux mots de langues différentes ont la même signification, et si oui, dans quelle mesure ?

Commençons par une définition négative. Nous ne dirons pas que deux mots de langues différentes ont la même signification s'ils renvoient à la même unité extralinguistique. Cette position semble être avancée par M. Lederer dans un ouvrage

traitant de la Théorie Interprétative, théorie de la traduction enseignée à l'ESIT (École Supérieure d'Interprétation et de Traduction) :

La Théorie Interprétative de la Traduction, corroborée par l'expérience, pose que ce sont les désignations des « choses »³ qui doivent être réexprimées. Le traducteur fait comprendre ce qu'est un fromage par un ou plusieurs mots appropriés à la « chose », ceux-ci ne correspondant pas obligatoirement à un mot donné de l'autre langue.

³ *De nos jours, on dit plus volontiers référent que chose.*

(Lederer 2006 : 73)

Cette idée se trouve résumée par S. Auroux *et al.* dans leur ouvrage *La philosophie du langage* : « Après tout il n'y a qu'un monde, et lorsque nous parlons des entités qui le constituent, quelle que soit la langue utilisée nous parlons bien de la même chose » (Auroux *et al.* 2004 : 192). Une telle position suppose une approche référentielle du langage, à laquelle nous n'adhérons pas. Affirmer qu'un mot réfère à quelque chose, c'est affirmer qu'il y a quelque chose à quoi l'on réfère. Or, nous ne saurions postuler l'existence d'objets du monde préalables à la langue, auxquels la langue donnerait, dans un second temps, un nom. R. Jakobson, dans le chapitre « Aspects linguistiques de la traduction » des *Essais de linguistique générale*, indique bien, en s'appuyant également sur l'exemple du mot « fromage » que la signification d'un mot n'est pas à chercher dans sa relation à un référent :

Le sens des mots français fromage, pomme, nectar, connaissance, mais, seulement, ou de n'importe quel autre mot ou groupe de mots est décidément un fait linguistique [...]. Contre ceux qui assignent le sens (le signifié) non au signe, mais à la chose elle-même, le meilleur argument, et le plus simple, serait de dire que personne n'a jamais goûté ni humé le sens de fromage ou de pomme. Il n'y a pas de signifié sans signe. On ne peut inférer le sens du mot fromage d'une connaissance non linguistique du roquefort ou du camembert sans l'assistance du code verbal. (Jakobson 1963 : 78-79 ; nous soulignons)

Dans son article « Sens, référence et existence : que faire de l'extra-linguistique ? », G. Kleiber affine l'approche référentielle du langage en précisant que l'on ne peut attribuer une existence objective aux objets du monde :

[...] ce que nous croyons être le monde réel n'est que le monde tel que nous le percevons ou tel que nous croyons qu'il est. [...] Nous croyons que ce monde existe avec son organisation ontologique et si l'on parle d'une expression référentielle comme renvoyant à telle entité du monde réel, peu importe que ce ne soit que dans notre modèle phénoménologique du monde : nous croyons que cette entité fait partie du monde réel, nous croyons qu'elle existe vraiment. (Kleiber 1997 : 13)

Cette précision n'est cependant pas suffisante. Non seulement les mots ne renvoient pas à des éléments du monde qui seraient objectifs, mais ils ne renvoient pas non plus à des éléments du monde que nous savons subjectifs. Cette croyance tire sa justification de l'idée que nous percevrions tous la même chose :

Cela se justifie d'autant plus que la conceptualisation ou la modélisation du monde apparaît comme objective, c'est-à-dire ne se trouve pas soumise aux variations subjectives d'un sujet percevant à l'autre, mais bénéficie d'une certaine stabilité intersubjective à l'origine de ce sentiment d'« objectivité » que peut dégager ce monde « projeté ». (ibid. : 13)

Affirmer ceci, c'est tirer un trait sur toutes les études comparatives qui montrent combien la perception du monde dépend de la langue et ainsi sur ce que l'on a coutume d'appeler « l'hypothèse Sapir-Whorf » :

Chaque langue est un vaste système de structures, différent de celui des autres [langues], dans lequel sont ordonnées culturellement les formes et les catégories par lesquelles l'individu non seulement communique, mais aussi analyse la nature, aperçoit ou néglige tel ou tel type de phénomènes ou de relations, dans lesquelles il coule sa façon de raisonner, et par lesquelles il construit l'édifice de sa connaissance du monde. (Whorf⁵⁴, cité par Mounin 1976 : 47)

⁵⁴ Benjamin Lee Whorf, *Language, thought and reality*, New York : Wiley & sons, Londres : Chapman & Hall, 1958, XII-278 p.

Comme le résume V. Nyckees dans son article « Les mots, les choses... et nous », « la structuration du monde opérée par les langues n'a d'autre fondement que les habitudes linguistiques » (Nyckees 1999-2000 : 26).

L'ordre « perception » – « croyance d'existence » doit alors être inversé : nous ne percevons pas des choses, que nous nommons ; la langue met à notre disposition un certain découpage du monde, qui est, tant que nous ne connaissons qu'une seule langue, le seul concevable. Le monde dans lequel nous évoluons n'est pas un monde perçu, c'est un monde parlé.

Une conception référentielle du langage ne saurait être convaincante dans la perspective qui est la nôtre, à savoir une approche comparée de la signification entre plusieurs lexiques, comme le signale F. Rastier : « Le postulat référentiel permet de maintenir un solide objectivisme et empêche de concevoir des sémantiques différenciées selon les langues » (Rastier 2007). Ainsi, la formation au comparatisme qu'avait reçue Saussure a été décisive pour l'élaboration de sa théorie sur la valeur du signe linguistique. Pour l'illustrer, apparaît d'abord dans le *Cours de linguistique générale* l'exemple désormais bien connu d'une différence entre le français et l'anglais : « mouton » et *sheep* ne sauraient avoir la même valeur, « parce qu'en parlant d'une pièce de viande apprêtée et servie sur la table, l'anglais dit *mutton* et non *sheep*. La différence de valeur entre *sheep* et *mouton* tient à ce que le premier a à côté de lui un second terme, ce qui n'est pas le cas pour le mot français » (Saussure 1995 : 160). Cette valeur différentielle des signes est également valable au sein d'une même langue : « des synonymes comme *redouter*, *craindre*, *avoir peur* n'ont de valeur propre que par leur opposition ; si *redouter* n'existait pas, tout son contenu irait à ses concurrents » (*ibid.* : 160). Au sein d'un même système linguistique, les mots ne se définissent pas par leur rapport au monde, mais par leur rapport aux autres mots du système :

Tout ce qui précède revient à dire que dans la langue il n'y a que des différences. Bien plus : une différence suppose en général des termes positifs entre lesquels elle s'établit ; mais dans la langue il n'y a que des différences sans termes positifs. Qu'on prenne le signifié ou le signifiant, la langue ne comporte ni des idées ni des sons qui préexisteraient au système linguistique, mais seulement des différences conceptuelles et des différences phoniques issues de ce système. (Saussure 1995 : 166)

La même idée est développée autrement dans les *Écrits de linguistique générale* :

C'est pourquoi vouloir épuiser les idées contenues dans un mot est une entreprise parfaitement chimérique, à moins peut-être de se borner à des noms d'objets matériels et d'objets tout à fait rares, par exemple l'aluminium, l'eucalyptus, etc. Déjà si l'on prend le fer et le chêne, on n'arrivera pas au bout de la somme de significations (ou d'emplois, ce qui est la même chose) que nous donnons à ces mots, et rien que la comparaison de fer avec deux ou trois mots comme acier, plomb, or ou métal, rien que la comparaison de chêne avec deux ou trois mots comme saule, vigne, bois ou arbre représente un infini travail. Quant à épuiser ce qui est contenu dans esprit par opposition à âme ou à pensée, ou ce qui est contenu dans aller par opposition à marcher, passer, cheminer, se porter, venir ou se rendre, une vie humaine pourrait sans exagération s'y passer. Or, comme dès l'âge de quinze ou seize ans nous avons un sens aiguisé de ce qui est contenu non seulement dans ces mots, mais dans des milliers d'autres, il est évident que ce sens repose sur le pur fait négatif de l'opposition des valeurs, vu que le temps matériellement nécessaire pour connaître la valeur positive des signes nous aurait cent fois et mille fois manqué. (Saussure 2002 : 77)

Il devient alors évident que les mots ne renvoient pas à des entités perçues :

Ainsi, dans une langue composée au total de deux signes, ba et la, la totalité des perceptions confuses de l'esprit viendra NECESSAIREMENT se ranger ou sous ba ou sous la. L'esprit trouvera, du simple fait qu'il existe une différence ba/la et qu'il n'en existe pas d'autre, un caractère distinctif lui permettant régulièrement de tout classer sous le premier ou sous un des deux chapitres (par exemple la distinction de solide et de non solide) ; à ce moment la somme de sa connaissance positive sera représentée par le caractère commun qu'il se trouve avoir attribué aux choses ba et le caractère commun qu'il se trouve avoir attribué aux choses la ; ce caractère est positif, mais il n'a jamais cherché en réalité que le caractère négatif qui pût permettre de décider entre ba et la ; il n'a point essayé de réunir et de coordonner, il a uniquement voulu différencier. (ibid. : 88)

Qu'est-ce alors que la signification d'un mot ?

Nous disons, avec F. Rastier, que cette question n'est pas fondamentale. Ce qui est à définir, c'est en réalité le *sens* d'un mot. L'auteur précise l'opposition entre les deux notions : « on peut appeler *signification* le contenu supposé invariant du mot et désigner par *sens* ses acceptions ou ses emplois en contexte » (Rastier 2003). Or, un mot n'apparaît jamais qu'en texte :

[...] le signe isolé est un artefact : le signe isolé n'est pas observé empiriquement et c'est une décision méthodologique d'isoler un signe. En revanche, les énoncés empiriques sont des textes oraux ou écrits, ou des passages de ces textes. (Rastier 2003)

Nous ne pouvons parler que *du* sens d'un mot dans *un* texte précis. Les locuteurs sont néanmoins tous capables de définir « ce que veut dire » un mot, sa signification, conçue comme valable pour tous ses emplois ou un certain nombre d'entre eux. La signification doit alors être comprise comme la somme des traits significatifs communs aux différents emplois du mot dans ses différentes occurrences. C'est ce qu'annonçait Saussure dans la citation donnée ci-dessus : « à ce moment la somme de sa connaissance positive sera représentée par le caractère commun qu'il se trouve avoir attribué aux choses *ba* et le caractère commun qu'il se trouve avoir attribué aux choses *la* ». F. Rastier précise :

En somme, la hiérarchie entre sens et signification pourrait être inversée. Le sens n'est pas de la signification déformée par le contexte : la signification ne serait plus un type diversement déformé dans ses occurrences qui constituent les sens, mais du sens normalisé car coupé de son contexte. Le type devient alors une collection d'accidents, un résumé conventionnel des occurrences retenues comme pertinentes pour sa définition. (ibid.)

Puisque nous n'avons affaire qu'à des occurrences en texte, ce sont à des sens que nous avons accès. La recherche de signification se trouve ainsi transposée au titre d'exercice artificiel, réalisé méthodiquement par les lexicographes, ou par les locuteurs dans leur diversité lorsqu'ils ont à réfléchir à « ce que veut dire » un mot isolé.

Le texte, et ainsi les sens, étant premiers, c'est d'un point de vue textuel que nous allons tenter de définir la proximité sémantique entre deux mots de langues différentes.

Nous avons, pour définir la proximité formelle directe, transposé des notions utilisées en analyse intralinguistique, à savoir l'homonymie et la paronymie, à une analyse interlinguistique. C'est ce qu'indique M. Ballard, dans l'introduction à son dictionnaire de faux amis anglais/français : il présente l'homonymie et la paronymie intralinguistiques pour ensuite présenter l'homonymie et la paronymie interlinguistiques « car [les deux phénomènes] procèdent du même principe et s'éclairent mutuellement » (Ballard 1999 : 9). Par exemple, au niveau intralinguistique, sont homonymes en français « sein » et « saint », en anglais « *cue* » et « *queue* » ; sont paronymes en français « conjoncture » et « conjecture », en anglais « *to rise* » et « *to raise* ». Au niveau interlinguistique, sont homonymes l'anglais « *bribe* » et le français « bribe » ; sont paronymes l'anglais « *actual* » et le français « actuel ».

C'est à l'aide de ce même principe, c'est-à-dire le fait d'éclairer un processus interlinguistique par des analyses déjà menées dans un cadre intralinguistique, que nous proposons de définir la similitude sémantique entre deux mots de langues étrangères. Au sein d'un même système linguistique, deux mots qui ont la même signification sont appelés « synonymes ». On sait toutefois à quel point cette notion est délicate à manier. La synonymie absolue est en effet extrêmement rare, pour les raisons que cite J. Picoche :

*On ne [...] trouve guère [de cas de synonymie absolue] que dans les langues scientifiques qui, au cours de leur élaboration, forgent parfois deux mots pour un seul et même concept. [...] Dans la plupart des autres cas, le phénomène de la synonymie est troublé par les faits suivants : 1) Les mots envisagés peuvent être polysémiques, et synonymes par une seule de leurs acceptions. Il peut même se faire que certains mots ne soient synonymes que dans des emplois métaphoriques et dans des contextes tout à fait particuliers : pain, vie, bifteck ne sont guère synonymes que comme compléments du verbe gagner ! 2) Un des mots peut être marqué par rapport à l'autre, c'est-à-dire comporter un trait pertinent de plus, appelé la **marque** : matou désigne le chat mâle adulte entier, alors que chat peut désigner le même animal ou toute l'espèce. 3) Les deux mots envisagés peuvent s'opposer par leur valeur stylistique : migraine et céphalée, synonymes par leur contenu sémantique, ne peuvent s'employer dans les mêmes contextes, l'un étant populaire, l'autre savant [...]* (Picoche 1997 : 100)

Pour les trois raisons données dans cette citation, on préférera parler de « parasynonymie ». J. Rey-Debove, dans son article « La synonymie ou les échanges de signes comme fondement de la sémantique » publié en 1997, appelle « quasi-synonymes » ces mots dont on ne peut pas dire qu'ils sont exactement synonymes (Rey-Debove 1997 : 93-95).

Deux mots ne pouvant être parfaitement synonymes, ils seront dits parasynonymes si la présence de l'un ou de l'autre dans un contexte identique par ailleurs donne des énoncés de même sens. C'est ce qu'illustre la citation suivante de V. Nyckees (il y utilise cependant le terme de synonymie, en ayant expliqué auparavant les raisons de son choix) :

Le seul critère décisif [...] réside dans la substitution en contexte : deux ou plusieurs termes ou expressions seront dits synonymes si la substitution de l'un à l'autre dans un même énoncé permet d'obtenir deux énoncés de même signification. Ainsi redouter et craindre : Jean craint les orages. Jean redoute les orages. (Nyckees 1998 : 180-181)

La parasynonymie intralinguistique peut donc être définie comme la relation qui unit deux mots X et Y qui, « étant chacun inséré dans le même énoncé ou entourage, $abcXd$ et $abcYd$ » (Rey-Debove 1966 : 86), permettent d'avoir deux énoncés de même sens. Par suite, la parasynonymie interlinguistique peut être définie comme la relation qui unit deux mots de langues différentes qui, le contexte étant équivalent par ailleurs, permettent d'avoir deux énoncés de même sens, c'est-à-dire deux mots dont l'un constitue la traduction de l'autre dans deux textes dont l'un est la traduction de l'autre.

Par exemple, de même que la substitution de « craindre » à « redouter » permet d'obtenir deux énoncés de même sens dans « Jean craint les orages » et « Jean redoute les orages », de même la substitution de « craindre » au verbe anglais *to fear* permet d'obtenir deux énoncés qui sont considérés comme ayant le même sens, c'est-à-dire qui sont une traduction l'un de l'autre, dans les énoncés « Jean craint les orages » et « Jean fears the storms ». Le passage d'une langue à l'autre entraîne inévitablement une acrobatie : pour déterminer l'équivalence de sens entre deux mots X et Y, il faut les insérer dans le même contexte, qui se trouve être ici, non pas $abcXd$ et $abcYd$, mais $abcXd$ et $a'b'c'Yd'$, où l'on remplace chacun des mots par son parasynonyme interlinguistique, en postulant ainsi une équivalence des contextes. Il faut donc se situer, pour définir la parasynonymie

interlinguistique, dans le cas où l'on choisit de traduire un mot par un autre, cas qui n'est pas le plus général.

Nous respectons l'approche textuelle que nous annonçons ci-dessus, puisque la parasynonymie interlinguistique consiste en une équivalence *en texte*. De même que deux parasynonymes intralinguistiques ne peuvent pas toujours se substituer l'un à l'autre, deux parasynonymes interlinguistiques ne peuvent pas toujours être la traduction l'un de l'autre. C'est une mise en équivalence de textes qui permet une mise en équivalence de mots.

En des termes quasiment similaires, R. Jakobson indique que la proximité sémantique qui peut exister entre deux mots au sein d'une même langue ou entre deux mots de langues différentes procède du même principe, celui de l'interprétation :

1) La traduction intralinguale ou reformulation (rewording) consiste en l'interprétation des signes linguistiques au moyen d'autres signes de la même langue.

2) La traduction interlinguale ou traduction proprement dite consiste en l'interprétation des signes linguistiques au moyen d'une autre langue. (Jakobson 1963 : 79⁵⁵)

Il nous faut maintenant définir ce que nous entendons par « mise en équivalence » de deux mots de langues différentes, ou, pour citer à nouveau R. Jakobson, ce qu'est l'« équivalence dans la différence » (*ibid.* : 80).

Une équivalence entre mots se définit, toujours selon notre approche, au sein de textes. Elle est ainsi instituée par un traducteur au moment où il rédige sa traduction ; il décide alors de traduire un mot A par un mot B, non parce qu'ils ont la même valeur (c'est impossible puisque A et B appartiennent à des systèmes linguistiques différents), mais parce que B rend le mieux compte, dans son système d'oppositions, de la valeur qu'a A à l'intérieur de son propre système d'oppositions. M. Ballard, dans son article intitulé « Éléments pour la structuration de l'équivalence : point de vue traductologique », définit les « éléments impliqués dans l'équivalence » comme des « formes dont la valeur peut être considérée comme substituable » (Ballard 2005 : 150). La mise en équivalence tient à la prise en compte aussi bien du contexte syntagmatique que du contexte paradigmatique. Le

⁵⁵ Il distingue également la « traduction intersémiotique ou transmutation [qui] consiste en l'interprétation des signes linguistiques au moyen de systèmes de signes non linguistiques. » (*op.cit.* : 79)

premier se définit « par des relations positionnelles » et le second « par des relations oppositionnelles » (Rastier 2005). En d'autres termes, B doit convenir dans son contexte aussi bien que A convient dans le sien, et le choix de B par rapport aux mots qui auraient pu être choisis par le traducteur doit être guidé par les raisons qui ont conduit au choix de A par rapport aux autres mots qui auraient pu être choisis par l'auteur du texte source.

Dans leur article « De la fidélité en matière de traduction littéraire : esquisse d'une méthode d'évaluation comparative », J. François et Y. Keromnès précisent davantage quels sont les différents niveaux d'équivalences à respecter :

[...] la tâche du traducteur consiste [...] à repérer les propriétés (sémantiques, discursives, pragmatiques, stylistiques, éventuellement rhétoriques) de chacune des unités linguistiques du texte-source DANS LEUR CHAMP PARAPHRASTIQUE et à sélectionner dans la production du texte-cible les unités linguistiques qui, chacune dans son champ paraphrastique, respectent le mieux ces propriétés. (François & Keromnès 1995 : 43)

A posteriori, une équivalence entre mots peut être dégagée lors de la comparaison de textes dont l'un est la traduction de l'autre : il s'agit de la technique de l'« alignement ». « Les techniques d'alignement [...] ont pour but de mettre en correspondance, par un traitement automatique, les portions de textes qui sont traductions les unes des autres » (Kraif 1999). Une fois que deux textes ont été alignés, le « repérage de traduction » (« *translation spotting* ») est une « tâche [qui] consiste simplement à déterminer, étant donné une unité du texte source, quelle est l'unité ou l'expression équivalente » (Kraif 2003)⁵⁶. Il s'agit, toujours dans le cas de la traduction d'un mot par un autre, de retrouver après coup l'équivalence qui avait été instituée par le traducteur.

Nous venons de voir que la proximité sémantique entre deux mots de langues différentes se définit « en texte », soit lors d'un choix de traduction, soit lors d'un alignement. Ainsi, nous dirons que les *sens* de deux mots de langues différentes peuvent être considérés comme proches directement lorsque l'un des deux mots constitue un équivalent de l'autre mot dans une traduction, et nous dirons, la signification se définissant comme un « résumé conventionnel » (Rastier 2003), que les *significations* de deux mots de

⁵⁶ Pour les difficultés à pratiquer un repérage de traduction au niveau du mot, voir Véronis & Langlais 2000 : 382-383.

langues différentes peuvent être conventionnellement considérées comme proches directement lorsque l'un des deux mots constitue régulièrement un équivalent de l'autre mot dans des traductions différentes.

Nous venons de définir la transparence directe des points de vue formel et sémantique. Cette transparence est directe parce qu'elle rend compte des mots *directement* compréhensibles par des lecteurs, c'est-à-dire des mots qui sont traditionnellement⁵⁷ considérés comme ayant une forme et une signification (quasi-)identiques de mots de leur langue. Ce chapitre avait pour objet de donner une définition précise de ce que signifie (quasi-)identité.

⁵⁷ Voir le chapitre 1.1. pour un examen des définitions traditionnelles de la transparence.

2.3. La transparence indirecte

Nous venons, dans le chapitre précédent, de définir la transparence directe. Nous avons pour cela donné les règles qui décident de la proximité formelle directe et de la proximité sémantique directe. Le présent chapitre sera consacré à la définition de la transparence indirecte, c'est-à-dire à l'extension de la notion de transparence. Comme dans le chapitre précédent, nous procéderons en deux temps : nous définirons la transparence indirecte d'abord par son versant formel puis par son versant sémantique.

2.3.1. La proximité formelle indirecte : l'ajustement formel

Nous avons vu dans la section 1.2.3. que des équipes travaillant sur l'intercompréhension ont mis en lumière le fait que la transparence ne se limite pas à la transparence directe. Les lois énoncées pour définir cette dernière, tirées de l'observation d'un dictionnaire de faux amis, donc de mots réputés proches formellement, ne suffisent pas à rendre compte de tout ce qui est mis à profit par les apprenants lorsqu'ils tentent de mettre un mot apparemment inconnu d'eux en rapport avec un mot qu'ils connaissent. Ils mettent notamment en œuvre une méthode que nous appelons l'« ajustement formel », qui consiste à manipuler certains phonèmes des mots, et que nous allons théoriser⁵⁸. À l'aide de cette méthode, un nombre plus important de mots que ceux qui sont immédiatement reconnaissables par un lecteur débutant sont considérés comme proches formellement. Puisque le nombre de mots reconnaissables augmente grâce à certaines manipulations, nous appelons cette situation la « proximité formelle indirecte ».

L'ajustement formel est une méthode qui peut permettre de reconnaître un mot qui paraît au premier abord incompréhensible. Il s'agit de mettre en pratique la « souplesse de lecture » que nous avons décrite p. 56, c'est-à-dire de ne pas se focaliser sur la forme telle qu'elle est exactement perceptible, mais de s'autoriser à changer un ou plusieurs phonèmes du mot jusqu'à aboutir à une forme qui fasse penser à un mot de la langue maternelle ou d'une autre langue étrangère connue, pour ainsi pouvoir proposer une interprétation pour ce mot. Ces changements sont évidemment soumis à certaines règles, que nous allons

⁵⁸ Voir aussi : Castagne 2007, Castagne *et al.* 2007 et Caure *et al.* 2008.

développer, sans quoi n'importe quelle forme pourrait conduire à n'importe quelle autre par le biais de transformations diverses.

Nous proposons certes une méthode de manipulation des phonèmes, mais uniquement pour les consonnes. Nous avons vu, lors de la définition de la proximité formelle directe, que les consonnes d'un mot constituent le point d'appui nécessaire et suffisant pour une reconnaissance de ce mot. Les voyelles peuvent apporter une aide supplémentaire, mais il n'est pas indispensable qu'elles soient identiques entre un mot étranger et un mot français pour que le mot étranger soit identifié. Nous partons du même point de départ, c'est-à-dire que nous nous concentrons sur les consonnes, et ce sont elles qui subissent les manipulations. Nous ne conseillons pas de faire abstraction des voyelles, puisque si elles sont identiques, elles constituent un appui de plus, mais il faut en premier lieu chercher à tirer parti des consonnes. L'ajustement formel consiste à remplacer les phonèmes consonantiques d'un mot *a priori* non compréhensible pour parvenir à la forme d'un mot connu. Il ne s'agit pas de remplacer les phonèmes consonantiques par tous les autres, car dans ce cas l'ajustement que nous proposons s'apparenterait plutôt à une loterie où seul le hasard laisse une chance de réussir, mais par ceux qui en sont proches, c'est-à-dire qui appartiennent à sa zone de proximité articulatoire.

Avant d'aller plus avant dans la présentation de cette méthode, rappelons que graphie et phonie sont nécessairement liées. Si les lecteurs peuvent faire des manipulations sur la forme phonologique des mots, c'est parce qu'ils associent la face signifiée (le phonème) à la face signifiante (le graphème) du signe graphique qu'ils ont sous les yeux. Nous avons exposé dans la partie « proximité formelle directe » pourquoi un graphème est « forcément » – selon le terme d'É. Benveniste – lié au phonème qu'il note (voir p. 103). Cependant, il s'agissait alors de graphèmes de la langue maternelle du lecteur. Dans notre optique, le même phénomène est également valable lors de la lecture de mots écrits en langue étrangère, puisque les apprenants ICE disposent toujours de la version lue des textes étudiés. De plus, les orthographe des trois langues germaniques qui nous intéressent sont majoritairement phonogrammiques, et ce particulièrement pour la notation des phonèmes consonantiques. Nous avons spécifié que l'attention devait être portée sur les consonnes, pour la proximité formelle directe aussi bien que pour la proximité formelle indirecte, et ce sont précisément pour les consonnes que les correspondances graphophonématiques sont les plus fortes, tant en anglais :

*Most cases of irregularity in the correspondences between graphemes and phonemes arise from inconsistencies in the mapping of vowels.*⁵⁹ (Berent & Perfetti 1995 : 148) ;

qu'en allemand :

La notation des phonèmes consonantiques présente moins de difficultés [que celle des phonèmes vocaliques] du fait qu'il y a dans les termes du fonds germanique (à l'exception des noms propres) une relative correspondance entre lettres et phonèmes (Schanen & Confais 1999 : 60) ;

et en néerlandais :

*The first problem for Dutch orthography is the lack of a sufficient number of letters for its sounds. This applies in particular to vowels: there are only five vowel letters (i,u,e,o,a) for the thirteen Dutch vowels, the diphthongs not included. This has the effect that there is no one-to-one correspondence between sounds and letters. [...] The spelling of the consonants is more straightforward.*⁶⁰ (Booij 1995 : 181)

Ces précisions sur la possibilité de jouer sur les phonèmes consonantiques de mots écrits en langue étrangère étant données, il est possible de présenter l'ajustement formel de façon plus précise. Pour ce faire, commençons par donner la liste des phonèmes consonantiques de chacune des langues qui nous intéressent. Nous les présentons sous forme de tableaux, qui reprennent la présentation et la terminologie du tableau 2005 (date de dernière révision) de l'Association de Phonétique Internationale dont nous reproduisons une copie ci-après⁶¹.

⁵⁹ Nous traduisons : « La plupart des cas d'irrégularité dans les correspondances entre graphèmes et phonèmes viennent des divergences dans la représentation des voyelles. »

⁶⁰ Nous traduisons : « Le premier problème de l'orthographe néerlandaise est le nombre insuffisant de lettres par rapport aux sons. Cela est particulièrement vrai pour les voyelles : il y a seulement cinq lettres vocaliques (i,u,e,o,a) pour les treize voyelles du néerlandais, diphtongues non comprises. Par conséquent, il n'y a pas de correspondance une à une entre les sons et les lettres. [...] L'orthographe des consonnes est plus nette. »

⁶¹ [http://www2.arts.gla.ac.uk/IPA/IPA_chart_\(C\)2005.pdf](http://www2.arts.gla.ac.uk/IPA/IPA_chart_(C)2005.pdf) (consultée le 08/01/09)

THE INTERNATIONAL PHONETIC ALPHABET (2005)

CONSONANTS (PULMONIC)

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Epi-glottal	Glottal
Nasal	m	ɱ	n		ɳ	ɺ	ɲ	ŋ	ɴ			
Plosive	p b	ɸ β	t d		ʈ ɖ	ʈ ɖ	c ɟ	k ɡ	q ɢ			
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	ħ ʕ	h ɦ
Approximant		ʋ	ɹ		ɻ	j	ɰ		ʁ			
Trill	ʙ		r						ʀ			
Tap, Flap		ⱱ	ɾ		ɽ							
Lateral fricative			ɬ ɮ		ɬ ɮ	ɬ ɮ	ɬ ɮ					
Lateral approximant			l		ɭ	ɭ	ɭ	ɭ				
Lateral flap			ɺ		ɺ	ɺ	ɺ	ɺ				

Where symbols appear in pairs, the one to the right represents a modally voiced consonant, except for murmured *f*. Shaded areas denote articulations judged to be impossible. Light grey letters are unofficial extensions of the IPA.

CONSONANTS (NON-PULMONIC)

Anterior click releases (require posterior stops)	Voiced implosives	Ejectives
⦿ Bilabial fricated	ɓ Bilabial	ʼ <i>Examples:</i>
ɮ Laminar alveolar fricated ("dental")	ɗ Dental or alveolar	ɸ' Bilabial
ɠ Apical (post)alveolar abrupt ("retroflex")	ɟ Palatal	ʈ' Dental or alveolar
ɰ Laminar postalveolar abrupt ("palatal")	ɠ Velar	ɰ' Velar
ɮ Lateral alveolar fricated ("lateral")	ɠ Uvular	ʂ' Alveolar fricative

CONSONANTS (CO-ARTICULATED)

- ɱ Voiceless labialized velar approximant
- ʋ Voiced labialized velar approximant
- ɰ Voiced labialized palatal approximant
- ɠ Voiceless palatalized postalveolar (alveolo-palatal) fricative
- ʒ Voiced palatalized postalveolar (alveolo-palatal) fricative
- ɧ Simultaneous x and ʃ (disputed)
- kp̚ ts̚ Affricates and double articulations may be joined by a tie bar

Nous allons maintenant présenter les phonèmes consonantiques des quatre langues dans quatre tableaux successifs en nous appuyant sur le tableau de l'API, que nous complétons pour que tous les traits pertinents soient visibles. Les phonèmes y sont placés comme ils l'ont été par l'API, par souci d'unification. Ainsi, par exemple, bien que la réalisation du /t/ français soit traditionnellement considérée comme étant dentale et celle du /t/ anglais comme étant alvéolaire, ils apparaissent au même endroit dans les tableaux, celui choisi par l'API. Cette présentation unifiée nous permettra ensuite de comparer les phonèmes entre eux.

Apparaîtront à la suite des tableaux les questions soulevées lors de leur élaboration.

		Labiale		Apicale			Dorsale			Glottale
		Labiale	Dentale	Dentale	Alvéolaire	Post-alvéolaire	Palatale	Vélaire	Uvulaire	
Occlusive	Sourde	p		t				k		
	Sonore	b		d				g		
	Nasale	m		n			ɲ	ŋ		
Affriquée	Sourde									
	Sonore									
Fricative	Sourde		f		s	ʃ			ʁ	
	Sonore		v		z	ʒ				
Approximante	Sonore						j ɥ	w		
Roulée	Sourde/Sonore									
Latérale approximante	Sourde/Sonore			l						

Tableau 1 : consonnes du français

		Labiale		Apicale			Dorsale			Glottale
		Labiale	Dentale	Dentale	Alvéolaire	Post-alvéolaire	Palatale	Vélaire	Uvulaire	
Occlusive	Sourde	p		t				k		
	Sonore	b		d				g		
	Nasale	m		n				ŋ		
Affriquée	Sourde					tʃ				
	Sonore					dʒ				
Fricative	Sourde		f	θ	s	ʃ			h	
	Sonore		v	ð	z	ʒ				
Approximante	Sonore						j	w		
Roulée	Sourde/Sonore			r						
Latérale approximante	Sourde/Sonore			l						

Tableau 2 : consonnes de l'anglais

		Labiale		Apicale			Dorsale			Glottale
		Labiale	Dentale	Dentale	Alvéolaire	Post-alvéolaire	Palatale	Vélaire	Uvulaire	
Occlusive	Sourde	p		t				k		
	Sonore	b		d				g		
	Nasale	m		n				ŋ		
Affriquée	Sourde		pf		ts					
	Sonore									
Fricative	Sourde		f		s	ʃ	ç	x	ɣ	h
	Sonore		v		z	ʒ				
Approximante	Sonore						j			
Roulée	Sourde/Sonore			r						
Latérale approximante	Sourde/Sonore			l						

Tableau 3 : consonnes de l'allemand

		Labiale		Apicale			Dorsale			Glottale
		Labiale	Dentale	Dentale	Alvéolaire	Post-alvéolaire	Palatale	Vélaire	Uvulaire	
Occlusive	Sourde	p		t				k		
	Sonore	b		d				g		
	Nasale	m		n			ɲ	ŋ		
Affriquée	Sourde									
	Sonore									
Fricative	Sourde		f		s	ʃ		x	ɣ	h
	Sonore		v		z	ʒ		y		
Approximante	Sonore		ʋ				j			
Roulée	Sourde/Sonore			r						
Latérale approximante	Sourde/Sonore			l						

Tableau 4 : consonnes du néerlandais

Constituer ces tableaux, c'est-à-dire décider des éléments qui devaient y apparaître, a soulevé des questions auxquelles il nous a fallu trouver des réponses :

- Nous y avons intégré les semi-consonnes (ou semi-voyelles ou glides) /w/ et /ɥ/, au même titre que le /j/ qui apparaît dans le tableau de l'API. Elles figurent dans cette charte de l'API en tant que « co-articulated consonants ».
- Nous faisons figurer dans ces tableaux les phonèmes et non les phones, puisque ce sont bien des phonèmes que transcrivent les graphèmes. Néanmoins, puisque nous voulons dégager des zones de proximité articulatoire, le /r/ apparaît deux fois dans les tableaux de l'allemand et du néerlandais, étant donné qu'il possède deux zones d'articulation bien distinctes à l'intérieur de ces langues, selon les aires géographiques : une zone dentale-alvéolaire et une zone uvulaire. Nous n'avons pas fait figurer toutes les réalisations possibles (/r/, /r̥/, /ʀ/, /ʁ/ et /ʁ̥/ par exemple pour le néerlandais), mais gardé les deux /r/ qui, non seulement représentaient les deux zones d'articulation, mais aussi que l'on retrouvait dans les autres langues. De même, nous avons noté le /r/ de l'anglais ainsi (bien qu'il puisse être noté /ɹ/ ou /r/) pour que soit bien visible la correspondance avec les autres langues. En effet, nous allons ensuite présenter un tableau récapitulatif des phonèmes des quatre langues, et avoir trois /r/ dans une même zone de proximité constitue une surabondance d'informations inutile pour notre propos.
- Nous avons également ajouté les affriquées, puisqu'elles sont considérées comme des phonèmes dans la majorité des ouvrages sur la phonologie de ces langues, et non comme des groupes consonantiques.
- Enfin, nous avons inclus les phonèmes qui n'apparaissent dans les langues que dans les mots d'origine étrangère (tel que le /ʒ/ en allemand), car ils sont connus des locuteurs au même titre que les mots d'origine de la langue concernée.

Nous donnons maintenant ci-après les phonèmes des quatre langues représentés en un seul tableau.

		Labiale		Apicale			Dorsale			Glottale
		Labiale	Dentale	Dentale	Alvéolaire	Post-alvéolaire	Palatale	Vélaire	Uvulaire	
Occlusive	Sourde	p		t				k		
	Sonore	b		d				g		
	Nasale	m		n			ɲ	ŋ		
Affriquée	Sourde		pf		ts	tʃ				
	Sonore					ʤ				
Fricative	Sourde		f	θ	s	ʃ	ç	x	ɣ	h
	Sonore		v	ð	z	ʒ		ʁ		
Approximante	Sonore		ʋ				j ɥ	w		
Roulée	Sourde/Sonore			r						
Latérale approximante	Sourde/Sonore			l						

Tableau 5 : consonnes des quatre langues

Avoir sous les yeux, réunies en un seul tableau, toutes les consonnes des quatre langues va nous aider à présenter l'ajustement formel de façon précise. Comme nous l'avons déjà stipulé, cet ajustement consiste à changer un ou plusieurs phonèmes consonantiques d'un mot jusqu'à aboutir à une nouvelle forme, qui, étant proche d'une forme en langue maternelle (ou dans une autre langue connue), suggère une interprétation. Il ne s'agit pas de remplacer une consonne par toutes les autres, mais par celles situées dans sa zone de proximité articulatoire. Le tableau ci-dessus permet justement de se représenter ces zones.

Si nous avons choisi de remplacer des consonnes par d'autres qui s'articulent en un lieu proche dans le chenal buccal, c'est parce que ce sont ces dernières qui sont senties comme étant en relation avec ces premières. Il y a évidemment une explication diachronique, selon laquelle un son évolue phonétiquement en un autre son qui n'en est pas extrêmement éloigné :

C'est qu'en effet le travail de transformation phonétique qui s'opère continuellement, au sein de n'importe quelle langue, reste cependant contenu dans certaines limites naturelles. Par exemple, en mettant les choses au pis, un t arrivera peut-être à travers beaucoup d'étapes à donner un l, mais un p a une chance égale à zéro de jamais donner l dans n'importe quel dialecte, même en quatre ou cinq mille ans. (Saussure 2002 : 268)

Cependant, puisque nous tenons à proposer une méthode fondée sur la synchronie actuelle, cette explication n'est pas recevable pour notre travail. Si les apprenants peuvent faire commuter des consonnes, c'est parce qu'il est possible de prendre conscience des proximités entre sons. Certaines de ces proximités sont très vite identifiées par eux (par ex. /p/-/b/), d'autres demandent une petite gymnastique de l'esprit (par ex. /b/-/v/), mais, à partir de quelques séances, ces manipulations ne constituent plus d'obstacles et sont quasiment parvenues au stade du réflexe.

Nous allons maintenant dresser la liste des consonnes avec lesquelles peuvent, dans le cadre de l'ajustement formel, commuter d'autres consonnes. Pour cela, nous donnons la « définition » de chacune des consonnes en traits articulatoires, en nous aidant du tableau 5 :

/p/ : ‘labiale’ ‘labiale’ ‘occlusive’ ‘sourde’
 /b/ : ‘labiale’ ‘labiale’ ‘occlusive’ ‘sonore’
 /m/ : ‘labiale’ ‘labiale’ ‘occlusive’ ‘nasale’
 /pf/ : ‘labiale’ ‘dentale’ ‘affriquée’ ‘sourde’
 /f/ : ‘labiale’ ‘dentale’ ‘fricative’ ‘sourde’
 /v/ : ‘labiale’ ‘dentale’ ‘fricative’ ‘sonore’
 /ʋ/ : ‘labiale’ ‘dentale’ ‘approximante’ ‘sonore’
 /θ/ : ‘apicale’ ‘dentale’ ‘fricative’ ‘sourde’
 /ð/ : ‘apicale’ ‘dentale’ ‘fricative’ ‘sonore’
 /t/ : ‘apicale’ ‘dentale ou alvéolaire ou post-alvéolaire’ ‘occlusive’ ‘sourde’
 /d/ : ‘apicale’ ‘dentale ou alvéolaire ou post-alvéolaire’ ‘occlusive’ ‘sonore’
 /n/ : ‘apicale’ ‘dentale ou alvéolaire ou post-alvéolaire’ ‘occlusive’ ‘nasale’
 /ts/ : ‘apicale’ ‘alvéolaire’ ‘affriquée’ ‘sourde’
 /s/ : ‘apicale’ ‘alvéolaire’ ‘fricative’ ‘sourde’
 /z/ : ‘apicale’ ‘alvéolaire’ ‘fricative’ ‘sonore’
 /tʃ/ : ‘apicale’ ‘post-alvéolaire’ ‘affriquée’ ‘sourde’
 /dʒ/ : ‘apicale’ ‘post-alvéolaire’ ‘affriquée’ ‘sonore’
 /ʃ/ : ‘apicale’ ‘post-alvéolaire’ ‘fricative’ ‘sourde’
 /ʒ/ : ‘apicale’ ‘post-alvéolaire’ ‘fricative’ ‘sonore’
 /r/ : ‘apicale’ ‘dentale ou alvéolaire ou post-alvéolaire’ ‘roulée’ ‘sourde ou sonore’
 /l/ : ‘apicale’ ‘dentale ou alvéolaire ou post-alvéolaire’ ‘latérale approximante’ ‘sourde ou sonore’
 /ɲ/ : ‘dorsale’ ‘palatale’ ‘occlusive’ ‘nasale’
 /ç/ : ‘dorsale’ ‘palatale’ ‘fricative’ ‘sourde’
 /j/ : ‘dorsale’ ‘palatale’ ‘approximante’ ‘sonore’
 /ɥ/ : ‘dorsale’ ‘palatale’ ‘labialisée’ ‘approximante’ ‘sonore’
 /k/ : ‘dorsale’ ‘vélaire’ ‘occlusive’ ‘sourde’
 /g/ : ‘dorsale’ ‘vélaire’ ‘occlusive’ ‘sonore’
 /ŋ/ : ‘dorsale’ ‘vélaire’ ‘occlusive’ ‘nasale’
 /x/ : ‘dorsale’ ‘vélaire’ ‘fricative’ ‘sourde’

/ɣ/ : ‘dorsale’ ‘vélaire’ ‘fricative’ ‘sonore’

/w/ : ‘dorsale’ ‘vélaire’ ‘labialisée’ ‘approximante’ ‘sonore’

/ʁ/ : ‘dorsale’ ‘uvulaire’ ‘fricative ou approximante’ ‘sourde ou sonore’

/h/ : ‘glottale’ ‘fricative ou approximante’ ‘sourde’

Nous pouvons maintenant définir les « champs phonologiques », c’est-à-dire les ensembles de phonèmes ayant des traits en commun. Nous nous inspirons ainsi du modèle sémantique, dans lequel un « champ » est « un ensemble de mots *distincts* partageant certains traits sémantiques communs » (Nyckees 1998 : 193)⁶². De la même façon, nous disons qu’un phonème a dans son champ tous les autres phonèmes partageant avec lui certains traits articulatoires. Plus précisément, une consonne a dans son champ toutes les autres consonnes qui ont en commun avec elle tous les traits sauf un ou deux, de sorte que l’on ne puisse pas faire varier plus de la moitié des traits définissant la consonne. En d’autres termes, nous faisons varier un ou deux traits de la définition d’une consonne (il s’agit alors de passer au niveau de l’hyperphonème⁶³) pour savoir quelles sont les autres consonnes de son champ, donc avec quelles consonnes elle peut commuter lors de l’ajustement formel. Le trait de l’organe articulant (‘labial’, ‘apical’, ‘dorsal’, ‘glottal’) ne peut, quant à lui, pas varier, sans quoi un phonème aurait dans son champ d’autres phonèmes qui en seraient trop éloignés. C’est pourquoi nous avons donné ce trait en premier dans la définition des phonèmes : ils sont déjà regroupés par leur organe articulant (arrive ensuite le point d’articulation pour compléter le lieu d’articulation et sont enfin donnés les deux modes d’articulation).

Le champ phonologique défini, nous pouvons dresser la liste précise des consonnes qui peuvent commuter avec telle autre. Les champs comprennent les consonnes des quatre langues puisque, nous le rappelons, en séance ICE les langues sont vues simultanément. Les apprenants développent donc une compétence plurilingue et comparative.

⁶² Voir la section « La proximité sémantique indirecte » pour une réflexion sur la notion de « champ sémantique ».

⁶³ Voir le passage sur l’hypersystème p. 87.

Consonne	Consonnes qui partagent ses traits articulatoires sauf un ou deux
/p/	/b/, /m/, /pf/, /f/
/b/	/p/, /m/, /v/, /v/
/m/	/p/, /b/
/pf/	/p/, /f/, /v/, /v/
/f/	/p/, /pf/, /v/, /v/
/v/	/pf/, /f/, /b/, /v/
/v/	/b/, /pf/, /f/, /v/
/θ/	/ð/, /t/, /d/, /n/, /ts/, /s/, /z/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/
/ð/	/θ/, /t/, /d/, /n/, /s/, /z/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/
/t/	/θ/, /ð/, /d/, /n/, /ts/, /s/, /z/, /tʃ/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/
/d/	/θ/, /ð/, /t/, /n/, /ts/, /s/, /z/, /tʃ/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/
/n/	/θ/, /ð/, /t/, /d/, /ts/, /s/, /z/, /tʃ/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/
/ts/	/θ/, /t/, /d/, /n/, /s/, /z/, /tʃ/, /tʃ/, /ʃ/, /r/, /l/
/s/	/θ/, /ð/, /t/, /d/, /n/, /ts/, /z/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/
/z/	/θ/, /ð/, /t/, /d/, /n/, /ts/, /s/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/
/tʃ/	/θ/, /t/, /d/, /n/, /ts/, /s/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/
/tʃ/	/ð/, /t/, /d/, /n/, /ts/, /z/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/
/ʃ/	/θ/, /ð/, /t/, /d/, /ts/, /s/, /z/, /tʃ/, /tʃ/, /ʒ/, /r/, /l/
/ʒ/	/θ/, /ð/, /t/, /d/, /s/, /z/, /tʃ/, /tʃ/, /ʃ/, /r/, /l/
/r/	/θ/, /ð/, /t/, /d/, /n/, /ts/, /s/, /z/, /tʃ/, /tʃ/, /ʃ/, /ʒ/, /l/
/l/	/θ/, /ð/, /t/, /d/, /n/, /ts/, /s/, /z/, /tʃ/, /tʃ/, /ʃ/, /ʒ/, /r/
/ɲ/	/ç/, /j/, /ɥ/, /k/, /g/, /ŋ/
/ç/	/ɲ/, /j/, /ɥ/, /k/, /x/, /ç/, /ʁ/
/j/	/ɲ/, /ç/, /ɥ/, /g/, /ç/, /ʁ/

/ɥ/	/ɲ/, /ç/, /j/, /g/, /ʎ/, /w/, /ʁ/
/k/	/ɲ/, /ç/, /g/, /ŋ/, /x/, /ʎ/, /w/, /ʁ/
/g/	/ɲ/, /j/, /ɥ/, /k/, /ŋ/, /x/, /ʎ/, /w/, /ʁ/
/ŋ/	/ɲ/, /k/, /g/, /x/, /ʎ/, /w/
/x/	/ç/, /k/, /g/, /ŋ/, /ʎ/, /w/, /ʁ/
/ʎ/	/ç/, /j/, /ɥ/, /k/, /g/, /ŋ/, /x/, /w/, /ʁ/
/w/	/ɥ/, /k/, /g/, /ŋ/, /x/, /ʎ/, /ʁ/
/ʁ/	/ç/, /j/, /ɥ/, /k/, /g/, /x/, /ʎ/, /w/
/h/	néant

Nous avons déjà précisé à quel point graphie et phonie étaient liées. C'est pourquoi il nous a semblé pertinent de regrouper des phonèmes qui peuvent se transcrire au moyen du ou des mêmes graphèmes. Voici les principales graphies possibles des phonèmes des quatre langues :

Consonne	Principales graphies possibles
/p/	<p>, , <pp>
/b/	, <bb>
/m/	<m>, <mm>
/pf/	<pf>
/f/	<f>, <ph>, <v>, <w>, <ff>, <gh>
/v/	<v>, <w>
/ʋ/	<w>
/θ/	<th>
/ð/	<th>
/t/	<t>, <tt>, <th>, <dt>, <d>
/d/	<d>, <ddh>, <dd>
/n/	<n>, <nn>

/ts/	<z>, <tz>, <t>, <c>
/s/	<s>, <c>, <ç>, <sc>, <sch>, <z>, <ss>, <ß>, <t>
/z/	<z>, <s>, <zz>, <ss>
/tʃ/	<ch>, <tch>, <t>, <tsch>, <c>
/dʒ/	<j>, <g>, <dg>, <gg>, <dj>
/ʃ/	<sj>, <ch>, <sh>, <sch>, <s>, <sk>
/ʒ/	<zj>, <g>, <j>, <ge>
/r/	<r>, <rr>, <rh>
/l/	<l>, <ll>
/ɲ/	<nj>, <gn>
/ç/	<ch>, <g>
/j/	<j>, <i>, <ij>, <y>, <ill>, <i>
/ɥ/	<u>
/k/	<k>, <c>, <ch>, <qu>, <ck>, <g>, <cc>, <cqu>, <q>
/g/	<g>, <gu>, <gg>, <gh>
/ŋ/	<ng>
/x/	<ch>, <g>
/ɣ/	<g>, <gh>
/w/	<oi>, <ou>, <w>
/ʁ/	<r>, <rr>, <rh>
/h/	<h>

Voici les champs que nous avons donnés ci-dessus, complétés ici par les phonèmes qui ont un lien phonographématique avec le phonème concerné, c'est-à-dire qui peuvent être notés par le même graphème (ces derniers sont soulignés) :

Consonne	Consonnes qui partagent ses traits articulatoires sauf un ou deux et celles qui ont un lien phonographématique avec celle
/p/	/b/, /m/, /pf/, /f/
/b/	/p/, /m/, /v/, /v/
/m/	/p/, /b/
/pf/	/p/, /f/, /v/, /v/
/f/	/p/, /pf/, /v/, /v/, /w/, /g/, /y/
/v/	/pf/, /f/, /b/, /v/, /w/
/v/	/b/, /pf/, /f/, /v/, /w/
/θ/	/ð/, /t/, /d/, /n/, /ts/, /s/, /z/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/
/ð/	/θ/, /t/, /d/, /n/, /s/, /z/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/
/t/	/θ/, /ð/, /d/, /n/, /ts/, /s/, /z/, /tʃ/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/
/d/	/θ/, /ð/, /t/, /n/, /ts/, /s/, /z/, /tʃ/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/
/n/	/θ/, /ð/, /t/, /d/, /ts/, /s/, /z/, /tʃ/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/
/ts/	/θ/, /t/, /d/, /n/, /s/, /z/, /tʃ/, /tʃ/, /ʃ/, /r/, /l/, /k/
/s/	/θ/, /ð/, /t/, /d/, /n/, /ts/, /z/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/, /k/
/z/	/θ/, /ð/, /t/, /d/, /n/, /ts/, /s/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/
/tʃ/	/θ/, /t/, /d/, /n/, /ts/, /s/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/, /ç/, /k/, /x/
/tʃ/	/ð/, /t/, /d/, /n/, /ts/, /z/, /tʃ/, /ʃ/, /ʒ/, /r/, /l/, /ç/, /j/, /k/, /g/, /x/, /y/
/ʃ/	/θ/, /ð/, /t/, /d/, /ts/, /s/, /z/, /tʃ/, /tʃ/, /ʒ/, /r/, /l/, /ç/, /k/, /x/
/ʒ/	/θ/, /ð/, /t/, /d/, /s/, /z/, /tʃ/, /tʃ/, /ʃ/, /r/, /l/, /ç/, /j/, /k/, /g/, /x/, /y/
/r/	/θ/, /ð/, /t/, /d/, /n/, /ts/, /s/, /z/, /tʃ/, /tʃ/, /ʃ/, /ʒ/, /l/, /ʁ/
/l/	/θ/, /ð/, /t/, /d/, /n/, /ts/, /s/, /z/, /tʃ/, /tʃ/, /ʃ/, /ʒ/, /r/
/ɲ/	/ç/, /j/, /ɥ/, /k/, /g/, /ŋ/
/ç/	/ɲ/, /j/, /ɥ/, /k/, /x/, /y/, /ʁ/, /ʃ/, /tʃ/, /ʒ/, /tʃ/, /g/

/j/	/ɲ/, /ç/, /ʎ/, /g/, /ɣ/, /ʁ/, /ʒ/, /ʝ/
/ʎ/	/ɲ/, /ç/, /j/, /g/, /ɣ/, /w/, /ʁ/
/k/	/ɲ/, /ç/, /g/, /ŋ/, /x/, /ɣ/, /w/, /ʁ/, /tʃ/, /s/, /ts/, /ʃ/, /ʒ/, /ʝ/
/g/	/ɲ/, /j/, /ʎ/, /k/, /ŋ/, /x/, /ɣ/, /w/, /ʁ/, /ʒ/, /ʝ/, /f/, /ç/
/ŋ/	/ɲ/, /k/, /g/, /x/, /ɣ/, /w/
/x/	/ç/, /k/, /g/, /ŋ/, /ɣ/, /w/, /ʁ/, /ʃ/, /tʃ/, /ʒ/, /ʝ/
/ɣ/	/ç/, /j/, /ʎ/, /k/, /g/, /ŋ/, /x/, /w/, /ʁ/, /ʒ/, /ʝ/, /f/
/w/	/ʎ/, /k/, /g/, /ŋ/, /x/, /ɣ/, /ʁ/, /v/, /ʋ/, /f/
/ʁ/	/ç/, /j/, /ʎ/, /k/, /g/, /x/, /ɣ/, /w/, /r/
/h/	néant

Les champs présentés ici sont établis de manière théorique, ce qui a pour conséquence que certains phonèmes en ont beaucoup d'autres dans leur champ. Ceci tient notamment au fait que /t/, /d/, /n/, /r/ et /l/ sont présentés comme dentaux, alvéolaires ou post-alvéolaires dans le tableau de l'API. Ils entrent donc en commutation possible avec un grand nombre de phonèmes. Même si ces champs reflètent des proximités articulatoires, nous savons par l'expérience d'ICE que les apprenants remarquent vite des régularités et apprennent à remplacer une consonne par quelques autres seulement de son champ, avec lesquelles elle commute préférentiellement. Nous tenons donc à restreindre nos champs pour que l'analyse du corpus, dans la troisième partie, prenne en compte le fait que les différentes commutations possibles au sein d'un champ ne sont pas toutes réalisées. Ce sont en effet les correspondances régulières qui sont mises à profit et qui doivent donc nous servir à définir la proximité formelle indirecte.

Nous allons, pour restreindre les champs donnés ci-dessus, nous appuyer sur les travaux de R. Geysen. Le but de son *Dictionnaire des formes analogues en 7 langues* est de présenter ensemble des mots du latin, de l'italien, de l'espagnol, du français, de l'anglais, du néerlandais et de l'allemand qui ont des formes analogues, pour tirer ensuite des « règles de correspondance qui découlent de ce groupement par analogie de forme » (Geysen 1990 : 7). Pour savoir s'il pouvait comparer des mots qui lui semblaient analogues

du point de vue de la forme, l'auteur s'est fondé sur leur étymologie. Si deux mots pouvaient être rapprochés par leur étymologie, alors il s'autorisait à comparer leur forme actuelle. Le fait que deux mots aient une même origine ou que l'un ait été emprunté à la langue de l'autre ne constitue ainsi qu'une caution *a posteriori* pour justifier un regroupement créé par analogies formelles. Nous nous appuyerons sur les correspondances consonantiques entre le français et les trois langues germaniques. Insistons sur le fait que les correspondances dégagées par l'auteur sont des correspondances en synchronie actuelle et qui ne reproduisent pas les évolutions phonétiques des consonnes envisagées. Ces correspondances sont données sous forme de tableaux, accompagnés de nombreux exemples, du type de celui-ci :

<i>Français</i>	<i>English</i>	<i>Nederlands</i>	<i>Deutsch</i>
p	f	v	v, f
Père	Father	Vader	Vater
Pied	Foot	Voet	Fuss

Les correspondances données sont des correspondances de lettres, que nous avons transformées en correspondances de phonèmes grâce aux exemples.

Les champs que nous voulons établir mettent en jeu des proximités articulatoires (ou phonographématiques) entre phonèmes en synchronie actuelle, pour rendre compte de la proximité telle qu'elle est ressentie par des apprenants qui n'ont éventuellement aucune connaissance en phonétique diachronique. Il ne s'agit donc pas de remplacer les champs présentés ci-dessus par les correspondances observées par R. Geysen, mais, pour restreindre ces champs établis théoriquement, de garder, dans le champ d'une consonne, les seules autres dont la correspondance avec la consonne envisagée est attestée, c'est-à-dire apparaît dans le *Dictionnaire des formes analogues en 7 langues*. De plus, puisque nous tenons à ce que les champs gardent leur justification théorique première, donc que les consonnes d'un champ soient proches d'un point de vue articulatoire, nous n'ajouterons pas les correspondances observées par R. Geysen et qui n'apparaissent pas dans les champs donnés ci-dessus. Ainsi, le facteur étymologique ne supprime en rien notre étude synchronique, mais restreint seulement ses résultats.

Les champs phonologiques définitifs sont les suivants :

Consonne	Consonnes de son champ phonologique
/p/	/b/, /pf/, /f/
/b/	/p/, /v/
/m/	pas de champ
/pf/	/p/
/f/	/p/, /v/, /v/, /ɸ/
/v/	/f/, /b/, /w/
/v/	/f/
/θ/	/t/, /d/
/ð/	/t/, /d/
/t/	/θ/, /ð/, /d/, /ts/, /s/
/d/	/θ/, /ð/, /t/, /ts/
/n/	/r/, /l/
/ts/	/t/, /d/, /s/
/s/	/t/, /ts/, /z/, /tʃ/, /ʃ/, /k/
/z/	/s/, /ʃ/
/tʃ/	/s/, /ʃ/, /ç/, /k/, /x/
/ç/	/d/, /ʒ/, /j/
/ʃ/	/s/, /z/, /tʃ/, /ç/, /k/, /x/
/ʒ/	/ç/, /j/, /k/, /g/
/r/	/n/, /l/
/l/	/n/, /r/
/ɲ/	pas de champ
/ç/	/k/, /x/, /ɣ/, /ʃ/, /tʃ/
/j/	/g/, /ʒ/, /ç/

/ɥ/	/w/
/k/	/ç/, /g/, /x/, /ɣ/, /ʃ/, /s/, /ʒ/, /z/
/g/	/j/, /k/, /x/, /ɣ/, /w/, /z/
/ŋ/	pas de champ
/x/	/ç/, /k/, /g/, /ʃ/, /ʒ/
/ɣ/	/ç/, /k/, /g/, /f/
/w/	/ɥ/, /g/, /v/
/ʁ/	pas de champ
/h/	pas de champ

Maintenant que la notion d'ajustement formel a été définie, nous pouvons caractériser la proximité formelle indirecte. De la même manière que des consonnes graphiques différentes qui notent le même phonème sont considérées comme ne remettant pas en cause la trame consonantique en proximité formelle directe, des phonèmes consonantiques différents appartenant au même champ phonologique sont considérés comme ne remettant pas en cause la trame consonantique en proximité formelle indirecte.

Nous formulons ainsi la règle suivante : *pour que deux mots soient proches formellement indirectement, ils doivent avoir la même trame consonantique, c'est-à-dire les mêmes consonnes apparaissant dans le même ordre, dont au moins une, dans un mot, appartient au champ phonologique de la consonne correspondante dans l'autre mot, et avec éventuellement dans un des mots une consonne de plus ou de moins au contact de ses consonnes initiales ou médianes, et la dernière consonne, si elle est au moins la troisième, éventuellement différente.* En d'autres termes, si le lecteur a à effectuer un ajustement formel sur au moins une des consonnes initiales et/ou médianes, alors les deux mots sont proches selon la proximité formelle indirecte.

Les données graphiques, qui ont été considérées en proximité formelle directe, sont capitalisables en proximité formelle indirecte. Ainsi, si une consonne d'un mot appartient au champ phonologique de la consonne correspondante dans l'autre mot considéré mais que ces deux phonèmes sont notés par le même graphème, alors les deux mots sont proches formellement directement. De même, s'il y a un ajustement formel à pratiquer sur une des

consonnes médianes d'un mot mais que les quatre premières lettres sont identiques entre les deux mots considérés, alors les deux mots sont également proches formellement directement (voir *supra* section 2.2.1.).

Donnons un exemple pour illustrer cette définition. Soit la phrase extraite d'un article du journal *Die Zeit* : « Merkel wird am Mittag in die US-Hauptstadt fliegen, um mit 19 anderen Staats- und Regierungschefs [...] über Konsequenzen aus der Weltfinanzkrise zu beraten »⁶⁴. Le mot qui nous intéresse est *Mittag*. À première vue, il ne ressemble à aucun mot français. Imaginons que, pour essayer de le comprendre, un apprenant pratique un ajustement formel. Nous ne donnons pas toutes les commutations possibles pour toutes les consonnes, car il est très improbable que cet apprenant tente toutes ces commutations avant de choisir celle qui convient. C'est la prise en compte du contexte qui le guide. Il s'agit de la première phrase de l'article, et il peut avoir reconnu que le journaliste relate (quoi ?) un déplacement (de qui ?) d'Angela Merkel (où) aux États-Unis. Puisqu'il lui manque l'indication de temps, il pourra reconnaître sous la forme *Mittag* le mot français « midi ». Entre *Mittag* et « midi », après avoir changé le /t/ en /d/, la trame consonantique est effectivement la même (puisque l'ajout d'une consonne finale en allemand concerne la troisième consonne), les deux mots sont donc proches formellement. Signalons que si le lecteur a des connaissances en anglais ou en néerlandais, il peut également penser à *midday* ou à *middag*.

L'ajustement formel est donc une méthode qui permet de « découvrir » sous une forme étrangère un mot qui est connu, soit parce qu'il appartient au lexique de la langue maternelle, soit parce qu'il appartient au lexique d'une autre langue étrangère connue.

Cette méthode de l'ajustement formel est d'autant plus recevable qu'il s'agit d'appliquer à la comparaison interlinguistique un principe que nous manions déjà tous dans notre langue maternelle. Ainsi, des oppositions de consonnes appartenant au même champ phonologique sont observables sur des paires de mots français entretenant une relation sémantique évidente. Nous proposons cette liste non exhaustive, qui illustre notre propos sur plusieurs champs :

⁶⁴ <http://www.zeit.de/online/2008/47/weltfinanzgipfel-empfehlungen-issing> (consultée le 27/11/08)
Nous traduisons : « Merkel se rendra ce midi dans la capitale des États-Unis pour discuter des conséquences de la crise financière mondiale avec 19 autres chefs d'États et de gouvernements. »

/v/-/b/ : hiver – hiberner
/b/-/p/ : double – dupliquer
/v/-/f/ : navette – nef
/s/-/t/ : force – fortifier
/d/-/t/ : perdition – perte
/ʃ/-/k/ : chair – carnassier
/g/-/ʒ/ : fugue – fugitif

À titre indicatif, signalons que ce type d'équivalences existe aussi dans les trois langues germaniques sur lesquelles nous travaillons :

- anglais :

/f/-/v/ *safe* – *save* (« sauf » – « sauver »)

/d/-/t/ *build* – *built* (« construire » – « construit »)

- allemand :

/f/-/pf/ *schaffen* – *Schöpfer* (« créer » – « créateur »)

/x/-/k/ *wachen* – *wecken* (« veiller » – « réveiller »)

- néerlandais :

/d/-/t/ *snijden* – *snit* (« couper » – « coupe »)

/f/-/v/ *graf* – *groeve* (« tombe » – « fosse »)

Nous ne pouvons certes pas affirmer que les apprenants ont conscience de ces manipulations dans leur langue, et que c'est ainsi qu'ils apprennent à pratiquer l'ajustement formel facilement. Il semble néanmoins que les locuteurs aient une conscience plus ou moins forte de la proximité entre phonèmes. C'est ce que nous voulons montrer en donnant trois illustrations d'ordres différents de cette idée.

G. Lukatela *et al.* ont réalisé des expériences qu'ils retracent dans leur article « Does visual word identification involve a sub-phonemic level? » publié en 2000. Leur hypothèse est que le code phonologique qui est assemblé lors de la lecture de mots écrits contient les traits qui composent les phonèmes particuliers. Lors de la première expérience, ils ont présenté à 72 sujets 24 mots cibles monosyllabiques, qui étaient précédés soit d'un non-mot dont la consonne initiale ne différait de celle du mot cible que par un seul trait (en l'occurrence le trait voisé/non voisé), soit par un non-mot dont la consonne initiale différait

de celle du mot cible par deux traits ou plus. Il s'agissait d'une tâche de décision lexicale. S'ils ont choisi de ne faire varier que le trait de voisement, c'est parce que la contribution des traits inférieurs aux phonèmes leur paraissait délicate à détecter :

*It seemed prudent, therefore, to identify and exploit that single-feature difference between the non-word test prime and its word target that rendered the two most similar. A difference in voicing seemed to satisfy the preceding requirement given that the acoustic spectra of two spoken rhyming words are most alike when the feature contrast between their initial phonemes is simply one of voicing.*⁶⁵ (Lukatela et al. 2000 : 43)

Les résultats de cette expérience montrent qu'un mot est plus vite identifié s'il est précédé d'un non-mot dont la première consonne ne diffère de celle du mot cible que par un seul trait (par exemple, *sea* est plus vite identifié s'il est précédé de *zea* que de *vea*, *did* est plus vite identifié s'il est précédé de *tid* que de *yid*, etc.). Il semblerait donc que les locuteurs aient une certaine conscience de la proximité entre phonèmes.

C. Kerbrat-Orecchioni consacre une partie de son ouvrage *La connotation* au symbolisme phonétique. Elle y fait le point sur un certain nombre d'expériences qui tendent à prouver que les locuteurs associent des caractéristiques à certains sons en fonction de leur articulation. Elle rappelle notamment les constatations suivantes :

Dans toutes les langues, les phonéticiens (pourtant en principe hommes de science) désignent les sons à l'aide de métaphores analogues : ils parlent de voyelles claires ou sombres, de consonnes mouillées ou liquides. Pourquoi ? (Kerbrat-Orecchioni 1984 : p. 29)

Des tests effectués sur des enfants de divers âges, de divers pays, et utilisant divers systèmes d'écriture, ont obtenu des résultats concordants : pour la grande majorité

⁶⁵ Nous traduisons : « Par conséquent, il nous a semblé prudent d'identifier et d'exploiter le trait différentiel qui rendait l'amorce, constituée par le non-mot, et le mot cible le plus similaires possible. Une différence de voisement semblait satisfaire à la condition précédente, étant donné que les spectres acoustiques de deux mots rimant l'un avec l'autre sont les plus proches lorsque la différence de trait entre leurs phonèmes initiaux n'est qu'une différence de voisement. »

des sujets, le [i] est perçu comme petit, clair et gentil ; le [u] est sombre et méchant ; le [a] gros et gras ; le [t] et le [k] sont durs et méchants, etc. (ibid. : 30)

En 1929, Sapir présente à des sujets anglais soixante paires de logatomes du type « mil/mal », en leur demandant lequel de ces deux pseudo-mots correspond à l'objet le plus grand. Cette expérience met en évidence le fait que les sujets établissent une proportionnalité parfaite entre le degré d'ouverture de la voyelle et la taille de l'objet correspondant. (ibid. : 30)

Pour expliquer ces faits, elle indique pour commencer que des raisons d'ordre acoustique ne sont pas à exclure :

[...] certains sons sont particulièrement propres à symboliser certaines réalités dans la mesure où ils produisent un effet acoustique analogue (les bilabiales fricatives imitent le souffle du vent) ; le [l] reproduit le bruit d'un liquide en mouvement, etc.). Mais cette explication n'est acceptable que dans les cas, très limités, où le dénoté est de nature sonore. (ibid. : 31)

Elle préfère néanmoins justifier les résultats de ces expériences par des raisons d'ordre articulaire :

L'explication articulaire semble beaucoup plus générale, et rend compte du fait que les enfants sourds démutés ont devant les tests qui viennent d'être décrits les mêmes réactions exactement que les autres. [...] l'articulation de tel ou tel son possède certaines propriétés visuelles, que l'on perçoit (inconsciemment) en se plaçant en quelque sorte à l'intérieur de la cavité buccale produisant le son, et que l'on projette métonymiquement, et par synesthésie, sur le son lui-même. Par exemple, lors de l'émission d'un [i], les organes phonatoires sont très resserrés, et la colonne d'air très mince, d'où l'impression que le [i] lui-même, et par un second glissement métonymique, que les objets correspondants (si toutefois le sémantisme du mot dans lequel figure la voyelle ne vient pas contrecarrer cette impression) sont petits ; au contraire, le [a] est « gros et gras » parce que la bouche est grande ouverte au moment de son articulation ; le [u] est obscur car la cavité buccale qui l'articule se présente comme un vaste espace clos ; les nasales sont « voilées » car

en passant par les fosses nasales le son s'estompe et perd de sa clarté ; les consonnes « mouillées » exigent pour leur articulation, cela a été prouvé expérimentalement, une abondante émission de salive, et les [t] et [k] une grande tension articulaire. (ibid. : 31-32)

Les locuteurs auraient ainsi conscience, à quelque degré que ce soit, de la façon dont les sons sont articulés. Il est donc tout à fait concevable qu'ils aient le même type de conscience de la proximité articulaire entre sons.

Cette relation de proximité qui lie entre eux certains phonèmes a déjà été observée et mise à profit. Ainsi, dans le système Pitman, qui est le système sténographique le plus répandu dans les pays de langue anglaise, les signes qui notent les phonèmes /p/ et /b/ sont très proches (il s'agit respectivement de \ et de \), ainsi que ceux qui notent /f/ et /v/, /t/ et /d/, /θ/ et /ð/, /s/ et /z/, /ʃ/ et /ʒ/, /k/ et /g/. Dans la méthode française Prévost-Delaunay, /s/ et /z/ sont notés par un signe cette fois-ci identique, ainsi que /f/ et /v/ d'une part et /g/ et /ʒ/ de l'autre.

Que la proximité entre phonèmes soit utilisée en sténographie, qui plus est dans des méthodes qui ont fait leurs preuves et dont on sait qu'elles permettent une relecture réussie, tend à montrer qu'il est possible de passer, de façon réglée, de certains sons, consonantiques en l'occurrence, à d'autres, sans perte d'efficacité dans la transmission du message. De la même façon, nous proposons avec l'ajustement formel de tirer parti de la proximité entre phonèmes.

2.3.2. La proximité sémantique indirecte : l'ajustement sémantique

Nous avons été amenée à préciser et à étendre la notion de proximité formelle pour rendre compte des procédures mises en œuvre par les apprenants pour accéder à des formes connues, ou à tout le moins des formes proches de formes connues. De même, nous allons ici étendre la notion de proximité sémantique et proposer une méthode que nous appelons l'« ajustement sémantique »⁶⁶, qui permet de rendre compte des procédures mises en œuvre

⁶⁶ Voir aussi : Castagne 2007, Castagne *et al.* 2007 et Caure *et al.* 2008.

par les apprenants pour accéder, du point de vue du sens cette fois-ci, à l'équivalent français lorsqu'ils essaient de comprendre un mot étranger. Par « mot français équivalent », nous entendons un mot qui constituerait, lors de la traduction d'un texte, et dans le cas où l'on choisit de traduire un mot par un autre, une traduction satisfaisante du mot étranger de départ.

Le parallèle entre la proximité formelle indirecte et la proximité sémantique indirecte peut être, dans un premier temps, formulé ainsi : de la même manière qu'une forme qui n'est pas exactement similaire à la forme d'un mot français permet tout de même à un lecteur, par le biais de quelques manipulations, de penser à ce mot, un mot qui n'a pas exactement la même signification que l'équivalent permet tout de même à un lecteur, par le biais de quelques manipulations, de penser à cet équivalent.

Avant d'entrer plus avant dans la présentation de l'ajustement sémantique et d'étendre le parallèle à d'autres notions, il nous faut signaler une différence fondamentale qui existe entre l'ajustement sémantique et l'ajustement formel. Cette différence entre les méthodes tient à la différence entre la nature des formes et celle des significations que nous avons développée dans la section 2.1.3.

L'ajustement formel se pratique, comme son nom l'indique, sur une forme, et donc en dehors de toute langue. Il s'agit en effet de remplacer des phonèmes par d'autres, et nous savons que les phonèmes sont dénués de signification – ils sont les unités non significatives de deuxième articulation dont le regroupement forme les unités significatives de première articulation, les morphèmes. Puisque l'ajustement formel se pratique sur un mot qui n'a pas pu être compris au premier abord, les manipulations ont lieu sur une simple forme, qui ne fait pas signe. En d'autres termes, cet ajustement se fait hors langue, sur une forme, et non sur un signifiant, qui, lui, est nécessairement lié à un signifié.

L'ajustement sémantique ne peut, quant à lui, se pratiquer qu'en langue. Nous avons déjà montré que l'analyse d'une signification ne peut être menée qu'en tant que signification d'un mot, donc au sein d'un système linguistique. Les manipulations que nous allons présenter et qui constituent l'ajustement sémantique ne peuvent donc être réalisées que sur un mot d'une langue précise. Puisque nous sommes dans le cadre de l'accès à des langues par des apprenants, la langue dont il est question ici et sur laquelle les apprenants peuvent réfléchir est leur langue maternelle (ou éventuellement une autre langue étrangère qu'ils maîtrisent). F.-J. Meissner *et al.* insistent sur le fait que les apprenants se

construisent de nouveaux systèmes de langue en appui sur le ou les systèmes qu'ils possèdent déjà :

[...] *l'apprenant d'une deuxième, troisième ou quatrième langue possède déjà un ou plusieurs systèmes : lexiques, grammaires, concepts, un savoir encyclopédique relié à ces langues - à partir desquels il émet ses hypothèses.* (Meissner et al. 2004 : 23)

La prise en compte de ces deux considérations – un lecteur n'a accès à une signification que par le biais d'un signifiant et il ne peut réfléchir qu'aux significations des mots de sa langue – nous permet de préciser ce que nous avons dit plus haut : l'ajustement sémantique consiste à jouer sur le sens du mot français identifié par sa ressemblance formelle avec le mot étranger jusqu'à aboutir à l'équivalent. Dans les faits, le processus est le suivant : le lecteur identifie une forme, c'est-à-dire reconnaît une forme comme étant celle d'un mot français ou comme étant proche de celle d'un mot français, ses capacités sémiologiques font correspondre un signifié à ce signifiant, et, si le mot français n'est pas adéquat à cet endroit du texte lu⁶⁷, alors le lecteur joue sur le sens du mot pour parvenir à un autre mot, qui, lui, convient.

Ce que nous décrivons, c'est-à-dire le cas où un mot a une forme similaire et une signification différente par rapport à un autre mot, nous amène à recharacteriser le processus de compréhension des faux amis.

En effet, lorsqu'il est question d'un faux ami, il est dit que c'est le sens du mot étranger qui diffère de celui du mot qui en est proche formellement en langue maternelle. Rappelons la définition linguistique que donne M. Ballard du faux ami : « Les faux amis sont des signes de deux langues différentes dont les signifiants sont en relation d'homonymie ou de paronymie et dont les signifiés diffèrent plus ou moins » (Ballard 1999 : 9). Or, comme nous l'avons dit, le lecteur peut être totalement débutant, et c'est en français qu'il réfléchit. Le sens du mot étranger est pour lui inaccessible. Même lorsqu'il a affaire à un mot transparent directement, et qu'il le comprend, c'est en réalité parce que le mot français qu'il reconnaît derrière le mot étranger (grâce à leur forme similaire) convient dans le contexte en question. H. Besse, dans son ouvrage *Méthodes et pratiques des*

⁶⁷ Nous développerons p. 168 les raisons qui font qu'un mot n'est pas adéquat dans un contexte.

manuels de langue, formule la même idée, pour une situation légèrement différente : il précise que seul un enseignant en langue (le « maître », dans son texte) peut donner une traduction d'un mot étranger, et non les apprenants (les « étudiants »), car connaître la signification d'un mot implique de connaître le système de la langue considérée. Le signifiant d'un mot étranger est donc mis en relation par les apprenants avec le signifié d'un mot de leur langue maternelle :

[S]eul peut véritablement traduire celui qui possède les deux langues entre lesquelles s'opère la traduction ; les étudiants n'en connaissant qu'une ne peuvent traduire, au sens précis de ce mot ; ils saisissent seulement que tel signe de leur L1 s'étiquette par tel signifiant de la L2, sans pouvoir restituer à ce signifiant le signifié qui est le sien à l'intérieur de la L2, puisque celui-ci est délimité et constitué essentiellement par des oppositions et des relations dont ils n'ont pas encore l'expérience : le signifiant étranger fourni par le maître les renvoie donc à un sens familier et à leur L1. (Besse 1985 : 21-22)

En fait, lors de la lecture d'un faux ami, le processus de compréhension du sens se passe en langue maternelle. Le lecteur ne peut donc pas juger si les « signes de langues différentes » ont des « signifiés qui diffèrent plus ou moins ». Les signes qui ont des « signifiés qui diffèrent plus ou moins » sont des signes de même langue. Le phénomène des faux amis est donc, du point de vue du sens, un phénomène intralinguistique.

En outre, notre approche se sépare du processus de compréhension des faux amis tel qu'il est décrit habituellement car un faux ami a une forme qui est « en relation d'homonymie ou de paronymie » avec un mot de la langue du lecteur. Nous avons, précédemment, défini l'homonymie et la paronymie comme relevant de la proximité formelle directe (voir section 2.2.1.). Ici, nous prenons également en considération la proximité formelle indirecte, puisqu'un mot proche formellement indirectement peut également être identifié par un apprenant (voir section 2.3.1.).

Trois mots sont impliqués lors de l'ajustement sémantique : 1/ le mot étranger, 2/ le mot français qui est suggéré par sa ressemblance formelle avec le mot étranger, et à partir duquel on pratique l'ajustement sémantique, et 3/ un autre mot français, auquel on aboutit après avoir pratiqué l'ajustement sémantique, et qui est l'équivalent du mot étranger.

Le mot étranger 1/ est proche formellement, directement ou indirectement, du mot français 2/ (c'est-à-dire qu'ils ont la même trame consonantique ou qu'ils commencent par les quatre mêmes premières lettres) ; le mot étranger 1/ est proche sémantiquement directement du mot français 3/, (c'est-à-dire que ce dernier constitue l'équivalent du mot étranger 1/ si on le traduit) ; et nous appelons « proximité sémantique indirecte » la relation qui unit le mot français 2/ et le mot français 3/. La proximité sémantique directe unit donc un mot étranger et un mot français, son équivalent, et la proximité sémantique indirecte unit deux mots français, dont les sens sont plus ou moins proches. Cette relation est indirecte puisqu'on pratique un ajustement sémantique pour passer d'un sens à l'autre.

Nous pouvons ainsi rendre compte des relations qui unissent les trois mots impliqués lorsqu'il y a ajustement sémantique par le schéma suivant, dans lequel les mots sont reliés deux à deux par leur proximité formelle ou significative :



gras : proximité formelle (directe ou indirecte)

italique : proximité sémantique directe

souligné : proximité sémantique indirecte

Le mot étranger 1/ apparaît en gras et en italique. Il est donc impliqué dans deux types de relations : il possède une ressemblance formelle avec le mot français 2/ et un sens directement proche du mot français 3/.

Le mot français 2/ apparaît en gras et souligné. Il est donc impliqué dans deux types de relations : il possède une ressemblance formelle avec le mot étranger 1/ et un sens indirectement proche du mot français 3/.

Le mot français 3/ apparaît en italique et souligné. Il est donc impliqué dans deux types de relations : il possède un sens directement proche du mot étranger 1/ et un sens indirectement proche du mot français 2/.

Notons qu'il n'est pas question de relation sémantique entre le mot étranger 1/ et le mot français 2/ (qui sont proches du point de vue de la forme), ni de relation formelle entre

le mot français 2/ et le mot français 3/ (qui sont proches, indirectement, du point de vue du sens).

Pour présenter précisément quelle est la relation qui unit le mot français 2/ et le mot français 3/, donc pour présenter la proximité sémantique indirecte, il nous faut développer ce qui est entendu lorsqu'il est question de deux mots de même langue qui ont des significations proches l'une de l'autre.

Les locuteurs d'une langue sont tous capables de dire, et ont probablement tous déjà dit, que tel mot « veut dire presque la même chose » que tel autre, ou qu'« il y a un rapport » entre deux mots. Ainsi, que deux mots soient proches du point de vue de leur signification est un fait intuitivement perceptible par chacun. Des expériences ont été menées dans le but de confirmer cette intuition, et de découvrir des constantes entre individus :

On demande alors de répondre à une question telle que : « Indiquez quel degré de parenté vous ressentez entre les significations des mots M et N. » On peut utiliser aussi bien les mots ordinaires de « ressemblance », de « distance » ou de « proximité » [...]. On fournit une échelle, de 1 à 7 ou de 1 à 9, avec une indication de correspondances entre nombres et degrés. Les participants répondent sans difficulté à de telles questions, et ils sont capables d'exprimer par une note leur sentiment concernant ces proximités. Il existe un raisonnable degré d'accord entre les participants sur ces estimations, non toujours en valeur absolue, mais sur leur ordre. Ainsi jugera-t-on que « docteur » est sémantiquement très proche de « médecin » (il en est même synonyme dans certains contextes), que « chirurgien » en est modérément proche, qu'« infirmier » en est un peu plus éloigné, qu'« hôpital » ou « ambulance » le sont encore un peu davantage, tandis qu'« autobus » et beaucoup d'autres mots en sont très éloignés. [...] Nous sommes là dans le champ très général de l'« analogie » sémantique, mais la stabilité statistique de ces estimations montre que nous avons tous une connaissance implicite des proximités sémantiques dans notre lexique mental. (Le Ny 2005 : 214)

Cette citation de J.-F. Le Ny rend compte d'expériences qui montrent que les locuteurs peuvent attribuer une gradualité à la proximité sémantique qui existe entre des mots qui leur sont présentés. D'autres expériences ont eu pour but de faire produire des

« associations lexicales » à des participants. Il apparaît alors que des mots peuvent être proches selon différents types d'associations sémantiques :

Certaines [associations] se font d'un concept vers un concept superordonné (« cheval » → « animal »), ou d'un superordonné à un infra-ordonné, ou hyponyme (« animal » → « cheval »), ou d'un concept à un autre de même niveau dans un arbre conceptuel (« cerise » → « fraise ») : on parle de « co-hyponymie » à ce propos. Mais on trouve communément aussi des associations qui vont d'un concept de catégorie d'objets à un concept de propriété de ces objets (« rhinocéros » → « corne »), ou qui vont dans le sens inverse (« corne » → « vache »), d'autres qui vont d'un concept de catégorie d'objets à un concept d'une autre catégorie d'objets qui coexistent avec les premiers dans une situation (« table » → « chaise » ou « cuiller » → « fourchette »), ou d'un concept de catégorie à celui d'une catégorie sémantiquement voisine, toujours à l'intérieur d'une situation (« infirmier » → « hôpital » ou, pour des noms abstraits, « maladie » → « santé »). Dans le cas de noms abstraits désignant une propriété ou une qualité, on trouve souvent des associations qui vont d'un concept vers un concept opposé (« blanc » → « noir »), ou simplement contrasté (« blanc » → « gris »). (ibid. : 225-226)

La psycholinguistique peut constater des rapports mentaux sémantiques entre mots. Pour les expliquer, la linguistique a créé la notion de « champ sémantique ». Sa définition est cependant tellement tributaire de l'auteur qui utilise cette appellation, que V. Nyckees la juge « excessivement polysémique » (Nyckees 1998 : 187).

Pour A. Niklas-Salminen, le champ sémantique est l'union d'un champ lexical et d'un champ notionnel :

On peut définir les champs sémantiques comme l'association d'un ensemble de termes du lexique (champ lexical) à une notion particulière (champ notionnel). [...] Les champs sémantiques ne comprennent que des mots appartenant à la même partie du discours : c'est ainsi que le champ sémantique des sièges, par exemple, ne regroupe que des substantifs. Le champ sémantique doit donc être opposé aux champs dérivationnels. [...]

Les champs sémantiques s'opposent aussi aux champs associatifs qui regroupent tous les mots gravitant autour d'une notion donnée [...] (Niklas-Salminen 1997 : 129)

F. Neveu, quant à lui, indique dans l'article « Champ sémantique » de son *Dictionnaire des sciences du langage* que le champ notionnel et le champ dérivationnel sont tous deux des types de champs sémantiques :

Les champs onomasiologiques, conçus selon une approche qui part du concept pour atteindre le signe linguistique qui lui correspond, forment des regroupements lexicaux sur la base de l'univers référentiel auquel renvoient les unités. Il s'agit donc de champs notionnels, qui marquent un domaine d'expérience, auxquels on fait correspondre un ensemble structuré de mots. On parle fréquemment de champ lexical lorsque les lexèmes qui constituent le champ appartiennent à une même classe grammaticale (substantifs, adjectifs, verbes, etc.), et de champ associatif lorsque le champ s'ouvre à l'hétérogénéité catégorielle.

Les champs sémasiologiques, conçus selon une approche qui part du signe pour accéder au concept, se construisent sur des critères linguistiques. Il s'agit de procéder à l'étude sémantique du lexique à partir de similarités morphologiques ou syntaxiques observées entre les unités. [...] Ainsi peuvent être dégagées, en diachronie, des familles de mots formés à partir d'un étymon, ou bien, en synchronie, des champs dérivationnels, faisant apparaître notamment des phénomènes d'affixation à partir d'un même lexème. (ex. courage, courageusement, encouragement, décourager, etc.). (Neveu 2004 : 62-63)

V. Nyckees précise deux autres significations possibles du champ sémantique, qui serait soit l'aire de significations d'un même mot, soit un regroupement de mots différents qui ont des traits sémantiques en commun (c'est d'ailleurs de cette dernière définition dont nous nous rapprochons le plus, comme nous le développerons plus loin) :

Nous en venons à présent aux relations entre les différents sens d'un même mot. La polysémie (opposée à la monosémie) et l'homonymie concernent en effet ce que l'on peut appeler l'aire sémantique d'un mot, c'est-à-dire l'ensemble des significations dont il est susceptible. On évitera en pareil cas d'utiliser l'expression

de champ sémantique qui peut aussi désigner un ensemble de mots distincts partageant certains traits sémantiques communs. (Nyckees 1998 : 193)

Un des premiers auteurs (avec G. Ibsen, W. Porzig et A. Jolles notamment) à réfléchir à la notion de champ, et dont la réflexion a été prédominante, fut J. Trier. Dans son article « Das sprachliche Feld. Eine Auseinandersetzung » datant de 1934, ce chercheur allemand définit le « champ linguistique » ainsi :

Felder sind die zwischen den Einzelworten und dem Wortschatzganzen lebendigen sprachlichen Wirklichkeiten, die als Teilganze mit dem Wort das Merkmal gemeinsam haben, daß sie sich ergliedern, mit dem Wortschatz hingegen, daß sie sich ausgliedern.⁶⁸ (Trier 1934 : 430)

Il insiste plus loin sur le fait que les champs tels qu'il les conçoit ne sont pas des ensembles fixes :

*[...] die Außengrenzen des Feldes [sind] offenbar recht ungewiß [...], die Zahl der Bestandteile [kann] unordentlicherweise zu- und abnehmen [...], und die Binnengrenzen, weit davon entfernt, als klare mathematische Grenzkonturen sich zu erweisen, [stellen] in Wahrheit vielmehr Überschneidungszonen und schwankende Übergangssäume [dar]*⁶⁹ (ibid. : 447)

Les champs seraient donc des ensembles de mots en relation les uns avec les autres, combinables entre eux pour constituer de nouveaux champs, plus importants, et ainsi de suite jusqu'à structurer le vocabulaire entier d'une langue. Cependant, J. Trier souligne lui-même la difficulté qu'il y a à circonscrire les champs, et ne précise pas de quel type est la relation qui unit les mots entre eux à l'intérieur d'un champ. C'est un problème soulevé par B. Pottier dans son article « Vers une sémantique moderne » :

⁶⁸ Nous traduisons : « Les champs sont les réalités linguistiques vivantes qui se situent entre les mots simples et l'ensemble du vocabulaire, et qui, en tant que totalités parcellaires, partagent, avec le mot, la caractéristique de s'assembler entre elles, et, avec le vocabulaire au contraire, celle de se subdiviser. ».

⁶⁹ Nous traduisons : « Les frontières extérieures du champ sont de toute évidence bien incertaines, le nombre des composants peut, de façon désordonnée, augmenter et diminuer, et les frontières intérieures, loin de s'avérer être des contours mathématiques clairs, constituent en vérité plutôt des zones de recoupement et des bordures de transition fluctuantes ».

Il ne peut exister de « champ » que dans la mesure où une limitation peut lui être assignée. Or c'est là un problème qui n'est généralement pas traité. Il suppose le phénomène d'association qui, s'il n'est pas défini linguistiquement, relève du domaine de la psychologie, où tout devient possible. (Pottier 1964 : 110)

J. Trier, influencé par Saussure, met toutefois l'accent sur le fait que les mots n'ont pas une signification qui leur est propre, mais une signification qui dépend de celle des autres mots du champ :

[ein] sprachliche[s] Feld, das sich ja eben nicht aus der Bedeutung eines einzelnen [...] Wortes nährt, sondern in dem das Bedeuten des einzelnen Wortes umgekehrt aus der Gliederung des Ganzen heraus lebt.⁷⁰ (Trier 1934 : 442)

L'idée selon laquelle un mot se définit par ses rapports avec les autres mots du système a déjà été développée (voir p. 120), et reprend la notion de « valeur » introduite par Saussure. Ce dernier indique qu'un mot doit être comparé « avec les valeurs similaires, avec les autres mots qui lui sont opposables » et que ce sont « les mots qui expriment des idées voisines [qui] se limitent réciproquement » (Saussure 1995 : 160). C'est ici la notion de « champ » qui est annoncée en filigrane. Une application des principes saussuriens à la théorie de J. Trier permet ainsi d'affiner la notion de « champ sémantique » : il s'agirait d'un ensemble de mots opposables entre eux et se définissant les uns par rapport aux autres, ensemble qui serait à son tour susceptible de se combiner avec d'autres ensembles pour en constituer un autre de niveau supérieur.

Ainsi, c'est la sémantique structurale qui va nous permettre de préciser les relations qui unissent les mots qui ont des significations proches les unes des autres, c'est-à-dire les mots qui appartiennent au même champ. La théorie des sèmes s'avère être la théorie idoine pour décrire les relations entre mots au sein de champs, comme le confirme cette citation de B. Pottier : « Le sème est le trait distinctif sémantique d'un sémème, relativement à un petit ensemble de termes réellement disponibles et vraisemblablement utilisables chez le locuteur dans une circonstance donnée de communication » (Pottier⁷¹, cité par Rastier 1987 : 33). Nous allons revenir sur les notions de « sème » et de « sémème », mais l'on

⁷⁰ Nous traduisons : « un champ linguistique qui ne se nourrit justement pas de la signification d'un mot unique, mais dans lequel la signification d'un mot unique vit au contraire de la structure du tout. »

⁷¹ B. Pottier, Sémantique et noémique, *Anuario de Estudios filológicos*, Universidad de Extremadura, Cáceres, 1980, p.169

reconnaît déjà dans cette définition les idées de valeur (« distinctif ») et de champ (« ensemble de termes »).

Un exposé des principes de la sémantique structurale va nous permettre de définir les relations précises qui unissent les mots qui ont des significations proches les unes des autres, et ensuite de donner notre définition du « champ sémantique », qui s'éloigne légèrement du « petit ensemble de termes réellement disponibles et vraisemblablement utilisables chez le locuteur dans une circonstance donnée de communication » de B. Pottier en ce qu'il est constitué d'un nombre plus important de mots.

La sémantique structurale est une théorie dont les outils permettent d'étudier le sens d'unités de différents niveaux, du morphème au texte. Un de ses plus importants représentants actuels est F. Rastier. Cette théorie a été développée dans les années 1960, grâce à la méthodologie mise au point par les phonologues structuralistes de l'École de Prague, par des chercheurs tels que A. J. Greimas ou B. Pottier :

Considérant que le plan de l'expression d'une langue est constitué d'écarts différentiels et qu'à ces écarts du signifiant doivent correspondre des écarts du signifié (interprétables comme des traits distinctifs de la signification), cette nouvelle approche trouve là un moyen d'analyser les unités lexicales manifestes (morphèmes ou assimilées) en les décomposant en ces unités sous-jacentes, plus petites (dites parfois minimales), que sont les traits sémantiques ou sèmes. (Greimas & Courtés 1979 : 326)

Tout comme, sur le plan du signifiant, l'analyse peut descendre sous le niveau du phonème et ainsi faire apparaître les traits articulatoires distinctifs, sur le plan du signifié, l'analyse peut descendre sous le niveau du morphème (ou du mot, ou de la lexie) et ainsi faire apparaître les traits sémantiques distinctifs, ou sèmes. Ces deux types de traits, articulatoires et sémantiques, n'ont pas d'existence propre, mais sont des outils d'analyse.

Le signifié d'un morphème est appelé « sémème ». Un sémème est composé de « sèmes », définis par opposition de sémèmes et qui n'ont ainsi qu'une valeur différenciatrice. Rappelons à ce propos cette citation des *Écrits de linguistique générale* : « il est évident que [l]e sens repose sur le pur fait négatif de l'opposition des valeurs, vu que le temps matériellement nécessaire pour connaître la valeur positive des signes nous aurait cent fois et mille fois manqué » (Saussure 2002 : 77). Ainsi, d'après la très fameuse

analyse de B. Pottier, le sème /avec bras/ n'est présent dans le sémème de « fauteuil » que parce qu'il permet d'opposer « fauteuil » à « chaise ». On ne trouvera pas dans son sémème tous les éléments de signification qu'on peut associer à « fauteuil » et qui ne l'opposeraient à aucun autre mot. Cette description est valable hors texte, lorsqu'il s'agit d'opposer des *significations*. Le sémème d'un mot y est alors composé de « sèmes inhérents », c'est-à-dire de sèmes « définitoires du type. Ils sont hérités par défaut du type dans l'occurrence, si le contexte n'y contredit pas » (Rastier 2005). Lorsqu'un mot est en texte, son sémème est également composé de sèmes « afférents », qui sont soit « socialement normés », c'est-à-dire qui « sont associés au type sans avoir de caractère définitoire » et qui « ne sont actualisés dans l'occurrence qu'en raison de prescriptions issues de son contexte », soit « contextuels », c'est-à-dire qui « résulte[nt] uniquement de propagations de sèmes en contexte » (*ibid.*). Nous aurons ultérieurement l'occasion de développer l'analyse des sémèmes en texte.

Le contenu d'un mot se définit donc par une liste finie de traits sémantiques. Cette conception de la signification permet de préciser ce que nous entendons par « proximité sémantique indirecte » : *deux mots d'une même langue seront considérés comme proches sémantiquement si leurs sémèmes possèdent au moins un sème en commun.*

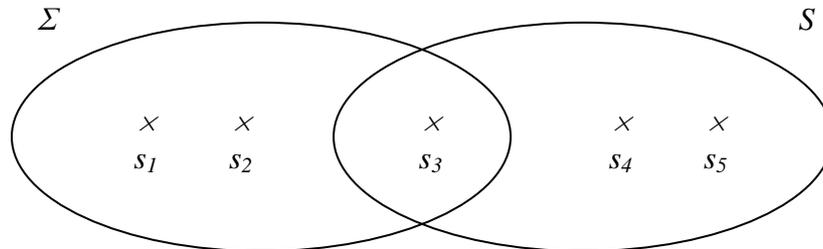
Par exemple, « amanite phalloïde » et « ciguë » sont proches sémantiquement par le biais du sème /vénéneux/. En outre, plus des mots auront de sèmes en commun, plus ils seront proches sémantiquement. On entrevoit là la possibilité d'établir un « gradient de proximité » rendant compte des différents degrés de proximité sémantique entre mots. À la lumière de ce que nous venons de dire, nous pouvons définir le champ sémantique d'un mot comme tous les autres mots dont le sémème possède au moins un sème en commun avec le sémème du mot étudié.

Cette analyse se retrouve sous la plume de J. Tournier, dans son ouvrage *Précis de lexicologie anglaise*, lorsqu'il décrit le mécanisme de la métaphore. Ce qu'il écrit convient parfaitement à notre propos :

On peut se représenter plus facilement le mécanisme général de la métaphore en combinant l'analyse sémique et la représentation ensembliste. Les sèmes sont les traits distinctifs auxquels aboutit l'analyse d'un signifié. Par exemple, le signifié [...] //mouton// comporte, entre autres, les sèmes /quadrupède/, /en forme de flocon/, /blanc/ (les moutons noirs sont beaucoup plus rares), etc. Le signifié

//nuage// comporte, entre autres, les sèmes /eau (en suspension)/, /en forme de flocon/, /blanc/ (les nuages très sombres sont plus rares), etc. [...]

Dans la métaphore, si l'on représente les sèmes de chaque signifié comme les éléments d'un ensemble, on a, au minimum, le diagramme suivant :



L'ensemble Σ (correspondant à un signifié) comporte les éléments (ou sèmes) s_1 , s_2 , s_3 et l'ensemble S (correspondant à un autre signifié) comporte les éléments (ou sèmes) s_3 , s_4 , s_5 . On constate que les deux ensembles ont au moins un sème commun, s_3 . Dès lors, on peut utiliser le signifiant correspondant à S pour désigner Σ . Métaphore veut dire « transfert » et ce transfert (de signifiant) est possible dès que deux signifiés ont un sème commun. [...] Nous appellerons **sème de transfert** tout sème commun qui permet le transfert métaphorique (s_3 dans le schéma ci-dessus). (Tournier 2004 : 137-138)

Nous appellerons après J. Tournier « sème de transfert » le sème qui fonde la proximité sémantique entre deux mots de même langue, c'est-à-dire le sème qui est commun à deux sémèmes.

Cette théorie du sème en commun, ou des sèmes en commun, est également exposée par J.-F. Le Ny pour rendre compte du processus de l'« amorçage sémantique ». Il en donne une première approche au début de son ouvrage *Comment l'esprit produit du sens*, lorsqu'il décrit le phénomène général de l'amorçage :

On présente successivement aux participants, sur un écran d'ordinateur, deux éléments-stimulus, par exemple une suite de deux mots (ou non-mots). On appelle « non-mot » une suite de lettres qui ressemble (on peut préciser à quel degré et de quelle façon) à un mot, mais qui n'en est pas un : par exemple « boutal ». Le participant doit dire le plus vite possible, en appuyant sur un bouton, si le second

élément est ou non un mot. On mesure son temps de réponse : c'est le « temps de décision lexicale ». L'« amorçage », comme phénomène de base, est le fait que la longueur de ce temps de décision dépend, toutes choses étant égales par ailleurs, de la nature du premier élément présenté, ou plus précisément de la relation qui existe entre ce premier élément et le second. Les choses sont particulièrement instructives quand cette relation est sémantique. Dans l'exemple traditionnel, le second élément est le mot « médecin ». On constate que, si le premier élément est « infirmier », ou « hôpital », le temps de décision sur « médecin » est plus court que si le premier élément est « confiseur » ou « carrefour », ou un non-mot. (Le Ny 2005 : 54)

F. Cordier précise, dans son article « De l'intérêt d'une technique d'amorçage sémantique dans l'étude des relations lexicales et conceptuelles » :

Il est un point sur lequel il y a consensus à l'heure actuelle. Il porte sur les processus stratégiques. Si l'amorce et la cible entretiennent un certain type de relation sémantique, si cette relation est attendue par le sujet, on s'attend à observer une facilitation. Dans le cas contraire, on s'attend à enregistrer une inhibition. (Cordier 1996 : 16)

J.-F. Le Ny affirme, dans une des dernières parties de son ouvrage, consacrée aux traits sémantiques, que la similarité sémantique peut s'expliquer par une théorie ensembliste :

Deux concepts A et B, repérés comme sémantiquement similaires, sont [...] analysables en 1. un noyau déterminé de traits communs, auquel s'ajoutent de part et d'autre 2. des traits spécifiques de A, et 3. des traits spécifiques de B. Si on raisonne en termes d'ensembles de traits, on dira que ces deux ensembles comportent une intersection. (Le Ny 2005 : 339)

Peu après, dans la sous-partie intitulée « Une nouvelle analyse de l'amorçage sémantique : la similarité sémantique conçue comme intersection d'ensembles de traits », il précise ce qui avait été annoncé au début de l'ouvrage :

Revenons encore une fois au résultat bien connu de nous : la suite de lettres « médecin », présentée au milieu d'autres mots ou de non-mots comme « bardonit », est jugée plus rapidement être un mot français lorsque « médecin » a été immédiatement précédé par « infirmier » que lorsqu'il l'a été par « autobus ». Dans l'interprétation restrictive du phénomène, on se contente de dire, plus ou moins intuitivement, que les représentations correspondant à « infirmier » et à « médecin » sont sémantiquement proches, [...] alors que celles qui correspondent à « autobus » et « médecin » sont éloignées l'une de l'autre. Dans l'interprétation théorique dominante, la propagation de l'activation (préactivation) atteint plus rapidement <médecin> dans le premier cas que dans le second à cause de ce facteur de proximité.

[...] on applique la conception présentée ci-dessus, et on considère que les significations de deux mots apparentés sémantiquement (d'« infirmier » et de « médecin »), ont une partie commune, due aux traits sémantiques qu'elles partagent. Si on présente les deux mots successivement, une certaine proportion des traits sémantiques communs se trouvera déjà dans un état actif au moment où le second mot sera présenté, perçu et interprété par l'esprit/cerveau du participant. [...] Ce ne sera pas le cas lorsque le premier mot présenté (par exemple « autobus ») sera sans relation sémantique avec le second, et où, par conséquent, aucun trait de la représentation de celui-ci ne pourra être activé par avance. (ibid. : 339-340)

Cette conception de la proximité sémantique par champs rend superflu le recours à des notions telles que l'hyponymie/hyponymie/co-hyponymie, la synonymie et la métonymie, qui sont, pour nos objectifs, des relations trop restreintes. Elles ne permettent de mettre en relation que des mots de même catégorie grammaticale, alors que la conception par traits en commun permet de rendre compte de la proximité entre « médecin » et « soigner » par exemple.

Après ce développement sur la signification d'un mot pensée comme ensemble de traits, qui nous a permis de définir ce que nous entendons par « proximité sémantique indirecte » (deux mots ont une signification proche si leurs sèmes possèdent au moins un sème en commun), nous pouvons revenir à l'exposé de l'ajustement sémantique, et préciser davantage le processus.

Un mot dans un texte en langue étrangère fait parfois penser à un mot en langue maternelle, sans que ce dernier convienne dans le contexte. C'est ce que nous avons indiqué au début de cette partie : une proximité formelle entre un mot étranger et un mot français fait penser à ce mot français, mais le sens de celui-ci demande à être adapté et un ajustement sémantique permet alors d'aboutir à l'équivalent. Nous savons maintenant qu'il faut, pour pratiquer cet ajustement, choisir parmi les mots qui sont proches sémantiquement, c'est-à-dire qui font partie du champ sémantique (il s'agit alors de passer au niveau de l'hypersémème⁷²), du mot évoqué en français par sa forme, et que nous appellerons désormais « mot-relais »⁷³. Néanmoins, ce mot-relais est proche sémantiquement d'un nombre considérable de mots, puisqu'il s'agit de tous les mots partageant avec lui au moins un sème. Comment donc choisir l'équivalent du mot étranger ? Quel est ou quels sont les sèmes de transfert du mot-relais à sélectionner ?

La recherche de l'équivalent du mot étranger, proche sémantiquement directement avec lui, est évidemment guidée par le contexte. Celui-ci met à disposition tellement d'informations que le choix, au lieu de devoir être effectué parmi des centaines de mots, est en réalité très fortement orienté. Nous empruntons à F. Rastier sa définition du « contexte » :

Nous définissons donc le contexte d'un sémème comme l'ensemble des sémèmes qui dans un texte donné entrent avec lui en relation d'incidence – quelle que soit la position des expressions qui les manifestent. Plus précisément, le contexte passif d'un sémème est l'ensemble des sémèmes sur lesquels il a une incidence, et son contexte actif est l'ensemble des sémèmes qui ont une incidence sur lui. (Rastier 1987 : 73)

Un rappel du rôle du texte, et notamment des isotopies, sur la constitution des sémèmes va nous permettre de préciser ce que nous entendons par « adaptation au contexte ».

Le processus de compréhension se faisant pendant la lecture d'un texte, nous devons commencer par rappeler comment un mot prend son sens quand il est en texte. Nous nous situons dans une approche gestaltiste, c'est-à-dire que c'est le texte qui donne

⁷² Voir le passage sur l'hypersystème p. 87.

⁷³ Nous suivons en cela la terminologie introduite par É. Castagne dans son article de 2007, et dont nous avons rappelé les points principaux p. 69.

son sens aux mots, et non l'addition de mots qui donne son sens au texte. Selon la formule de F. Rastier, « le global [...] détermine le local et le *recompose* » (Rastier 2002a), et selon la vision gestaltiste, « la perception est d'emblée perception de relations, car une relation n'est pas nécessairement le fruit d'une opération intellectuelle, ni même postérieure à la perception des termes qu'elle relie » (Rosenthal & Visetti 2001).

Chaque sémème situé en texte influence les autres sémèmes du texte et est influencé par eux, et ce de trois façons possibles : par l'activation, l'inhibition et la propagation de sèmes. L'activation de sèmes permet d'actualiser des sèmes afférents, qui ne sont donc actualisés que si le contexte le demande ; par exemple, le sème /arrive en retard/ qui n'est pas définitoire dans le sémème de « train » est activé dans la phrase « Mon train m'a fait louper ma correspondance ». L'inhibition de sèmes virtualise des sèmes inhérents ; par exemple, le sème /ingérer/ est inhibé dans le sémème de « dévorer » dans la phrase « Paul a dévoré Marie des yeux toute la soirée ». Enfin, la propagation de sèmes actualise des sèmes afférents en contexte ; par exemple, le sème /néfaste/ se propage dans le sémème de *Women* dans le titre de l'ouvrage de G. Lakoff *Women, Fire, and Dangerous Things*. Pour résumer :

On retient donc que la nature et le nombre des sèmes d'un sémème varie selon ses occurrences ; et, plus précisément :

- (i) Tout sème peut être virtualisé par le contexte.*
- (ii) Tout sème n'est actualisé qu'en fonction du contexte.*
- (iii) Aucun sème n'est actualisé en tout contexte. (Rastier 1987 : 82)*

Selon les mots de J.-E. Tyvaert, c'est la « textualisation » qui permet d'analyser un sémème :

[...] des indications sémantiques multiples (et toujours artificiellement ordonnées dans l'ordre de l'analyse lexicale) interagissent d'un mot à l'autre dans tout le fragment textuel examiné [...], et transforment une signification artificielle « hors texte » du mot isolé (construite en fonction de critères totalement indépendants de l'élaboration du sens dans le texte qui doit se faire lors de la rencontre effective du mot dans un texte) en une signification effective « en texte » du mot en fonction de l'ordonnement des indications établi selon l'importance de leurs co-occurrences dans le texte examiné. C'est donc par un processus que nous proposons d'appeler

« textualisation » que ces indications ainsi classées participent à la définition du mot lui permettant de contribuer au sens du texte. (Tyvaert 2005 : 114)

Les sèmes d'un texte s'interdéfinissent, si bien qu'un sème se trouve toujours dans plusieurs sèmes, le tout créant une « isotopie » :

Fondamentalement, une isotopie est instituée par une série de relations d'identité entre sèmes. Ces relations induisent des relations d'équivalence entre sèmes. [...]

Cela conduit à un déplacement de problématique. En général, on considère l'isotopie comme une forme remarquable de combinatoire sémique, un effet de la combinaison des sèmes. Ici au contraire, où l'on procède paradoxalement à partir du texte pour aller vers ses éléments, l'isotopie apparaît comme un principe régulateur fondamental. Ce n'est pas la récurrence de sèmes déjà donnés qui constitue l'isotopie, mais à l'inverse la présomption d'isotopie qui permet d'actualiser des sèmes, voire les sèmes. (Rastier 1987 : 11-12)

La notion de « présomption d'isotopie » est capitale, car elle met au centre de l'interprétation le processus de lecture, et donc le lecteur lui-même. En effet, le lecteur part du principe, implicite, qu'un texte est cohérent. Nous retrouvons cette idée dans la définition de l'isotopie sémantique selon A. J. Greimas : « [...] l'isotopie sémantique [...] rend possible la lecture uniforme du discours, telle qu'elle résulte des lectures partielles des énoncés qui le constituent, et de la résolution de leurs ambiguïtés qui est guidée par la recherche d'une lecture unique » (Greimas & Courtés 1979 : 197). Nous modulons cette définition en précisant d'une part, avec F. Rastier, que ce n'est pas l'isotopie qui rend possible la lecture uniforme du texte, mais la lecture uniforme du texte qui crée l'isotopie, et d'autre part qu'il n'y a pas recherche d'une lecture unique, car la lecture unique, unifiante, est selon nous un processus inconscient, à l'œuvre de façon automatique dès que le lecteur commence à lire ou écouter les premiers mots d'un texte.

Nous pouvons à ce propos faire appel aux notions de « contrat de lecture » et d'« horizon d'attente », qui, pour appartenir à la terminologie de la théorie littéraire, ne rendent pas moins compte de processus engagés lors de la lecture de tous types de textes. La lecture d'un texte est toujours précédée de la lecture de ce que G. Genette appelle le « paratexte », c'est-à-dire « un certain nombre de productions, elles-mêmes verbales ou

non, comme un nom d'auteur, un titre, une préface, des illustrations, dont on ne sait pas toujours si l'on doit ou non considérer qu'elles [...] appartiennent [au texte], mais qui en tout cas l'entourent et le prolongent, précisément pour le présenter » (Genette 1987 : 198). Le paratexte donne donc des informations, qui, loin d'être anecdotiques, situent le texte par rapport à un genre et ainsi commandent un certain type de lecture :

Eu égard à sa fonction de présentation, le paratexte est le lieu où se noue explicitement le contrat de lecture. Pour éclairer cette notion, rappelons qu'on ne lit pas tous les textes de la même manière : un roman policier ne suscite pas les mêmes attentes qu'un roman historique ; un roman réaliste ne respecte pas les mêmes règles qu'un roman fantastique. Le paratexte, en donnant des indications sur la nature du livre, aide le lecteur à se placer dans la perspective adéquate. (Jouve 2007 : 8)

Le contrat de lecture, créé par le paratexte, implique un « horizon d'attente » du lecteur. L'auteur s'engageant à respecter certaines règles propres au genre, le lecteur s'attend à voir ces règles respectées et oriente sa lecture pour trouver dans le texte ce qu'il s'attend à y trouver :

Toutes les indications données par le texte avant que ne commence la lecture dessinent un champ de possibles que le lecteur identifie plus ou moins consciemment. Si cet horizon d'attente est déçu par le texte, il y a violation du pacte de lecture et la communication ne fonctionne plus. [...] La lecture est structurée par des conventions qui, pour n'être pas explicites, n'en pèsent pas moins lourdement sur notre relation au texte. (ibid. : 8)

L'horizon d'attente du lecteur favorise ainsi l'identification d'isotopies. C'est cette notion d'isotopie qui va nous permettre de préciser davantage le fonctionnement de l'ajustement sémantique.

Une fois qu'un mot-relais est identifié par sa proximité formelle avec le mot étranger, l'apprenant adapte le sens de ce mot-relais en cherchant un mot qui non seulement possède au moins un sème en commun avec ce mot-relais, mais aussi qui ne rompt pas l'isotopie, ou une des isotopies, du texte. Le ou les sèmes de transfert à garder dans le sémème du mot-relais sont également ceux qui sont compatibles avec une isotopie.

En d'autres termes, si le mot identifié crée une allotopie⁷⁴, puisque la lecture d'un texte ne souffre pas d'incompatibilité, alors le lecteur rétablit l'isotopie en sélectionnant le ou les sèmes compatibles, qui deviennent ainsi des sèmes de transfert, et choisit un mot dont le sémème contient ces sèmes de transfert, ainsi que d'autres qui ne rompent pas l'isotopie. Toute cette procédure n'est sûrement pas consciente chez les apprenants ; elle n'en reste cependant pas moins valable. Ajoutons que des arguments d'ordre phraséologique doivent également intervenir dans le choix de l'équivalent du mot étranger.

Nous pouvons maintenant développer davantage le parallèle que nous avons commencé à faire entre la proximité formelle indirecte et la proximité sémantique indirecte : tout comme, lors d'un ajustement formel, un apprenant cherche à remplacer un phonème par un autre qui en est proche, c'est-à-dire qui possède des traits articulatoires en commun avec lui, lors d'un ajustement sémantique, un apprenant cherche à remplacer un mot par un autre qui en est proche, c'est-à-dire qui possède des traits sémantiques en commun avec lui.

Reprenons les différentes étapes de l'ajustement sémantique en les illustrant par un exemple. Soit le texte suivant :

« Vancouver Opera is pleased to announce the appointment of Jonathan Darlington as its Principal Conductor. He will take up his new duties in the 2002-2003 season and will make his début as Principal Conductor at a gala concert February 27, 2003 celebrating the orchestra's 25th anniversary. He will also conduct the final production of the 2002/2003 season, Puccini's *La Bohème*. »⁷⁵.

Nous portons notre intérêt sur le mot *conductor*. Ce mot anglais est proche formellement directement du mot français « conducteur ». Cependant, le recrutement d'un « conducteur » pour une production de l'Opéra de Vancouver paraît pour le moins étrange. En effet, l'isotopie de la musique peut être repérée dans le texte, dans les sémèmes des mots *Opera*, *concert*, *orchestra*, « Puccini » et « La Bohème » (et, par propagation, dans *season*). Le mot « conducteur » crée donc, lors de l'interprétation de ce texte, une

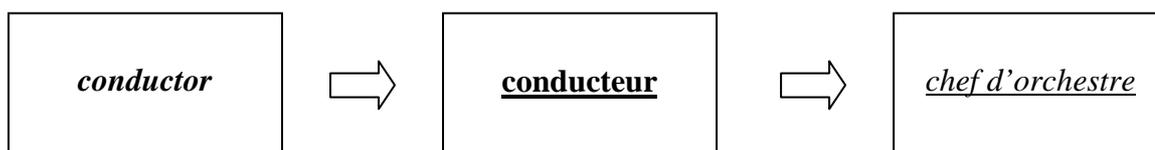
⁷⁴ « relation de disjonction exclusive entre deux sémèmes (ou deux complexes sémiques) comprenant des sèmes incompatibles ; par extension, rupture d'isotopie » (Rastier 2002b)

⁷⁵ http://www.vancouveropera.ca/pr/pr_28_VO_Appoints-.pdf

Nous traduisons : « L'Opéra de Vancouver est heureux d'annoncer le recrutement de Jonathan Darlington en tant que chef d'orchestre principal. Il prendra ses nouvelles fonctions pour la saison 2002-2003 et fera ses débuts en tant que chef d'orchestre principal au concert gala du 27 février 2003 organisé en l'honneur du 25^{ème} anniversaire de l'orchestre. Il dirigera également la dernière production de la saison 2002/2003, *La Bohème* de Puccini. »

allotopie, par son sème /véhicule/. Il y a alors sélection des sèmes du sémème de « conducteur » qui sont compatibles avec le contexte, à savoir /personne/ et /qui dirige/. Ces sèmes acquièrent le statut de sèmes de transfert. La recherche d'un mot dont le sémème possède également ces sèmes, et dont les autres sèmes sont compatibles avec l'isotopie de la musique, mène rapidement au mot « chef d'orchestre ».

Le schéma donné au début de cette section est vérifié : *conductor* et « chef d'orchestre » sont proches sémantiquement directement (ils peuvent être traduction l'un de l'autre) ; *conductor* et « conducteur » sont proches formellement ; et enfin « conducteur » et « chef d'orchestre » sont proches sémantiquement indirectement, par le biais des sèmes de transfert /personne/ et /qui dirige/.



gras : proximité formelle (directe)

italique : proximité sémantique directe

souligné : proximité sémantique indirecte

Nous avons vu, dans la section sur la proximité formelle indirecte, que l'ajustement formel correspondait, dans le cadre d'une analyse interlinguistique, à un procédé qui existe en français (dans un échange du type /v/-/b/ : « h**i**ver »-« h**i**berner »). Nous pouvons faire la même remarque à propos de l'ajustement sémantique. Nous l'avons décrit ci-dessus en tant que méthode d'aide à la compréhension de textes rédigés en langues étrangères, mais les lecteurs pratiquent également cette adaptation du sens quand ils lisent dans leur langue. Rappelons d'ailleurs que nous avons emprunté l'appellation « sème de transfert » à J. Tournier, qui a forgé cette notion pour décrire le processus, intralinguistique, de compréhension des métaphores. De même, pour R. Martin, la similitude entre le sens propre et le sens métaphorique d'un mot repose sur l'identité d'au moins un des sèmes spécifiques entre les deux sémèmes considérés. De la sorte, les deux acceptions de « impasse » sont formalisées ainsi :

Σ^1 : « Rue /S¹/ sans issue /s¹₁/ »

Σ^2 : « Situation /S²/ sans issue /s²₁/ favorable /s²₂/ »,

Σ symbolisant un sémème, S son archisémème (le genre prochain) et $s_1, s_2 \dots$ ses différents sèmes spécifiques. La relation qui unit les deux acceptions d'« impasse » est fondée sur l'identité $s^1_1 = s^2_1$, c'est-à-dire sur le sème /sans issue/ qui est partagé par les deux sémèmes (Martin 1983 : 70).

Ajoutons que c'est le même phénomène qui est à l'œuvre quand, pendant la production d'un énoncé, le locuteur précise sa pensée en la reformulant. Ainsi, si un locuteur déclare « Nous avons bu le champagne dans de magnifiques coupes, ou plutôt de magnifiques flûtes », on peut décrire le passage de « coupe » à « flûte » par la sélection du sème de transfert /récipient servant à boire/ qui est commun aux deux sémèmes, et par l'ajout du sème /étroit/, qui n'appartient qu'au sémème de « flûte ».

2.4. Bilan

Il convient maintenant de résumer ce qui a été écrit dans cette partie sur la transparence.

Nous nous étions fixé les objectifs suivants : définir la transparence directe puis indirecte par des critères formels et sémantiques synchroniques et objectifs, qui permettent de fixer un seuil entre les mots transparents et les mots opaques, et, par suite, de fixer un seuil entre les mots transparents directement et les mots transparents indirectement.

La transparence directe se définit formellement interlinguistiquement par une même trame consonantique ou quatre premières lettres identiques. Les règles précises que nous avons tirées sont les suivantes :

- pour que deux mots soient proches formellement directement, ils doivent avoir la même « trame consonantique », c'est-à-dire les mêmes consonnes apparaissant dans le même ordre, avec éventuellement dans un des mots une consonne de plus ou de moins au contact de ses consonnes initiales ou médianes, et la dernière consonne, si elle est au moins la troisième, éventuellement différente ;

- si les deux mots comparés sont composés de quatre lettres ou plus, il suffit qu'ils aient au moins les quatre premières en commun pour être considérés comme proches formellement directement. Si l'un des deux mots a moins de quatre lettres, il suffit que le mot le plus long commence par toutes les lettres du mot le plus court pour qu'ils soient considérés comme proches formellement directement.

La transparence directe se définit sémantiquement interlinguistiquement par une équivalence en texte. Si un mot constitue l'équivalent de l'autre dans une traduction, alors ils sont transparents sémantiquement directement.

On pourrait s'étonner que nous n'ayons pas fait appel à l'analyse en sèmes pour définir la proximité sémantique directe, alors que nous l'avons utilisée pour la proximité sémantique indirecte. Mais nous ne pouvons définir la proximité sémantique directe par une identité des sèmes ou d'un grand nombre de sèmes, car la proximité sémantique directe relie un mot étranger et son équivalent dans une autre langue. Or, l'analyse sémique d'un mot se pratique par la comparaison de ce mot avec les autres mots de son système. On définit ainsi le sémème du mot considéré, et, le sémème n'étant valable qu'au sein d'un système linguistique, on ne peut mettre en relation les sémèmes de deux mots de langues

différentes. De plus, les sèmes ne sont pas des universaux mais des traits spécifiques à chaque langue :

[...] les sèmes sont uniquement définis par des rapports entre sémèmes, qui dépendent des structures lexicales de chaque langue particulière. [...] En somme, on ne peut affirmer que les sèmes soient des universaux, à moins de fonder cette thèse sur leur assimilation à des idées innées, ou à des qualités du réel extralinguistique. On fera ici l'économie de ces présupposés coûteux et inutiles, car les sèmes sont des unités linguistiques, et les universaux, qu'ils soient substantiels ou formels, ne sont concevables qu'au niveau le plus abstrait, dont l'existence est purement métalinguistique. (Rastier 1987 : 28)

Nous ne pouvons donc pas comparer les sèmes d'un mot d'une langue avec les sèmes d'un mot d'une autre langue. C'est pourquoi nous avons défini la proximité sémantique directe par une équivalence en texte, c'est-à-dire par la relation qui unit un mot et son équivalent dans deux textes dont l'un constitue la traduction de l'autre. Les deux faces de la transparence directe sont ainsi définies par l'application de principes dégagés pour une analyse intralinguistique à une analyse interlinguistique : l'homonymie et la paronymie pour la face formelle, et la synonymie et la parasynonymie pour la face significative.

La transparence indirecte se définit formellement interlinguistiquement par une même trame consonantique avec recours à l'ajustement formel. La règle est la suivante : pour que deux mots soient proches formellement indirectement, ils doivent avoir la même trame consonantique, c'est-à-dire les mêmes consonnes apparaissant dans le même ordre, dont au moins une, dans un mot, appartient au champ phonologique de la consonne correspondante dans l'autre mot, et avec éventuellement dans un des mots une consonne de plus ou de moins au contact de ses consonnes initiales ou médianes, et la dernière consonne, si elle est au moins la troisième, éventuellement différente.

La transparence indirecte se définit sémantiquement intralinguistiquement par le partage d'au moins un sème entre le mot évoqué formellement en langue maternelle et l'équivalent du mot étranger. Nous avons utilisé la terminologie suivante : le « mot-relais », qui est proche formellement du mot étranger, est proche sémantiquement de l'équivalent par l'intermédiaire d'au moins un « sème de transfert ».

Nous avons fait en sorte de nous tenir à nos objectifs, à savoir d'abord la plus grande objectivité possible (l'intuition étant ainsi exclue en tant que mode de comparaison) :

- pour caractériser la proximité formelle directe, nous avons mis au jour le critère de la trame consonantique par l'observation d'une liste de paires de mots qui ont pour seule caractéristique commune d'être proches du point de vue de la forme ;

- pour caractériser la proximité formelle indirecte, nous avons établi les champs phonologiques par la prise en compte des traits pertinents partagés par les phonèmes des langues ;

- pour caractériser la proximité sémantique directe, nous avons fait reposer la similitude de sens entre deux mots de langues différentes sur l'observation d'une équivalence en texte ;

- pour caractériser la proximité sémantique indirecte, nous avons fondé la relation sémantique qui unit deux mots d'une même langue sur le partage d'au moins un sème.

Nous avons également tenu à définir la transparence en synchronie actuelle (la prise en compte de la diachronie étant ainsi exclue) :

- pour caractériser la proximité formelle directe, nous avons observé et comparé des formes actuelles ;

- pour caractériser la proximité formelle indirecte, nous avons fait reposer les équivalences entre phonèmes sur des propriétés articulatoires – et non sur des lois de phonétique historique ;

- pour caractériser la proximité sémantique directe, nous avons fait appel à une équivalence en texte – et non à un étymon commun ou une relation d'emprunt ;

- pour caractériser la proximité sémantique indirecte, nous avons exploité la comparaison de sémèmes, qui définissent le sens d'un mot à un moment donné.

Nous voulions enfin que notre définition de la transparence permette de fixer un seuil entre les mots transparents et les mots opaques : en transparence directe aussi bien qu'en transparence indirecte, le critère formel de la trame consonantique suffit à fixer ce seuil. Soit les mots considérés ont la même trame consonantique, soit ils ont des trames différentes. Ce même critère formel suffit à fixer le seuil entre les mots transparents directement et les mots transparents indirectement : soit il n'y a pas besoin d'avoir recours

à un ajustement formel pour définir l'identité des trames consonantiques des mots considérés, soit un ajustement formel est nécessaire.

Grâce aux principes dégagés, nous sommes maintenant en mesure de comparer entre eux des mots appartenant à des lexiques étrangers et de dire s'ils sont transparents ou non. Illustrons donc les règles définies dans cette deuxième partie par des exemples⁷⁶.

Pour qu'il y ait transparence directe, il faut qu'il y ait proximité formelle directe *et* proximité sémantique directe. Par exemple, le verbe EN *change* (« changer »), le nom DE *Partei* (« parti ») et l'adjectif NL *lang* (« long ») sont transparents directement par rapport au français ; le verbe DE *singen* (*sing*) et le nom NL *peer* (*pear*) sont transparents directement par rapport à l'anglais ; le nom NL *spin* (*Spinne*) est transparent directement par rapport à l'allemand.

Pour qu'il y ait transparence indirecte, il faut qu'il y ait :

- i) proximité formelle directe et proximité sémantique indirecte, ou bien
- ii) proximité formelle indirecte et proximité sémantique directe, ou bien
- iii) proximité formelle indirecte et proximité sémantique indirecte.

Par exemple :

- i) proximité formelle directe et proximité sémantique indirecte :

le verbe EN *learn* (« apprendre » ; mot-relais : « lire » ; sème de transfert : 's'instruire'), le nom DE *Jahr* (« an » ; mot-relais : « jour » ; sème de transfert : 'durée') et le nom NL *deel* (« plancher » ; mot-relais : « dalle » ; sème de transfert : 'sol') sont transparents indirectement par rapport au français.

- ii) proximité formelle indirecte et proximité sémantique directe :

le nom EN *road* (« route » ; échange /d/-/t/), le verbe DE *suchen* (« chercher » ; échange /z/-/ʃ/) et l'adjectif NL *enig* (« unique », échange /x/-/k/) sont transparents indirectement par rapport au français.

- iii) proximité formelle indirecte et proximité sémantique indirecte :

⁷⁶ Nous donnons, quand cela est possible, des exemples pour chacune des paires de langues, pour montrer que les règles peuvent s'appliquer à toutes les combinaisons, et les exemples pour les couples de langues EN/FR, DE/FR et NL/FR sont tirés de l'analyse du corpus, présentée dans la troisième partie de ce travail. Néanmoins, puisque nous n'avons pas appliqué l'analyse à d'autres couples de langues, seuls des exemples par rapport au français sont donnés pour la transparence indirecte, qui nécessite pour sa détermination une méthodologie bien précise.

l'adjectif EN *safe* (« sûr » ; mot-relais : « sauver » ; échange /f/-/v/ ; sème de transfert : 'danger'), le nom DE *Teil* (« partie » ; mot-relais : « cellule » ; échange /t/-/s/ ; sème de transfert : 'constitutif') et le verbe NL *lijden* (« souffrir » ; mot-relais : « plainte » ; échange /d/-/t/ ; sème de transfert : 'douleur') sont transparents indirectement par rapport au français.

Avoir caractérisé la transparence indirecte révèle tout son intérêt. Des mots tels que EN *bath*/DE *Bad* ou DE *liefern*/FR « livrer » et d'autres mots tels que EN *look*/FR « regarder » (par le biais de FR « loucher ») ou NL *drinken*/FR « boire » (par le biais de FR « trinquer ») n'auraient pas pu relever du phénomène de la transparence sans la prise en compte d'ajustements formels, dans le premier cas, et d'ajustements sémantiques, dans le second, dans la définition de la transparence.

3. De l'interlexique à l'hyperlexique

Nous avons exposé dans la partie précédente quelles sont les règles qui permettent de décider si un mot étranger doit être considéré comme transparent pour un francophone ou non, et fourni des explications quant au processus de compréhension au niveau lexical tel qu'il se déroule en texte. Nous avons d'abord décrit la transparence directe (chapitre 2.2.), rendant compte des mots immédiatement compréhensibles, puis la transparence indirecte (chapitre 2.3.), rendant compte des mots dont la compréhension est possible après un ou plusieurs ajustements.

Nous souhaitons maintenant appliquer ces règles à l'analyse d'un corpus et déterminer le pourcentage de mots transparents entre l'anglais et le français, entre l'allemand et le français et entre le néerlandais et le français. L'ensemble des mots transparents directement correspondra à l'interlexique, et l'ensemble des mots transparents directement et indirectement correspondra à l'hyperlexique. Nous traiterons les trois langues germaniques dans l'ordre anglais-allemand-néerlandais, selon leur degré de familiarité pour les Français, qui étudient très majoritairement l'anglais, puis pour certains l'allemand, et enfin pour un très petit nombre d'entre eux le néerlandais⁷⁷. Nous pourrons ensuite revenir sur ce choix d'apprentissage, à la lumière de ce que nos analyses nous auront permis d'observer.

Les règles déterminant la transparence directe et la transparence indirecte formulées dans la deuxième partie de notre travail sont des règles définissant la transparence « en texte ». Ainsi la proximité sémantique directe relève-t-elle d'une équivalence entre deux textes dont l'un est la traduction de l'autre, la proximité formelle indirecte d'un ajustement formel dont le choix est guidé par le contexte et la proximité sémantique indirecte d'une sélection des sèmes d'un sémème qui sont conformes aux isotopies du texte. Pour appliquer ces règles, il faudrait donc un corpus constitué de textes. Or, un corpus de textes nous a semblé être un outil inapproprié, et ce pour deux raisons. D'abord parce que, en texte, les mots ont un *sens*, et non une *signification*⁷⁸, et que nulle part nous ne pourrions nous procurer la détermination du sens d'un mot en texte (notre critère de la plus grande objectivité possible nous empêchant de le décrire nous-même). Nous ne pouvons par définition pas vérifier le sens d'un mot étranger dans un dictionnaire : un dictionnaire ne peut recenser tous les sens d'un mot, chaque sens apparaissant en texte étant unique ; il

⁷⁷ Selon une étude menée par le Ministère de l'Éducation Nationale sur les langues vivantes dans le second degré à la rentrée 2008, l'anglais est choisi dans 94 % des cas comme LV1 et l'allemand dans moins de 8 % des cas. Les autres langues représentent 2 % des premières langues. Cette étude est disponible à la page suivante : http://media.education.gouv.fr/file/2009/84/9/chap4-16_117849.pdf (consultée le 16/10/09).

⁷⁸ Voir p. 121 pour un développement sur la différence entre « sens » et « signification ».

présente des « résumé[s] conventionnel[s] » (Rastier 2003), à savoir des significations, valables « hors texte ». Ensuite, d'un point de vue pratique, et dans le même ordre d'idée que l'argument précédent, l'analyse de textes étrangers impliquerait que nous soyons en possession de traductions d'articles de presse du type de ceux utilisés dans la méthode ICE⁷⁹, pour pouvoir comparer les textes sources aux textes cibles, ce qui n'est pas le cas. Nous travaillerons donc sur un corpus de mots isolés, que nous étudierons hors texte. Le corpus sera présenté au début de cette troisième partie (3.1. Présentation du corpus).

Au début des deux chapitres suivants (3.2. L'interlexique et 3.3. L'hyperlexique), nous transposerons à une analyse hors texte les règles de la transparence en texte formulées dans la deuxième partie. La méthodologie ainsi fixée sera ensuite appliquée au corpus de façon systématique. Nous n'affirmons donc nullement que les mots qui auront été considérés comme transparents, directement ou indirectement, par l'application de cette méthode d'analyse correspondent aux mots qui pourront être compris par des lecteurs francophones. Il s'agira uniquement des mots qui répondent aux règles fixées. C'est pourquoi les résultats chiffrés que nous tirerons de cette analyse ne constitueront ni une surévaluation de la capacité de compréhension d'un lecteur, sous prétexte qu'un lecteur n'aurait pas obligatoirement songé à toutes les procédures mises en œuvre, ni une sous-évaluation, sous prétexte que la méthodologie effectue des choix et qu'un lecteur aurait pu envisager des voies de compréhension supplémentaires. Les listes de mots obtenues ne reflétant pas les mots dont nous disons qu'ils sont compréhensibles par des apprenants mais les mots auxquels on aboutit par l'application stricte de la méthodologie, on ne s'étonnera pas de ce que la qualification de certains mots étrangers comme transparents puisse paraître contre-intuitive. De même, certaines voies qui auront été empruntées pour accéder à la compréhension de mots étrangers par le biais d'ajustements formels et/ou sémantiques pourront parfois sembler étonnantes. Il convient donc de garder à l'esprit que l'appartenance de tel ou tel mot à la catégorie des mots transparents n'est pas de notre ressort, mais tient à l'application méthodique des règles d'analyse que nous présenterons dans cette partie et qui découlent directement des définitions formulées dans la deuxième partie.

⁷⁹ Voir p. 44 pour une présentation de la méthode.

3.1. Présentation du corpus

Nous avons constitué notre corpus grâce à la base de données de CELEX (*Dutch Centre for Lexical Information*⁸⁰, Baayen *et al.* 1995), développée conjointement par l'université de Nimègue, l'*Institute for Dutch Lexicology* de Leyde, le *Max Planck Institute for Psycholinguistics* de Nimègue, et l'*Institute for Perception Research* de Eindhoven. CELEX Database comporte des données d'ordres orthographique, morphologique, phonologique, syntaxique et fréquentiel sur les lexiques de l'anglais, de l'allemand et du néerlandais. Nous l'avons choisie car elle répond à nos exigences, par sa fiabilité⁸¹ et le fait que nous puissions y trouver, de façon unifiée et cohérente, toutes les données dont nous avons besoin.

Pour les trois langues germaniques qui nous intéressent, nous avons choisi des listes de mots lemmatisés. Le choix en faveur des lemmes s'est imposé puisque, pour évaluer la proximité formelle et la proximité sémantique entre les mots étrangers du corpus et des mots français, nous aurons à comparer ces mots étrangers avec leurs équivalents dans des dictionnaires bilingues, puis à chercher ces équivalents français dans un dictionnaire monolingue (voir chapitres 3.2. et 3.3. sur le développement sur la méthodologie d'analyse), et que les entrées des différents types de dictionnaires sont constituées par des formes lemmatisées.

Nous avons ensuite croisé les informations de catégorie grammaticale et de fréquence, de sorte à avoir des listes de mots lemmatisés, donnés avec leur catégorie et rangés par ordre décroissant de fréquence. Cette première indication s'est avérée nécessaire pour savoir avec quelles formes françaises comparer les formes pluricatégorielles (pour savoir, par exemple, s'il faut comparer le mot EN *book* qui a une fréquence de 7 780 avec les noms français « livre », « bouquin » ou avec les verbes « réserver », « retenir »). Le critère de fréquence, quant à lui, a été retenu pour une raison majeure. Il s'agit, grâce à cette étude, de savoir quel pourcentage du lexique d'une langue peut être considéré comme potentiellement compréhensible. Par conséquent, nous avons sélectionné les mots dont la probabilité d'être rencontrés est la plus haute, puisque le cadre de cette analyse est l'*intercompréhension*. En effet, comme le signale P. Bogaards,

⁸⁰ Voir la page d'informations : <http://www ldc.upenn.edu/Catalog/docs/LDC96L14/README>

⁸¹ Cette base de données, constituée par des chercheurs, est considérée comme la base de travail de référence dans de très nombreuses études, par exemple dans Harm & Seidenberg 1999, Barzilay & McKeowm 2001 et Namer 2003.

[s]ous l'angle de la communication, il est évident que la fréquence est plus importante pour la compréhension que pour la production [...] : pour comprendre ce qui est dit ou écrit, on dépend des mots que les autres choisissent d'employer, alors qu'en production l'apprenant peut choisir les éléments qu'il a à sa disposition, et qui ne sont pas nécessairement les plus fréquents. (Bogaards 1994 : 114-115)

Ce sont donc les mots les plus fréquents qui ont été retenus. Pour l'anglais, les données d'ordre fréquentiel de CELEX Database proviennent du corpus COBUILD, mis au point par l'université de Birmingham en 1991, d'après des sources majoritairement écrites⁸², et qui contient 17,9 millions de mots. Pour l'allemand, ces données proviennent du corpus MANNHEIM mis au point par l'« Institut für deutsche Sprache » en 1984, d'après des sources majoritairement écrites⁸³, et qui contient 6 millions de mots. Enfin, pour le néerlandais, ces données proviennent du corpus INL mis au point par l'« Instituut voor Nederlandse Lexicologie » de Leyde en 1995, d'après des sources exclusivement écrites⁸⁴, et qui contient 42,38 millions de mots.

Nous avons limité le nombre de mots à analyser par langue à 1 000. Les différentes études menées sur le sujet montrent qu'un nombre relativement restreint de mots couvre la plus grande partie de n'importe quel texte. Par exemple, les recherches conduites sur le TLF mettent en lumière le fait que « moins de 1 000 mots couvrent 90 % du corpus, que les 5 ou 6 000 suivants en couvrent 8 % et que tout le reste ne couvre que 2 % » (Picoche 2000 : 199). P. Bogaards résume les résultats des publications sur la notion de « couverture lexicale » par les points suivants :

- *les dix mots les plus fréquents représentent le quart ou plus des occurrences ;*
- *il en faut une cinquantaine pour couvrir la moitié des occurrences ;*
- *il suffit d'une centaine de mots pour arriver à plus de 60 % des occurrences ;*
- *avec les mille premiers mots, les estimations varient de 70 % [...] à 90 % [...] ;*

⁸² Il s'agit de journaux, de magazines, de livres, de brochures et de rapports. Les sources orales sont constituées de réunions, d'interviews et de conversations.

⁸³ Il s'agit de traités, de biographies, d'essais, de modes d'emploi, de programmes de partis, de discours, de lettres, de prospectus, de poèmes, d'éditoriaux, de recensions, de publicités, de contes, de communiqués de presse, de romans, de décrets et de comptes rendus de recherches. Les sources orales sont constituées de pièces radiophoniques et d'interviews.

⁸⁴ Il s'agit de livres, de magazines, et majoritairement de journaux.

- pour atteindre une couverture de 95 %, il faut, selon les auteurs, 2 000 mots [...], 3 000 [...], 4 000 [...], 4 500 [...] ou 5000 [...]. (Bogaards 1994 : 117)

La limite fixée à 1 000 mots nous a ainsi semblé être suffisante pour couvrir les mots les plus susceptibles d'être rencontrés dans tout type de texte.

Nous avons également limité notre corpus en n'y incluant que les mots lexicaux. Les parties du discours correspondant aux mots lexicaux et aux mots grammaticaux ne font pas consensus dans les grammaires. Ce sur quoi les linguistes s'accordent (Nève 1996 : 135-136 ; Riegel *et al.* 1998 : 536 ; Apothéloz 2002 : 11 ; Martinet 2008 : 118), c'est le critère sur lequel on doit les opposer : les mots lexicaux appartiennent à des « classes ouvertes », c'est-à-dire des classes dans lesquelles le nombre d'éléments peut varier, tandis que les mots grammaticaux appartiennent à des « classes fermées », c'est-à-dire des classes dans lesquelles le nombre d'éléments est stable (les ajouts et les retraits sont extrêmement rares⁸⁵). En synchronie, les mots lexicaux, de par leur classe ouverte, sont en grand nombre, c'est-à-dire que chaque mot lexical a dans son paradigme un nombre très important de substituts, et si leur nombre peut varier, c'est parce que des mots peuvent être créés, empruntés ou tomber en désuétude. Les mots grammaticaux, de par leur classe fermée, sont en petit nombre, c'est-à-dire que chaque mot grammatical a dans son paradigme un nombre restreint de substituts ; ils sont plutôt courts car souvent utilisés, et si leur nombre ne varie pratiquement pas, c'est parce qu'ils appartiennent au système fonctionnel de la langue. On peut imaginer qu'un locuteur invente le mot lexical « *carabouillir » ; en revanche, il y a une chance quasi-nulle qu'il invente un mot grammatical :

[Les morphèmes grammaticaux] forment des paradigmes « figés » dont la modification entraînerait un bouleversement de pans entiers du système grammatical : il suffit d'imaginer les effets de l'introduction dans le français contemporain d'une septième personne verbale ou d'un troisième genre ! (Riegel et al. 1998 : 536)

⁸⁵ Ce fut par exemple le cas au cours du XIV^e siècle, lors de la disparition progressive des articles de cas sujet, entraînée par la disparition de la déclinaison en ancien français (Picoche & Marchello-Nizia 1998 : 224).

Nous suivrons ainsi la classification la plus répandue s'appuyant sur la distinction classe ouverte/classe fermée en donnant l'étiquette de mot lexical aux noms, verbes, adjectifs qualificatifs et adverbes du français, de l'anglais, de l'allemand et du néerlandais, et celle de mot grammatical aux autres parties du discours⁸⁶.

Nous avons retenu les seuls mots lexicaux parce que ce sont eux qui portent la plus grande partie de la charge sémantique d'un texte. Les mots grammaticaux, également appelés « mots outils » ou « mots vides » (Tesnière 1959 : chap. 28 ; Bogaards 1994 : 115 ; Mortureux 2001 : 10 ; Ledegen & Rossi 2002), sont considérés comme ayant la syntaxe des phrases à charge, la mise en relation des informations portées par les mots lexicaux. C'est ce que souligne P. Bogaards :

[...] il n'est pas impossible de faire passer un message sans employer [les mots grammaticaux] : on les supprime dans les télégrammes ou les titres de journaux, dans des situations donc où il est important d'être bref et parfaitement clair ainsi que dans les moments de panique ou de grande émotion. Les mots outils sont aussi volontiers absents dans le parler des petits enfants et des étrangers. (Bogaards 1994 : 115-116)

Il résume cette idée dans le paragraphe suivant :

[...] leur utilité communicative reste limitée. Leur fonction réelle semble plutôt se situer au niveau grammatical et se mettre au service de la correction et de la précision. (ibid. : 116)

Or, nous voulons par cette étude évaluer le nombre de mots qui sont transparents pour des lecteurs habitués aux principes de l'intercompréhension, parmi lesquels l'approximation joue un rôle phare. C'est pourquoi nous nous concentrerons sur les mots lexicaux, également appelés « mots pleins », qui permettent de véhiculer les informations principales d'un texte.

⁸⁶ Voir notamment Niklas-Salminen 1997 : 23 ; Picoche 1997 : 56 ; Delesalle & Gary-Prieur 1976 : 24 ; Lehmann & Martin-Berthet 2005 : 4 ; Riegel *et al.* 1998 : 536.

Le petit test suivant confirmera cette idée. Soit une phrase extraite du *Précis de lexicologie française* (Picoche 1997 : 45). Nous la retranscrivons dans un premier temps en ne donnant que les mots grammaticaux, les mots lexicaux étant remplacés par des points de suspension :

« On ... d'... l'... des ... qu'une ... à la ... des ..., et ... l'... des ... par un ... dans des ... ».

Sans surprise, aucune information ne peut être inférée.

Voici maintenant la même phrase, où ne sont donnés que les mots lexicaux, les mots grammaticaux étant remplacés par des points de suspension :

« ... conviendra ... appeler lexique ... ensemble ... mots ... langue met ... disposition ... locuteurs, ... vocabulaire ... ensemble ... mots utilisés ... locuteur donné ... circonstances données ».

Tel qu'il l'avait été annoncé auparavant, le message manque de précision, mais le lecteur parvient sans difficulté à saisir les informations données par l'auteur.

Notre corpus est ainsi constitué des 1 000 noms, verbes, adjectifs et adverbes les plus fréquents des trois langues germaniques qui nous intéressent (nous avons rejeté les noms propres car ils n'apparaissent pas dans tous les dictionnaires, ainsi que les mots donnés par CELEX Database mais qui n'apparaissent pas dans les dictionnaires, car nous ne pouvions pas les analyser). Plus précisément, la répartition de chaque partie du discours selon la langue est la suivante (les données sont présentées en valeurs absolues puis en pourcentages arrondis à l'unité supérieure) :

	ANGLAIS		ALLEMAND		NEERLANDAIS	
	Absolu	Relatif au total	Absolu	Relatif au total	Absolu	Relatif au total
Adjectifs	131	13 %	173	17 %	172	17 %
Adverbes	115	12 %	119	12 %	117	12 %
Noms	496	50 %	472	47 %	398	40 %
Verbes	258	26 %	236	24 %	313	31 %
Total	1 000	100 %	1 000	100 %	1 000	100 %

Il est intéressant de constater que la répartition du nombre de mots de chaque catégorie est plus ou moins la même dans les trois langues : les noms représentent la grande majorité des corpus ; arrivent ensuite les verbes, les adjectifs et les adverbes.

Nous allons, dans les deux chapitres suivants, appliquer les règles de transparence dégagées dans la deuxième partie de notre travail au corpus que nous venons de présenter. Nous regroupons les deux chapitres suivants par le type de transparence considéré (la transparence directe qui rend compte de l'interlexique puis la transparence globale qui rend compte de l'hyperlexique), pour exposer la méthodologie, commune aux trois langues, choisie pour l'analyse des corpus.

3.2. L'interlexique

Nous allons dans ce chapitre calculer le nombre de mots anglais, allemands et néerlandais transparents directement par rapport au français. Pour ce faire, nous appliquerons les principes de la proximité formelle directe et de la proximité sémantique directe à notre corpus, c'est-à-dire aux 1 000 mots lexicaux les plus fréquents des trois langues germaniques citées. Les mots transparents directement dans les paires de langues considérées correspondront aux réalisations de l'interlexique de ces langues.

Exposons pour commencer la méthodologie choisie pour identifier le nombre de mots des trois langues germaniques transparents directement par rapport au français.

Nous avons déjà insisté⁸⁷ sur le fait que, pour la compréhension d'un signe, le plan du signifiant était primordial lors de la lecture en langue étrangère. Ce n'est qu'une fois que l'apprenant a identifié une forme qu'il peut faire une hypothèse sur le sens. En bref, pas de proximité sémantique sans proximité formelle. Pour analyser le corpus, nous aurions donc voulu, dans un premier temps, pouvoir comparer chacun des mots étrangers avec des mots français du point de vue de la forme, pour repérer dans un second temps des similitudes sémantiques parmi les mots dont la forme aurait été identifiée comme proche. Un programme informatique permettant de trouver quelles sont toutes les formes proches d'une autre forme, en fonction des règles données dans le chapitre « proximité formelle directe », aurait été un outil idéal pour mener cette recherche. Dans l'incapacité de développer un tel programme, nous procéderons dans l'ordre inverse.

La proximité sémantique directe a été définie, dans la partie précédente, selon une approche textuelle. Par l'application de méthodes identifiées pour des procédés intralinguistiques à des procédés interlinguistiques, la parasynonymie interlinguistique a été définie, comme la parasynonymie intralinguistique, comme une équivalence en texte. Deux mots de langues différentes ont donc des significations considérées comme proches directement si l'un constitue régulièrement l'équivalent de l'autre dans des textes dont les uns sont la traduction des autres. Pour notre analyse hors texte, nous considérerons donc comme proches sémantiquement directement deux mots dont l'un est proposé comme équivalent possible de l'autre dans un dictionnaire bilingue. Ce principe qui pourrait sembler anodin correspond en réalité à l'aboutissement de notre réflexion menée dans la

⁸⁷ Voir section 2.1.3., « Du statut inégal de la forme et de la signification ».

section 2.2.2. sur la proximité sémantique directe. De même que, à l'intérieur d'une langue, la signification d'un mot est considérée comme un regroupement des traits communs à ses différents sens en texte, et que cette signification, ce « résumé conventionnel » (Rastier 2003), est notée dans les dictionnaires monolingues, les significations de deux mots de langues différentes sont considérées comme proches par un résumé des équivalences en texte, et cette proximité est notée dans les dictionnaires bilingues. Le dictionnaire bilingue joue ainsi le même rôle de *mémoire de communauté* des unités lexicales qui s'équivalent dans plusieurs textes de langues différentes que le dictionnaire monolingue d'une même unité lexicale dans plusieurs textes d'une même langue.

Les dictionnaires bilingues ont été choisis en fonction de la reconnaissance dont ils jouissent et du fait qu'ils donnent une transcription phonétique des mots-entrées (dont nous avons besoin pour l'établissement de la trame consonantique). Il s'agit du *Robert & Collins* pour l'étude du corpus anglais, du *Langenscheidt Großwörterbuch* pour l'étude du corpus allemand et du *Robert & Van Dale* pour l'étude du corpus néerlandais.

Pour analyser le corpus, nous chercherons donc d'abord les mots étrangers dans un dictionnaire bilingue, outil qui nous permettra de savoir quels sont les équivalents des mots étrangers, donc les mots français qui en sont proches sémantiquement directement. Ensuite, s'il y a parmi les équivalents proposés un mot qui, en plus d'être proche sémantiquement, est proche formellement directement du mot étranger considéré, alors ce mot étranger sera considéré comme transparent directement (puisqu'il est proche formellement directement et proche sémantiquement directement d'un mot français), et appartiendra à l'ensemble des mots rendant compte de l'interlexique des deux langues envisagées. Quatre cas de figure sont donc possibles :

- soit le dictionnaire bilingue ne propose qu'un seul équivalent pour le mot étranger cherché, et cet équivalent est proche formellement directement du mot étranger ; ce dernier est alors considéré comme transparent directement.

- soit le dictionnaire bilingue propose plusieurs équivalents, et un de ces équivalents est proche formellement directement du mot étranger ; ce dernier est alors considéré comme transparent directement.

- soit le dictionnaire bilingue ne propose qu'un seul équivalent pour le mot étranger cherché, et cet équivalent n'est pas proche formellement directement du mot étranger ; ce dernier n'est alors pas considéré comme transparent directement.

- soit le dictionnaire bilingue propose plusieurs équivalents, et aucun de ces équivalents n'est proche formellement directement du mot étranger ; ce dernier n'est alors pas considéré comme transparent directement.

Les exemples suivants, empruntés aux différentes langues, illustrent ces cas de figure :

- En face du nom EN *advantage*, le dictionnaire bilingue donne pour seul équivalent le nom français « avantage ». Puisque l'un constitue l'équivalent de l'autre, ils sont proches sémantiquement directement, et puisqu'ils ont la même trame consonantique (une consonne initiale diffère, mais elle est au contact d'une autre consonne), ils sont proches formellement directement. *Advantage* est donc transparent directement par rapport au français.

- En face du nom NL *kapitein*, le dictionnaire bilingue donne pour équivalents possibles les noms français « capitaine » et « commandant ». Les deux équivalents sont proches sémantiquement directement du mot étranger, et l'un, « capitaine », en est proche formellement directement (les graphèmes <k> et <c> notent tous deux le phonème /k/). *Kapitein* est donc transparent directement par rapport au français.

- En face du nom EN *summer*, le dictionnaire bilingue donne pour seul équivalent le nom français « été ». Ils sont proches sémantiquement directement mais ils ne sont pas proches formellement directement (ils n'ont pas la même trame consonantique ni ne commencent par les quatre mêmes premières lettres). *Summer* n'est donc pas transparent directement par rapport au français.

- En face du nom DE *Gewinn*, le dictionnaire bilingue donne pour équivalents possibles les noms français « bénéfice », « profit », « gain » et « lucre ». Ils sont tous proches sémantiquement directement du mot étranger, mais aucun d'eux n'en est proche formellement. *Gewinn* n'est donc pas transparent directement par rapport au français.

L'organisation du corpus d'une part et des notices des dictionnaires d'autre part appelle quelques dernières remarques.

Les indications de fréquence du corpus sont données pour des lemmes, présentés avec leur catégorie grammaticale. Nous savons ainsi à quelle entrée chercher pour les formes pluricatégorielles. En revanche, nous ne savons pas à quelle signification l'indication de fréquence correspond pour les cas d'homonymie et de polysémie. Comme

le signale justement T. Szende dans son article « L'information sémantique en lexicographie bilingue (hongrois-français) » :

Si la fréquence de tel ou tel mot peut être facilement vérifiée à l'aide de logiciels aujourd'hui courants, il n'existe pas de données précises concernant la fréquence d'emploi des différentes valeurs et acceptions des mots. (Szende 2000 : 70)

Les trois dictionnaires bilingues que nous utilisons observent la même organisation : ils présentent les mots qu'ils considèrent comme homonymes dans des notices différentes et les différentes acceptions d'un mot qu'ils considèrent comme polysémique dans des subdivisions d'une même notice. Par exemple, le *Robert & Collins* établit une distinction entre ces trois noms :

bow¹ : arc

bow² : salut

bow³ : proue

et sépare les différentes acceptions de *bowl* :

bowl : **1.** (= *container*) (*gen*) bol ; (*larger*) saladier, jatte ; (*for water*) cuvette ; (*for fruit*) coupe ; [*of beggar*] sébile ; (*US Sport*) championnat, coupe. **2.** [*of wineglass*] coupe ; [*of pipe*] fourneau ; [*of spoon*] creux ; [*of lamp*] globe ; [*of lavatory, sink*] cuvette. **3.** (*Geog*) bassin, cuvette

De même, le *Langenscheidt Großwörterbuch* distingue la présentation des homonymes :

Bank¹ : banc, banquette

Bank² : banque

de celle du polysème :

Bau : **1.** (*Hauses etc.*) construction. **2.** (*Dramas, Atoms, Körpers etc.*) structure ; (*Körpers etc. auch*) constitution ; (*Lebewesens auch*) organisation ; (*Musikinstrumentes*) facture. **3.** (*Baustelle*) chantier

Enfin, le *Robert & Van Dale* distingue lui aussi la présentation de :

club¹ : club (association)

club² : club (fauteuil)

de celle de :

complex : **1.** ensemble, complexe. **2.** (psycho.) complexe

Nous suivrons donc la même organisation que les dictionnaires, et considérerons comme des mots différents les mots qui ont des notices différentes et comme des acceptions d'un même mot les subdivisions d'une même notice (par exemple, le nom allemand *Bank* sera analysé deux fois tandis que *Bau* ne le sera qu'une fois). Ainsi, il suffira qu'un des équivalents proposés dans une de ces subdivisions soit proche formellement avec le mot étranger pour que le mot soit considéré comme transparent. Notons que nous ne prendrons pas en compte les équivalents donnés pour des acceptions qui sont précédées des mentions « vieilli » ou « archaïque », puisque ces acceptions n'existent plus dans la langue actuelle, et que nous tenons à déterminer quels sont les mots qui relèvent de la transparence en synchronie actuelle.

Il est toutefois noté dans l'introduction du *Robert & Van Dale* que « les mots ayant la même orthographe, mais appartenant à des classes grammaticales différentes constituent des entrées séparées » et qu'il « en est de même si, à l'intérieur d'une même classe grammaticale, un mot appartient à plusieurs sous-catégories » (*Robert & Van Dale* 2007 : IX-X). Nous gardons tout de même le même principe de découpage en homonymes et polysèmes pour ce dictionnaire, car sa consultation laisse apparaître que des critères d'ordre sémantique jouent un rôle au moins aussi important pour le découpage des entrées que le critère syntaxique donné. Ainsi, nous trouvons deux entrées *legering¹* (« campement ») et *legering²* (« alliage »), qui sont pourtant tous deux des substantifs féminins précédés de l'article *de*, et ayant un pluriel en *-en* ; et il n'y a, inversement, qu'une seule entrée pour le verbe *heten*, qui est pourtant soit transitif soit intransitif.

Enfin, il faut rappeler que cette étude est une étude hors texte et que nous cherchons des équivalences strictement mot à mot (la prise en compte d'un contexte, si minime soit-il, donnant un *sens* au mot considéré, alors que nous nous intéressons dans cette partie de

notre travail aux *significations* des mots). Nous examinerons les mots donnés dans le corpus par le biais des entrées correspondantes dans les dictionnaires bilingues, et ne nous intéresserons, pour l'observation des équivalents proposés, ni aux locutions et exemples d'emploi, ni, pour la recherche des mots du corpus, aux formes composées des verbes. Les mots pour lesquels ne sont pas proposés d'équivalents stricts mais qui sont seulement traduits au sein de locutions seront ainsi considérés comme opaques, puisqu'il n'existe pas d'équivalence hors texte entre deux mots considérés isolément. Si toutefois un équivalent proposé est une lexie composée où un seul morphème est lexical et que ce morphème est proche formellement directement du mot étranger, alors ce dernier sera considéré comme transparent directement. Par exemple, le verbe DE *liegen* est considéré comme transparent directement par rapport au verbe FR « être allongé », ainsi que le verbe EN *join* par rapport au verbe FR « se joindre à ».

Les tableaux de présentation de l'analyse, pour chaque langue, sont organisés en 4 colonnes : apparaissent successivement la fréquence du mot étranger considéré, sa forme lemmatisée, sa catégorie grammaticale et l'équivalent français transparent directement avec le mot étranger. Les mots sont rangés par ordre de fréquence décroissante.

Insistons sur le fait que les mots étrangers présents dans les tableaux ne seront pas assurément compris par des francophones. Il s'agit uniquement des mots qui respectent les règles de la transparence directe. De même, certains équivalents proposés pourront paraître surprenants ; mais s'ils sont proposés par le dictionnaire bilingue et qu'ils sont proches formellement directement du mot étranger, alors c'est par ce biais que le mot étranger sera considéré comme transparent directement.

Les tableaux d'analyse des 1 000 mots de chaque langue apparaissent en annexes. Nous présentons ici une description détaillée des dix premières lignes d'un des tableaux⁸⁸, celui de l'anglais (la méthodologie étant la même pour toutes les langues), pour expliquer comment nous avons les avons constitués et comment il faut les comprendre.

Nous présenterons ensuite, dans chaque section, les statistiques tirées de l'étude des tableaux, et établirons en fin de chapitre un bilan comparatif des données recueillies.

⁸⁸ Nous avons ajouté ici une numérotation des lignes pour les besoins de l'explication. Cette numérotation n'apparaît pas dans les tableaux des annexes.

	Fréquence	Lemme	Catégorie grammaticale	Équivalent français
1.	69206	will	V	vouloir
2.	41842	make	V	marquer
3.	40954	so	ADV	si
4.	38013	know	V	connaître
5.	37195	one	N	un
6.	26543	people	N	peuple
7.	24308	just	ADV	juste
8.	19158	new	A	nouveau
9.	18686	too	ADV	trop
10.	16496	also	ADV	aussi

Les indications de fréquence révèlent combien de fois les mots ont été relevés lors de la constitution du corpus utilisé dans CELEX Database. Par exemple, dans le corpus anglais constitué de 17,9 millions de mots, le verbe *will* a été relevé 69 206 fois.

Les paragraphes suivants présentent la façon dont nous avons analysé chacun de ces dix mots, qui sont les dix premiers mots transparents directement de l'anglais par rapport au français. Les numéros des paragraphes indiquent à quelle ligne du tableau nous nous référons.

1. Le *Robert & Collins*, dans la notice de *will*, propose majoritairement des traductions d'exemples d'emploi. Néanmoins, un équivalent hors texte est proposé : « vouloir ». Puisque le verbe français est considéré dans le dictionnaire bilingue comme un équivalent possible du mot anglais, les deux mots sont proches sémantiquement directement. De plus, ils sont proches formellement directement car ils ont la même trame consonantique : les graphèmes <w> et <v>, bien que différents, ne rompent pas la trame car ils peuvent tous deux noter, en français, le même phonème, /v/ ; on trouve ensuite les graphèmes <ll> et <l>, qui notent le même phonème, /l/ ; enfin, il y a suppression d'une consonne finale, le /ʁ/, en anglais par rapport au français, mais puisqu'il y a déjà deux consonnes équivalentes entre les deux mots (<w>/<v> et <ll>/<l>), cette suppression ne remet pas en cause la trame consonantique. Ainsi, puisque le verbe *will* est proche

formellement directement d'un équivalent (dont il est donc proche sémantiquement directement), il est considéré comme transparent directement par rapport au français.

2. De nombreux équivalents sont proposés en face du verbe *make* : « faire », « construire », « fabriquer », « gagner », « se faire », « réaliser un bénéfice net de », « rapporter », « marquer », « arriver à », « attraper », « avoir », « obliger », « forcer » et « monter ». Tous sont considérés comme proches sémantiquement directement du verbe anglais. L'un d'entre eux, « marquer », en est, de plus, proche formellement directement : les deux mots ont la même première consonne, <m>, puis une seconde consonne équivalente, les graphèmes <k> et <qu> notant le même phonème /k/ ; il y a une consonne médiane en plus en français, le <r>, mais elle est au contact d'une autre consonne ; enfin, le <r> final français ne se prononce pas comme consonne, il ne rompt donc pas l'identité des trames consonantiques. Le verbe *make* est ainsi proche sémantiquement directement et proche formellement directement d'un mot français, « marquer » ; il est donc considéré comme transparent directement.

3. En face de l'adverbe *so* sont proposés les équivalents suivants : « si », « tellement », « aussi », « ainsi », « comme ceci », « comme cela », « de cette façon ». Le premier équivalent, « si », est proche formellement directement du mot anglais : ils ont la même trame consonantique, qui se résume ici au <s>, et ils sont proches sémantiquement directement puisque l'un est proposé comme équivalent de l'autre. *So* est donc considéré comme transparent directement. Notons que *so* est également transparent directement par rapport à « aussi » et « ainsi ». Nous avons cependant choisi l'équivalent « si » parce que moins de manipulations formelles sont nécessaires pour passer de *so* à « si » que pour passer de *so* à « aussi » ou « ainsi »

4. En face du verbe *know* sont proposés les équivalents suivants : « connaître », « savoir », « reconnaître ». *Know* est proche formellement directement du premier équivalent, dont il est aussi proche sémantiquement directement : les graphèmes <k> et <c> peuvent tous deux noter en français le phonème /k/ (ils sont donc équivalents) ; on observe ensuite les graphèmes <n> et <nn> qui notent le même phonème, /n/ ; il y a enfin un groupe consonantique de moins à la fin du mot anglais par rapport au français, <tr>, mais puisqu'il y a déjà eu deux consonnes équivalentes entre les deux mots (<k>/<c> et <n>/<nn>), cette suppression ne remet pas en cause la trame consonantique. Parce qu'il est

proche formellement directement et proche sémantiquement directement d'un mot français, « connaître », *know* est considéré comme transparent directement.

5. En face du nom *one*, un seul équivalent est proposé, « un ». Ils sont proches sémantiquement directement et proches formellement directement : ils ont la même trame consonantique, qui se résume ici au <n>. *One* est donc considéré comme transparent directement.

6. En face du nom *people* sont proposés les équivalents suivants : « gens », « personnes », « population », « habitants », « famille », « parents », « peuple », « nation », « race ». Le nom anglais est proche sémantiquement directement et proche formellement directement de l'équivalent « peuple » : ils ont exactement la même trame consonantique. *People* est donc considéré comme transparent directement. Il est également proche formellement directement de « population » (ajout en français de la consonne <t> mais qui est précédée de trois consonnes équivalentes à l'anglais, <p>, <p> et <l> ; et le <n> final ne se prononce pas en tant que consonne), mais nous avons choisi l'équivalent « peuple » qui présente moins de différences formelles.

7. En face de l'adverbe *just* sont proposés les équivalents suivants : « juste », « exactement », « ne...que », « simplement », « spécialement », « absolument », « tout simplement ». *Just* est proche sémantiquement directement et proche formellement directement d'un mot français, « juste », qui est un de ses équivalents dans le dictionnaire bilingue et qui a la même trame consonantique (de plus, les deux mots commencent par les quatre mêmes premières lettres). *Just* est donc considéré comme transparent directement.

8. En face de l'adjectif *new* sont proposés les équivalents suivants : « nouveau », « neuf », « frais », « fraîchement trait », « pas fait ». Le mot anglais est proche sémantiquement directement de ces équivalents, et il est proche formellement directement de l'un d'eux, « nouveau ». Ils ont en effet la même trame consonantique : ils commencent par la même consonne, <n>, et présentent ensuite des consonnes équivalentes, <w> et <v>, parce qu'elles peuvent toutes deux noter en français le phonème /v/. *New* est donc considéré comme transparent directement.

9. En face de l'adverbe *too* sont proposés les équivalents suivants : « trop », « par trop », « aussi », « en plus », « par-dessus le marché », « de plus ». Le premier des équivalents, « trop », est proche formellement directement du mot anglais car ils ont la même trame consonantique : ils commencent par la même consonne, <t> ; la consonne de plus en français, <r>, est au contact d'une autre consonne ; la dernière consonne en français, <p>, ne se prononce pas. *Too* est donc considéré comme transparent directement.

10. En face de l'adverbe *also* sont proposés les équivalents suivants : « aussi », « également », « de plus ». *Also* est proche sémantiquement directement, par définition, de tous ces équivalents, et il est proche formellement directement du mot français « aussi » car les deux mots ont la même trame consonantique : les graphèmes <s> et <ss> notent le même phonème, /s/, et la consonne de plus en anglais, le <l>, est au contact d'une autre consonne. *Also* est donc considéré comme transparent directement.

Nous allons maintenant présenter, dans les sections suivantes, les résultats que nous avons tirés de l'analyse des tableaux de la transparence directe.

Ces résultats sont évidemment valables uniquement pour le corpus utilisé et en fonction des règles avec lesquelles il a été analysé. Les hypothèses que nous pourrions avancer après observation des résultats sont donc également dépendantes du corpus et de son mode d'analyse.

3.2.1. Détermination de l'interlexique anglais/français

Le tableau suivant donne le nombre de mots du corpus anglais transparents directement par rapport au français, de façon globale puis pour chaque catégorie grammaticale, ainsi que la proportion de mots transparents par rapport à l'interlexique anglais/français et par rapport au corpus anglais. Rappelons que le corpus anglais est constitué de 1 000 mots, répartis en 131 adjectifs, 115 adverbes, 496 noms et 258 verbes.

	Absolu	Relatif à l'interlexique	Relatif au corpus
Total	528	100 %	53 %
Adjectifs	77	15 %	8 %
Adverbes	25	5 %	3 %
Noms	310	59 %	31 %
Verbes	116	22 %	12 %

Le premier chiffre intéressant apparaissant grâce à ce tableau est le nombre total de mots transparents directement en anglais : 528 sur 1 000, soit 53 %. Plus de la moitié des 1 000 mots lexicaux les plus fréquents de l'anglais répondent donc aux règles de la transparence directe. Un tel taux confirme voire dépasse ce que l'on sait sur la relation qui unit l'anglais au français : comme H. Walter le rappelle dans son ouvrage *Honni soit qui mal y pense : l'incroyable histoire d'amour entre le français et l'anglais*, les deux langues ont connu tellement d'échanges au cours de leur histoire qu'elles possèdent un grand nombre d'éléments lexicaux en commun. Nous pouvons maintenant chiffrer ce nombre, du moins pour les 1 000 premiers mots lexicaux et en fonction de nos règles d'analyse.

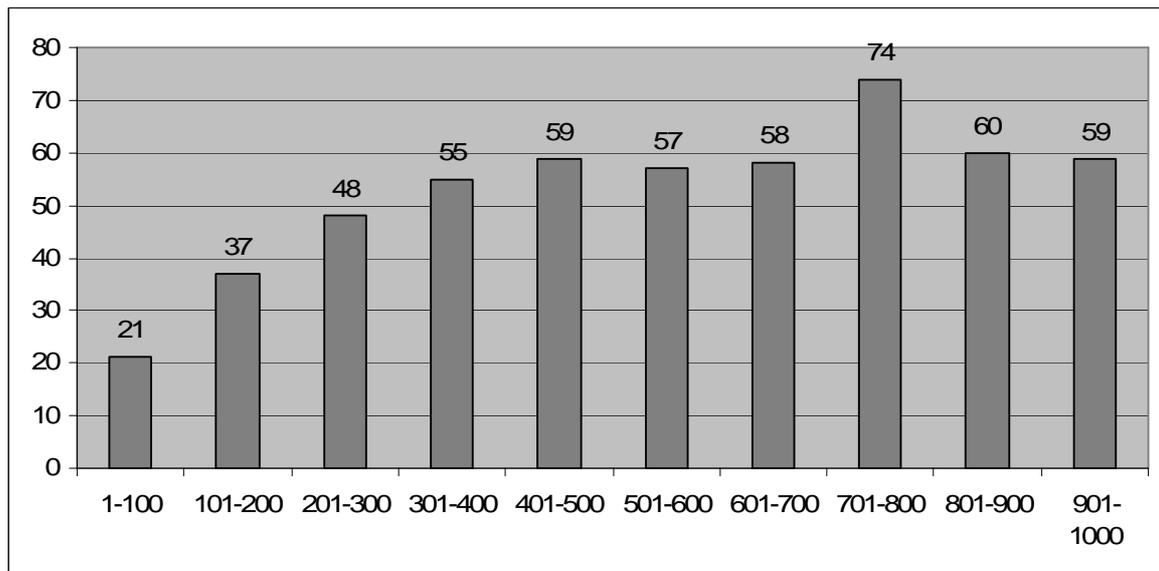
L'autre information que nous retenons de ces calculs est la répartition des catégories grammaticales par rapport à la transparence directe. Parmi les mots transparents directement, il y a 59 % de noms, soit une très large majorité. Arrivent derrière les verbes avec 22 %, les adjectifs avec 15 %, et enfin les adverbes avec seulement 5 %. Les noms transparents directement représentent ainsi non seulement les 3/5 des mots transparents directement, mais aussi quasiment le tiers de tout le corpus.

Le tableau suivant permet de savoir si certaines catégories sont plus susceptibles de transparence directe que d'autres.

	Absolu	Relatif à la catégorie
Adjectifs	77	59 %
Adverbes	25	22 %
Noms	310	62 %
Verbes	116	45 %

En comparant ce tableau et le précédent, on remarque que les noms sont à la fois les plus nombreux par rapport à tous les mots transparents directement, et la catégorie dont la plus grande proportion des items est transparente directement. En revanche, si les noms étaient suivis des verbes quant au nombre de mots transparents directement, ce sont désormais les adjectifs qui, peu après les noms, sont les plus susceptibles d'être transparents directement. Notons enfin que seuls 22 % des adverbes sont transparents directement. Compte tenu des réserves avancées avant cette section quant à la prudence requise pour faire des hypothèses à partir des résultats, il semble que les noms constituent, dans la mesure où ils peuvent être identifiés comme tels, un appui pour la compréhension par des francophones de textes rédigés en anglais.

Il nous a également semblé intéressant de savoir si le taux de transparence directe était en relation avec la fréquence, notre hypothèse étant que les mots les plus souvent utilisés sont également ceux qui subissent le plus de modifications, donc ceux qui se singularisent le plus par rapport aux autres langues. Nous avons donc calculé le nombre de mots transparents directement, toutes catégories confondues, par tranches de 100 mots (de 1 à 100, de 101 à 200, etc.). Le graphique suivant illustre les résultats (les tranches de mots par fréquence décroissante apparaissent sur l'axe des abscisses, et le nombre de mots transparents directement sur l'axe des ordonnées) :



Les mots les plus fréquents sont visiblement les moins transparents, et le taux de mots transparents augmente de la première à la cinquième tranche. L'hypothèse semble ainsi se vérifier : si 1/5 des 100 premiers mots sont transparents directement, leur proportion atteint presque 60 % des mots de la cinquième tranche. À partir de la sixième en revanche, on n'observe plus d'évolution nette, et le taux de mots transparents directement oscille peu, excepté lors du pic de 74 % de transparence directe dans la tranche 701 à 800.

3.2.2. Détermination de l'interlexique allemand/français

Le tableau suivant donne le nombre de mots du corpus allemand transparents directement par rapport au français, de façon globale puis pour chaque catégorie grammaticale, ainsi que la proportion de mots transparents par rapport à l'interlexique allemand/français et par rapport au corpus allemand. Rappelons que le corpus allemand est constitué de 1 000 mots, répartis en 173 adjectifs, 119 adverbes, 472 noms et 236 verbes.

	Absolu	Relatif à l'interlexique	Relatif au corpus
Total	222	100 %	22 %
Adjectifs	44	20 %	4 %
Adverbes	7	3 %	1 %
Noms	150	68 %	15 %
Verbes	21	9 %	2 %

Bien que la quantité de mots transparents directement par rapport au français soit inférieure en allemand qu'en anglais, il faut noter qu'elle constitue néanmoins un peu plus d'1/5 du corpus allemand. Parmi ces mots transparents directement, plus des deux tiers sont des noms. Ils représentent ainsi environ 1/6 du corpus allemand total.

Si les noms transparents directement représentent 15 % du corpus allemand, le corpus est en revanche constitué d'une quantité très faible de mots transparents directement appartenant à d'autres catégories. Les adjectifs, les verbes et les adverbes transparents directement ne représentent en effet respectivement que 4 %, 2 % et 1 % du corpus.

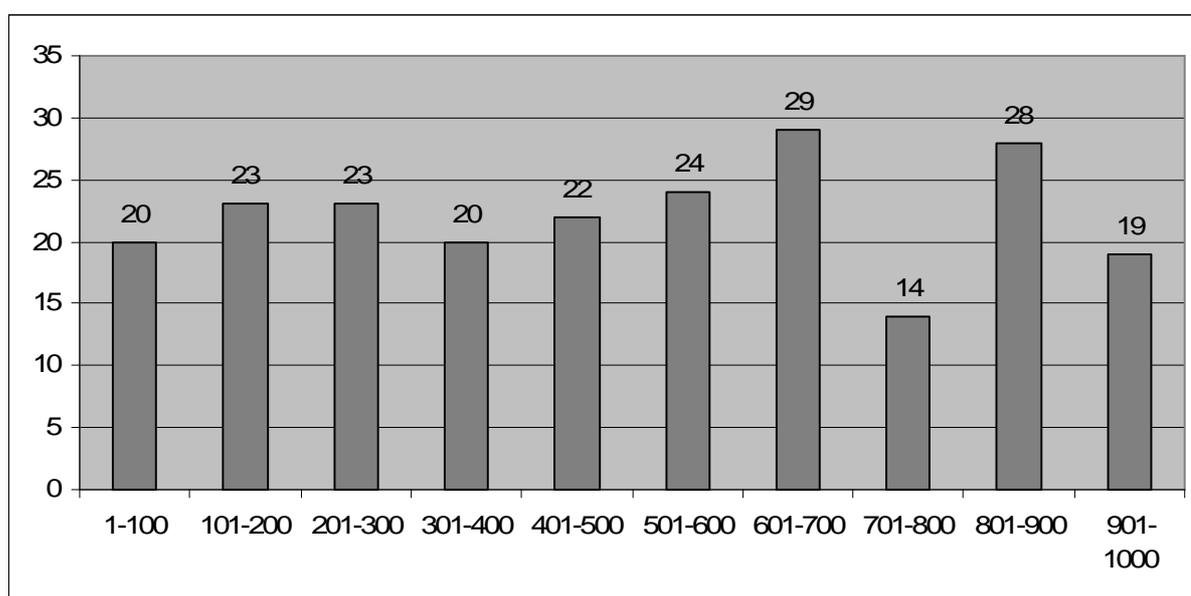
Le tableau suivant présente la proportion de mots transparents directement au sein de chaque catégorie :

	Absolu	Relatif à la catégorie
Adjectifs	44	25 %
Adverbes	7	6 %
Noms	150	32 %
Verbes	21	9 %

Nous avons vu dans le tableau précédent que les catégories qui constituent la transparence directe en allemand sont, par fréquence décroissante, les noms, les adjectifs, les verbes et les adverbes. Nous observons ici que la proportion de mots transparents directement au sein de chaque catégorie suit le même ordre : les noms du corpus allemand sont proportionnellement les plus transparents directement, suivis des adjectifs, des verbes

et des adverbes. Ainsi, non seulement seuls 2 % des mots transparents directement sont des verbes, mais en plus seuls 9 % de ces derniers sont transparents directement. Il semble donc que les verbes, qui sont des éléments fondamentaux pour la construction des phrases, puissent constituer une difficulté majeure pour des lecteurs francophones débutant en allemand, et demandent à être compris, quand ils sont rencontrés pour la première fois, par le processus de l'inférence.

La proportion de mots transparents directement par tranches de 100 mots est illustrée dans le graphique suivant :



Nous n'observons pas ici de tendance forte. Si le taux de mots transparents directement augmente de la première à la deuxième tranche et, plus généralement, de la première à la septième tranche (de 20 % à 29 %), il ne subit pas moins des fluctuations pour les tranches intermédiaires. L'hypothèse selon laquelle les mots les plus fréquents sont également les plus opaques ne se vérifie donc que partiellement. Notons également le chiffre de la tranche 701 à 800, où la transparence directe atteint son niveau le plus bas (14 %).

3.2.3. Détermination de l'interlexique néerlandais/français

Le tableau suivant donne le nombre de mots du corpus néerlandais transparents directement par rapport au français, de façon globale puis pour chaque catégorie grammaticale, ainsi que la proportion de mots transparents par rapport à l'interlexique néerlandais/français et par rapport au corpus néerlandais. Rappelons que le corpus néerlandais est constitué de 1 000 mots, répartis en 172 adjectifs, 117 adverbes, 398 noms et 313 verbes.

	Absolu	Relatif à l'interlexique	Relatif au corpus
Total	208	100 %	21 %
Adjectifs	40	19 %	4 %
Adverbes	9	4 %	1 %
Noms	133	64 %	13 %
Verbes	26	13 %	3 %

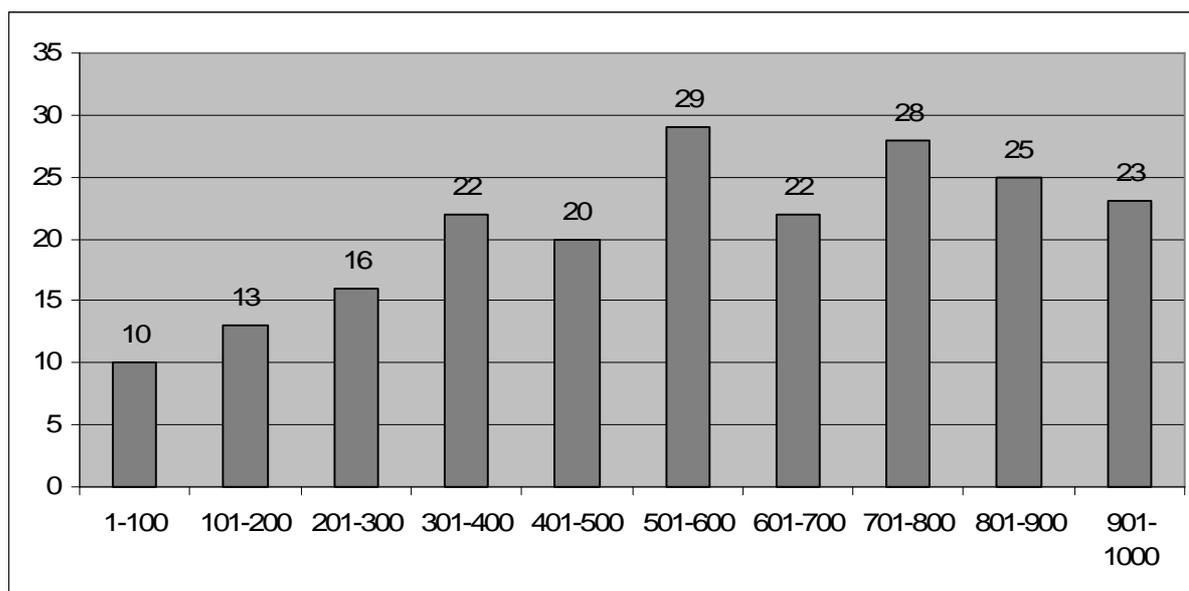
Deux informations principales se dégagent. D'abord, 1/5 des 1 000 mots lexicaux les plus fréquents du néerlandais sont transparents directement par rapport au français. Il semble ainsi que des apprenants francophones qui tenteraient de comprendre des textes rédigés dans cette langue germanique pourraient bénéficier lors de leur processus de compréhension d'un nombre non négligeable de mots identiques ou très proches de mots de leur langue. Ensuite, on note que les noms transparents directement représentent presque les deux tiers des mots transparents directement, et presque le huitième du corpus néerlandais total.

Le tableau suivant présente le pourcentage de transparence directe au sein de chaque catégorie grammaticale :

	Absolu	Relatif à la catégorie
Adjectifs	40	23 %
Adverbes	9	8 %
Noms	133	33 %
Verbes	26	8 %

Proportionnellement, ce sont également les noms qui sont les plus transparents, suivis des adjectifs. Arrivent ensuite les verbes et adverbes, dont 1/12 sont transparents directement.

Le nombre de mots transparents directement en fonction de la fréquence apparaît dans le graphique suivant :



La progression est ici nettement visible : le nombre de mots transparents directement augmente de la première à la quatrième tranche (de 10 % à 22 %), et connaît ensuite les valeurs les plus élevées dans les sixième (29 %) et huitième (28 %) tranches. L'hypothèse que nous avons énoncée et selon laquelle les mots les plus fréquents sont les moins transparents se vérifie donc pour le néerlandais.

3.2.4. Bilan comparatif des interlexiques

Le tableau suivant présente de manière synthétique les résultats chiffrés de la transparence directe des trois langues germaniques par rapport au français :

	ANGLAIS			ALLEMAND			NEERLANDAIS		
	Absolu	Relatif à l'interlexique	Relatif au corpus	Absolu	Relatif à l'interlexique	Relatif au corpus	Absolu	Relatif à l'interlexique	Relatif au corpus
Total	528	100 %	53 %	222	100 %	22 %	208	100 %	21 %
Adjectifs	77	15 %	8 %	44	20 %	4 %	40	19 %	4 %
Adverbes	25	5 %	3 %	7	3 %	1 %	9	4 %	1 %
Noms	310	59 %	31 %	150	68 %	15 %	133	64 %	13 %
Verbes	116	22 %	12 %	21	9 %	2 %	26	13 %	3 %

Il apparaît que l'anglais est de loin la langue qui possède le plus de mots transparents directement (environ deux fois et demie de plus qu'en allemand et en néerlandais). Le taux de transparence directe est relativement important en allemand et en néerlandais (1/5 environ de chacun de ces deux corpus est constitué de mots transparents directement), et il dépasse les 50 % en anglais.

De plus, ce tableau récapitulatif permet de voir à quel point l'allemand et le néerlandais sont similaires. Cette similarité est la plus remarquable par les valeurs en pourcentage de leurs taux de transparence directe, globale et par catégories, relativement aux corpus.

Les trois langues sont semblables par le fait que ce sont les noms qui constituent nettement la part la plus importante des mots transparents directement, et ce particulièrement en allemand où la proportion dépasse les deux tiers. La différence principale concerne les verbes : ils représentent une part bien moins importante en allemand et en néerlandais (respectivement 2 % et 3 % des corpus) qu'en anglais (12 % du corpus).

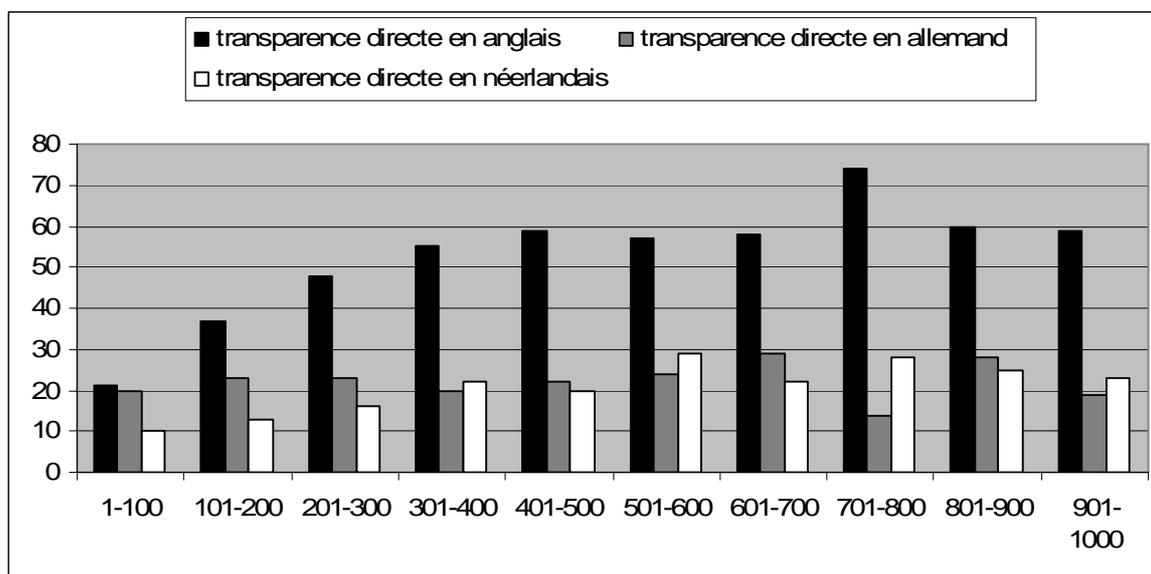
Le tableau suivant rappelle le pourcentage de mots transparents directement au sein de chaque catégorie grammaticale, dans les trois langues germaniques :

	ANGLAIS		ALLEMAND		NEERLANDAIS	
	Absolu	Relatif à la catégorie	Absolu	Relatif à la catégorie	Absolu	Relatif à la catégorie
Adjectifs	77	59 %	44	25 %	40	23 %
Adverbes	25	22 %	7	6 %	9	8 %
Noms	310	62 %	150	32 %	133	33 %
Verbes	116	45 %	21	9 %	26	8 %

Nous constatons que les catégories ne sont pas également transparentes dans les différentes langues. Si les noms sont toujours les plus transparents directement, ils le sont environ deux fois plus en anglais que dans les autres langues. La catégorie suivante la plus transparente, dans les trois langues, est celle des adjectifs. Viennent ensuite les verbes, qui constituent la différence la plus importante entre l'anglais d'une part et l'allemand et le néerlandais de l'autre. Environ la moitié d'entre eux sont transparents directement en anglais, tandis qu'à peine 1/10 le sont dans les deux autres langues. Les adverbes sont à la fois les représentants les plus faibles des mots transparents dans les trois langues et la catégorie dont le moins grand nombre d'items est transparent directement.

Ces résultats, qui sont, rappelons-le, dépendants du corpus utilisé et des règles avec lesquelles il a été analysé, laissent apparaître deux informations principales. D'une part, l'enseignement des langues germaniques en France aurait beaucoup à gagner en prenant en compte le fait qu'une partie non négligeable de leur lexique est proche de mots français. Que le taux de mots transparents directement soit d'1/5, comme c'est le cas en allemand et en néerlandais, ou d'1/2 comme c'est le cas en anglais, il est suffisamment important pour qu'on enseigne aux apprenants à chercher le sens par eux-mêmes de textes qu'ils lisent dans ces langues. Sans même avoir pris de cours, on pourrait accéder au sens d'une partie importante des mots lexicaux, qui sont précisément ceux qui portent majoritairement la charge sémantique. D'autre part, il semble que lorsqu'on lit dans ces trois langues, on doive en premier lieu essayer de fixer notre attention sur les noms, dans la mesure où on est capable les repérer en tant que tels, qui sont la source majoritaire de transparence directe.

Le graphique suivant présente de façon synthétique l'évolution du taux de transparence directe en fonction de la fréquence pour les trois langues :



La supériorité de l'anglais quant au nombre de mots transparents directement est également visible sur ce graphique. Notons toutefois qu'il y a quasiment autant de mots transparents directement dans les cent mots les plus fréquents de l'anglais et de l'allemand. Dans la suite du graphique, la relation est plus ou moins constante : l'anglais dépasse de beaucoup l'allemand et le néerlandais, qui ont des valeurs comparables. Le néerlandais et l'anglais s'opposent néanmoins à l'allemand pour ce qui est de l'évolution : une progression au début pour atteindre un palier et garder une constance relative d'un côté, et pas de tendance évolutive notable de l'autre. Notons également que l'écart principal du pourcentage de mots transparents directement en anglais et en allemand est observable dans la même tranche, la tranche 701-800, vers le haut en anglais et vers le bas en allemand. Nous ne pouvons attribuer ce fait qu'au hasard, mais peut-être demanderait-il à être mis en relation avec d'autres données.

Nous voulions, dans ce chapitre, appliquer les règles de la transparence directe à notre corpus. Pour cela, nous avons élaboré une méthodologie qui nous permettait de transposer à une analyse hors texte les principes de transparence directe pensée en texte que nous avons formulés dans la deuxième partie de ce travail.

Les résultats obtenus laissent apparaître que l'anglais est plus transparent directement par rapport au français que les deux autres langues germaniques, et que

l'allemand et le néerlandais ont un comportement tout à fait similaire face à la transparence directe.

La méthodologie employée ici ne nécessitait que l'utilisation de dictionnaires bilingues, puisqu'il s'agissait pour nous d'examiner les équivalents proposés par ces dictionnaires en face des mots étrangers, et de regarder si les équivalents étaient proches formellement directement des mots étrangers. Dans le chapitre suivant, consacré aux hyperlexiques des langues considérées, nous aurons à développer une méthodologie d'analyse nous permettant de comparer les équivalents français proposés par les dictionnaires bilingues à d'autres mots français, lesquels jouent le rôle de mots-relais, proches formellement des mots étrangers et proches sémantiquement des équivalents français.

3.3. L'hyperlexique

Nous allons dans ce chapitre calculer le nombre de mots anglais, allemands et néerlandais transparents par rapport au français. L'ensemble des mots transparents entre deux langues (anglais-français, allemand-français, néerlandais-français) rendra compte de l'hyperlexique de ces deux langues. Nous avons calculé dans le chapitre précédent le nombre de mots transparents directement entre chacune de ces paires de langues. Pour déterminer leur hyperlexique, nous devons maintenant calculer le nombre de mots transparents directement *et* indirectement. Pour ce faire, nous appliquerons les principes de la proximité formelle directe et indirecte et de la proximité sémantique directe et indirecte à notre corpus.

Exposons quelle sera la méthodologie pour connaître le nombre de mots des trois langues germaniques transparents indirectement pour un francophone.

La remarque faite au début du chapitre sur l'interlexique quant à la primauté de la forme (p. 189) reste tout aussi valable pour ce chapitre, mais nous nous voyons toujours dans l'obligation de comparer les mots par leur face significative, grâce à des dictionnaires bilingues.

Pour analyser le corpus, nous nous intéresserons dorénavant à tous les mots qui n'ont pas été considérés comme transparents directement. Nous les chercherons dans les mêmes dictionnaires bilingues que ceux qui ont permis de vérifier la transparence directe. En face de chaque mot sont proposés des équivalents français, qui sont donc tous proches sémantiquement directement du mot étranger. Si l'un des équivalents proposés est proche formellement indirectement du mot étranger, alors ce dernier est transparent indirectement.

La méthode pour évaluer la proximité sémantique indirecte est plus complexe. Si aucun des équivalents proposés n'est proche formellement du mot étranger, il est toutefois possible que l'un de ces équivalents soit proche sémantiquement indirectement d'un autre mot français, qui, lui, serait proche formellement (directement ou indirectement) du mot étranger. Pour chaque mot qui n'aurait dans ses équivalents aucun mot proche formellement, il nous faudra donc partir à la recherche d'un éventuel mot-relais. Rappelons que le mot-relais et l'équivalent sont proches sémantiquement indirectement par l'intermédiaire d'au moins un sème de transfert.

Pour identifier une proximité sémantique indirecte, nous avons donc besoin des sèmes des mots. Pour avoir accès à ces sèmes – de mots français rappelons-le, puisque l’ajustement sémantique se pratique en langue maternelle – nous avons fait le choix de travailler sur des définitions de dictionnaire. La définition du dictionnaire monolingue sera donc considérée comme une suite de sèmes. Ce parti pris théorique peut, dans un premier temps, surprendre. Les définitions lexicographiques sont en effet de courts textes, rédigés avec des mots (on parle d’ailleurs de la circularité des définitions de dictionnaire, dans lesquelles on définit des mots avec d’autres mots, qui sont à leur tour définis ailleurs dans le dictionnaire), et non des analyses sémiques. Toutefois, les lexicologues travaillant sur la dictionnaire sont d’accord pour attribuer aux éléments d’information des définitions un statut équivalent à celui de sèmes du mot défini :

La lexicographie s’attache à décrire les composantes du lexique. Ce sont les vocables et leurs relations (entre eux). L’analyse lexicographique est constituée d’une analyse sémique qui décompose les vocables en traits sémantiques ou éléments définitionnels et d’une analyse syntagmatique qui enregistre les possibilités combinatoires de chaque vocable. (Loffler-Laurian 2000 : 137)

On sait en effet que le sens lexical a un caractère décomposable et différentiel. La meilleure façon de définir un mot c’est de répertorier les sèmes (unités de sens réduites) que l’on estime indispensables pour sa description et, il est clair également que chaque sème (collection de sèmes) se distingue de tout autre par au moins un sème. (Szende 2000 : 73)

Ces affirmations s’expliquent : un dictionnaire de mots se distingue d’un dictionnaire encyclopédique en ce que le premier informe « sur des **mots** et le second sur la **connaissance du monde** » (Le Robert 2002 : x). On trouve donc dans les notices du dictionnaire de mots des informations d’ordres morphologique, syntaxique, étymologique et sémantique. Pour définir le mot et non la chose, la définition du dictionnaire de mots donne les informations nécessaires au bon emploi du mot défini, et non toutes les informations connues sur la chose. Par exemple, le *Petit Robert* donne après l’entrée « lion » la définition suivante : « Grand mammifère carnivore, grand félin à pelage fauve, à crinière brune et fournie, à queue terminée par une grosse touffe de poils, vivant en Afrique et en Asie » ; tandis que l’*Encyclopædia Universalis* donne la définition suivante :

« Le lion (*Panthera leo*) est caractérisé par un corps puissant, une tête très large et couronnée d'une crinière foncée, de fortes pattes et une longue queue se terminant par un toupet de poils noirs. Comme la plupart des félidés, ses griffes sont rétractiles. Son pelage ras est de couleur ocre. Le mâle mesure de 2,6 à 3,30 mètres de longueur (queue de 60 cm à 1 mètre non comprise), 1,20 mètre de hauteur et son poids varie de 150 à 250 kilogrammes. La femelle, la lionne, est nettement plus petite. L'espérance de vie est de quinze à vingt ans [etc.] ».

La liste des traits suffisant à différencier un mot d'un autre (tels que ceux que donne le dictionnaire de mots) est également une définition valable de « sémème », les traits concernés étant plus ou moins équivalents aux sèmes. La possibilité de postuler cette équivalence est très clairement explicitée par J. Picoche :

*La définition linguistique ne retient que ce qui est utile pour le fonctionnement correct du langage et non ce qui est nécessaire pour la connaissance exhaustive du type de référent auquel renvoie le mot. La **spécificité**, c'est-à-dire l'indication des traits distinctifs, est sa règle principale. Elle met entre parenthèses, à des fins de meilleure intelligibilité, un grand nombre de caractères des référents. Elle se contente de ceux que révèle l'analyse sémique, confirmée par le fait que l'absence de l'un d'eux entraîne les locuteurs à refuser d'utiliser le terme en question, traits pertinents propres à marquer la différence entre deux mots classés sous le même hyperonyme, et qui constituent son essence même. (Picoche 1997 : 140)*

Elle conclut, un peu plus loin : « [la définition linguistique] est donc la forme lexicographique habituelle de l'analyse sémique » (*ibid.* : 141).

Nous nous permettons donc d'attribuer aux différents traits d'une définition linguistique le statut de sèmes du mot étudié. Deux mots français seront ainsi considérés comme proches sémantiquement indirectement lorsqu'ils posséderont au moins un trait en commun dans leur définition respective.

Pour savoir quels sont les mots qui sont proches sémantiquement indirectement d'un autre mot, c'est-à-dire qui appartiennent à son champ sémantique, nous utiliserons la version électronique du *Nouveau Petit Robert 2007*. Ainsi, quand aucun des équivalents proposés dans le dictionnaire bilingue pour un mot étranger ne sera proche formellement de ce dernier, nous chercherons le premier équivalent dans le *Petit Robert* électronique.

Nous faisons le choix de travailler à partir du premier équivalent pour limiter les recherches, et parce que nous faisons l'hypothèse que c'est l'équivalent le plus fréquent qui est donné en premier (cependant, seul le *Langenscheidt Großwörterbuch* donne cette indication : « Les traductions françaises sont rangées de telle façon que les mots et locutions les plus courantes se trouvent au début » (Sachs & Villatte 1990 : xvii)). La recherche dans le *Petit Robert* nous donnera la définition du mot français équivalent au mot étranger. Ensuite, pour savoir quels sont les mots qui appartiennent au champ sémantique de cet équivalent, nous utiliserons une des fonctionnalités de la version électronique du *Petit Robert*, à savoir la « Recherche par critères ». Cette fonctionnalité permet, entre autres, de savoir quelles sont les entrées qui possèdent tel ou tel mot dans leur définition. Il nous sera ainsi permis de taper dans le champ de recherche un trait de la définition de l'équivalent et de savoir quels sont les mots qui partagent ce trait avec lui. Si, dans cette liste de mots, un des mots est proche formellement du mot étranger, alors ce dernier est transparent indirectement par rapport au français (parce que le mot étranger est proche formellement d'un mot français qui, lui, est proche sémantiquement indirectement de l'équivalent).

Avant que la méthodologie ne soit définie plus précisément, il nous semble intéressant de présenter les travaux menés par M.-C. L'Homme et J. Dancette, présenté dans leur article « Modélisation des relations sémantiques dans un dictionnaire spécialisé bilingue » publié en 2001, portant sur les relations sémantiques unissant les mots d'un dictionnaire. Elles ont pris pour outil de travail un dictionnaire bilingue anglais/français spécialisé dans le domaine de la distribution, et ont modélisé les relations sémantiques existant entre les différents mots définis dans ce dictionnaire grâce à leur définition :

Le modèle que nous avons conçu pour la mise en évidence des relations sémantiques est un système de traits. Les traits, dans ce modèle, sont tout simplement des étiquettes qui rendent compte d'une caractéristique sémantique. Le modèle doit permettre l'accès, sur demande de l'utilisateur, à diverses relations sémantiques : hyperonymie, hyponymie, antonymie, parenté sémantique ; et divers types d'équivalence interlinguistique. (L'Homme & Dancette 2001 : 2)

Comme dans notre étude, la recherche d'un trait permet de faire apparaître tous les mots ayant ce trait dans leur définition, c'est-à-dire de faire apparaître ce qu'elles appellent un « réseau conceptuel » (proche de notre champ sémantique) :

Le format que nous proposons, même s'il ne s'agit pas d'un langage formel à proprement parler, repose déjà, par sa conception, sur l'hypothèse selon laquelle, dans un dictionnaire électronique, l'appel des traits sémantiques d'un terme peut générer l'énumération de toute une série de termes corrélés dans une langue ou dans une autre, contribuant ainsi à l'apparition de réseaux conceptuels. (ibid. : 9)

Ainsi, en cherchant les traits « établissement commercial » et « prix réduits », on obtient dix-huit termes, tels que « marché aux puces », « magasin d'usines », « restaurant minute », « club-entrepôt », etc. Huit traits (comme celui renseignant sur la politique de prix) permettent de définir les mots désignant des établissements commerciaux.

Pour constituer le modèle, les auteurs ont dû normaliser les expressions utilisées dans le dictionnaire. Par exemple, elles ont attribué le trait « prix réduit » aux expressions « marges relativement réduites », « politique de bas prix », « prix inférieur au prix demandé dans les magasins », « à bon marché », « solde », etc.

Que ce modèle de recherche de mots ayant en commun un ou plusieurs traits sémantiques repose sur les mêmes principes que notre étude nous conforte dans le choix de notre méthodologie. Cependant, il existe une différence essentielle entre les deux approches, tenant au corpus analysé. M.-C. L'Homme et J. Dancette travaillent sur un dictionnaire très spécialisé, tandis que nous travaillons sur le *Petit Robert*, dictionnaire recensant une très grande partie du lexique français (comme le montre l'indication « 60 000 mots et leurs 300 000 sens » apparaissant sur la couverture du *Petit Robert*). La détermination des traits définitoires n'en est donc que plus complexe. En effet, pour leur modèle, les auteurs cités ont travaillé sur un dictionnaire contenant « environ 350 articles de base faisant apparaître quelque 4 000 termes traités et indexés dans les deux langues » (*ibid.* : 3), et dont les termes appartiennent tous à un domaine très restreint, celui de la distribution. Les traits servant à opposer les mots les uns aux autres sont donc très spécifiques, puisqu'ils opposent des mots au sein d'un domaine très spécialisé. Dans le *Petit Robert* au contraire, les traits permettent de définir les mots de toute la langue, et non d'un domaine précis. Ils sont donc pour certains très généraux. Comme le précise

F. Rastier dans la partie de son ouvrage *Sémantique interprétative* consacrée au nombre de sèmes nécessaires pour décrire une langue, « [l]e seul de point de repère dont nous disposons est purement empirique : c'est le nombre des sèmes du corpus décrit » (Rastier 1987 : 30). Les mots du *Petit Robert* ne sont certes pas tous définis par rapport à l'ensemble de la langue. Par exemple, la signification de « fourchette » est définie par « ustensile de table, dont on se sert pour piquer les aliments » ; elle est donc définie relativement au taxème des ustensiles de table, auquel appartient également « cuillère », dont la signification est définie par « ustensile de table ou de cuisine formé d'un manche et d'une partie creuse, qui sert à transvaser ou à porter à la bouche des aliments liquides ou peu consistants ». En revanche, une des acceptions de « forme » est définie par « être ou objet confusément aperçu et dont on ne peut préciser la nature » : il semble ici difficile de déterminer précisément l'ensemble des êtres ou objets au sein duquel cette signification est définie. Nous ne pourrions donc pas, contrairement à M.-C. L'Homme et J. Dancette, travailler à partir d'un nombre fini et déterminé de traits.

Avant même de sélectionner les traits, nous devons sélectionner la bonne définition du *Petit Robert*. Si l'équivalent indiqué en face d'un mot étranger est susceptible d'avoir des homonymes ou différentes acceptions, les dictionnaires bilingues donnent la plupart du temps une indication de domaine qui nous permet de savoir quelle définition de l'équivalent chercher. Par exemple, en face du nom néerlandais **band**¹ (l'exposant signifiant que le dictionnaire propose plusieurs homonymes, et en l'occurrence qu'il s'agit du premier des homonymes), le *Robert & Van Dale* propose « (mus.) groupe ». Nous ne regarderons donc que la définition du *Petit Robert* « Ensemble de musiciens et de chanteurs appartenant à une même formation », qui situe le mot « groupe » dans le domaine de la musique. S'il n'y a pas d'indication de domaine dans le dictionnaire bilingue et que l'équivalent est défini selon plusieurs acceptions, nous nous aiderons des indications données par le *Petit Robert* au début de chacune des définitions. Ainsi, nous ne prendrons pas en compte les définitions données après l'indication « **VX** », car celle-ci signifie « vieux (mot, sens ou emploi de l'ancienne langue, incompréhensible ou peu compréhensible de nos jours et jamais employé, sauf par effet de style : archaïsme) » (Le Robert 2002 : XXIX), ni celles données après l'indication « **REGION.** », car celle-ci signifie « régional (mot ou emploi particulier au français parlé dans une ou plusieurs régions [France, pays francophones], mais qui n'est pas d'usage général ou qui est senti comme propre à une région) » (*ibid.* : XXVIII). Nous regarderons en priorité les définitions

qui ne sont précédées d'aucune indication (ou parfois de **Cour.**), car ce sont celles-ci qui donnent la signification commune, c'est-à-dire la signification actuelle et la plus fréquente, et ensuite seulement, si besoin est, les définitions marquées ou spécialisées (celles par exemple qui sont précédées de **Techn.**, **Didact.**, etc.).

Précisons maintenant comment nous lancerons nos recherches. Chacune des différentes étapes, numérotées pour plus de lisibilité, va être présentée dans les pages suivantes de façon théorique puis illustrée par un exemple et une capture d'écran du *Petit Robert* électronique exposant la méthode suivie.

1) Les notices du *Petit Robert* sont ainsi construites que les renvois y jouent un rôle prépondérant. Ces renvois ne donnent pas seulement les synonymes du mot défini, mais aussi des mots dont la signification est en relation avec celle du mot défini, c'est-à-dire des mots qui appartiennent à son champ sémantique. Notre premier réflexe sera donc de regarder s'il y a parmi les renvois de la notice de l'équivalent un mot proche formellement du mot étranger. Si c'est le cas, cela signifie qu'il existe un mot proche sémantiquement de l'équivalent et proche formellement du mot étranger ; le mot étranger sera donc considéré comme transparent indirectement.

Par exemple, le nom EN *country* est considéré comme transparent indirectement car il y a dans la notice de son équivalent, « pays », un renvoi dont il est proche formellement, à savoir « contrée » :

The screenshot shows the 'Le Petit Robert de la langue française' software interface. The search bar contains the word 'pays'. The main window displays the entry for 'pays - n. m.'. The entry includes several sections: a list of related terms on the left, a main text area with definitions and examples, and a list of numbered definitions at the bottom. The word 'contrée' is highlighted in a box in the fourth definition.

Le Petit Robert de la langue française

Fichier Édition Recherche Vues Aide

Le nouveau Petit Robert de la langue française 2007

recherche par critères

nomenclature complète

entrées formes locutions composés

orth.

prononcer

plan étymologie renvois et contraires citations exemples et expressions homonymes

pays - n. m.

et de Montaigne. **Prov.** *Nul n'est prophète* en son pays.*
Loc. *Avoir le mal du pays* : être triste loin de son pays.
 → **nostalgie.** *Retourner au pays.*

◆ **LE PAYS DE qqch.** : terre d'élection, milieu particulièrement favorable à, riche en. *La France est le pays du vin. L'Allemagne, pays de la musique.* → **patrie.** *C'est le pays du jazz.*

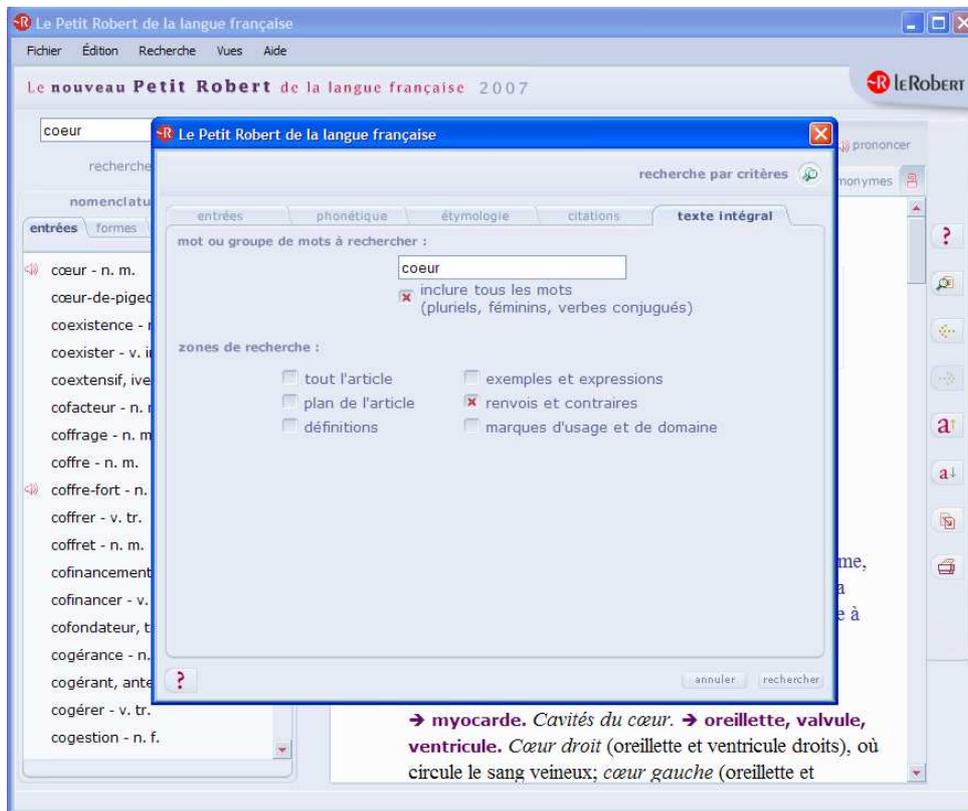
▫ **Fig.** *Le pays du rêve, des songes.* → **domaine, royaume.** « *Alice au pays des merveilles* », œuvre de Lewis Carroll.

4 Région géographique, plus ou moins nettement limitée considérée surtout dans son aspect physique. → **contrée, endroit, 1. lieu, région.** *Les pays chauds, froids, tempérés. Pays plat, de montagnes. Les Pays-Bas. Pays de forêts, de vignes, d'élevage. Voir du pays* : voyager.
 ▫ *Pays de cocagne**.

5 Petite ville; village. *Il habite un petit pays, un pays perdu au fin fond de l'Auvergne.* → fam. **bled, 2. patelin, trou.**

2) S'il n'y a pas de mot proche formellement du mot étranger parmi les renvois de l'équivalent, nous chercherons une occurrence de l'équivalent lui-même dans les renvois de toutes les autres notices du dictionnaire, pour savoir quels sont les mots qui ont l'équivalent dans leur champ sémantique. C'est la fonctionnalité « Recherche sur critères » du *Petit Robert* électronique qui nous permet de lancer une recherche parmi les renvois des différentes notices. S'il y a parmi les résultats de la recherche un mot proche formellement du mot étranger, alors ce dernier fonctionnera comme mot relais et le mot étranger sera considéré comme transparent indirectement.

Par exemple, le nom NL *hart* est considéré comme transparent indirectement. En effet, en cherchant son équivalent, « cœur », parmi les renvois des autres notices du dictionnaire :



on observe parmi les résultats de la recherche qu'il y a un mot, « hardiesse », dont la notice contient « cœur » en tant que renvoi, et qui est proche formellement de *hart* :

Le Petit Robert de la langue française

Fichier Édition Recherche Vues Aide

Le nouveau Petit Robert de la langue française 2007

LeROBERT

coeur

recherche par critères

résultats de la recherche

résultats

- duramen - n. m.
- dureté - n. f.
- entrailles - n. f. pl.
- entraîn - n. m.
- fort, forte - adj. et n. m.
- frisette - n. f.
- généreux, euse - adj.
- générosité - n. f.
- gilet - n. m.
- grâce - n. f.
- hardiesse** - n. f.
- intracardiaque - adj.
- intrigue - n. f.
- mèche - n. f.
- orée - n. f.
- poitrine - n. f.
- raison - n. f.
- sein - n. m.

42 réponses

féminins et pluriels

hardiesse - n. f. prononcer

plan étymologie renvois et contraires citations exemples et expressions homonymes

hardiesse [ˈardjɛs] nom féminin

ÉTYM. 1361; *ardiesce* fin XII^e; de *hardi*

I Littér.

- 1 Qualité d'une personne hardie, de ce qui est hardi.
→ assurance, audace, bravoure, **cœur**, courage, énergie, fermeté, intrépidité. *Avoir, montrer de la hardiesse. Faire preuve de hardiesse. Répondre avec hardiesse.* « Il faut une grande hardiesse pour oser être soi » (Delacroix).
- 2 Pêj. Il a la hardiesse de soutenir cela ! → effronterie, impudence, insolence, témérité.
- 3 Hardiesse du style. → nouveauté, originalité. Une grande hardiesse de pinceau. → vigueur.

3) Si les deux premières recherches s'appuyant sur les renvois s'avèrent non concluantes, nous procédons à d'autres recherches qui mettent en jeu la double nature des traits des définitions. Dans le *Petit Robert* comme dans les autres dictionnaires monolingues, les mots sont définis grâce à d'autres mots, dont nous avons vu qu'ils avaient un statut plus ou moins équivalent à celui des sèmes. Un mot apparaissant dans le *Petit Robert* est ainsi susceptible d'être lu, en différents lieux du dictionnaire, comme un mot à part entière ou comme une indication sur un mot. En tirant parti de cette double nature, nous chercherons ensuite s'il y a un mot proche formellement du mot étranger dans la définition même de l'équivalent. Si c'est le cas, cela signifiera qu'il existe un mot français, avec le statut de sème en l'occurrence, mais qui est aussi un mot défini ailleurs dans le dictionnaire, dont la signification est en relation avec celle de l'équivalent et qui est proche formellement du mot étranger ; le mot étranger sera donc considéré comme transparent indirectement.

Par exemple, l'adjectif DE *besser* est considéré comme transparent indirectement, car il y a dans la définition de son équivalent, « meilleur », un trait sémantique dont il est proche formellement, à savoir « bonté » :

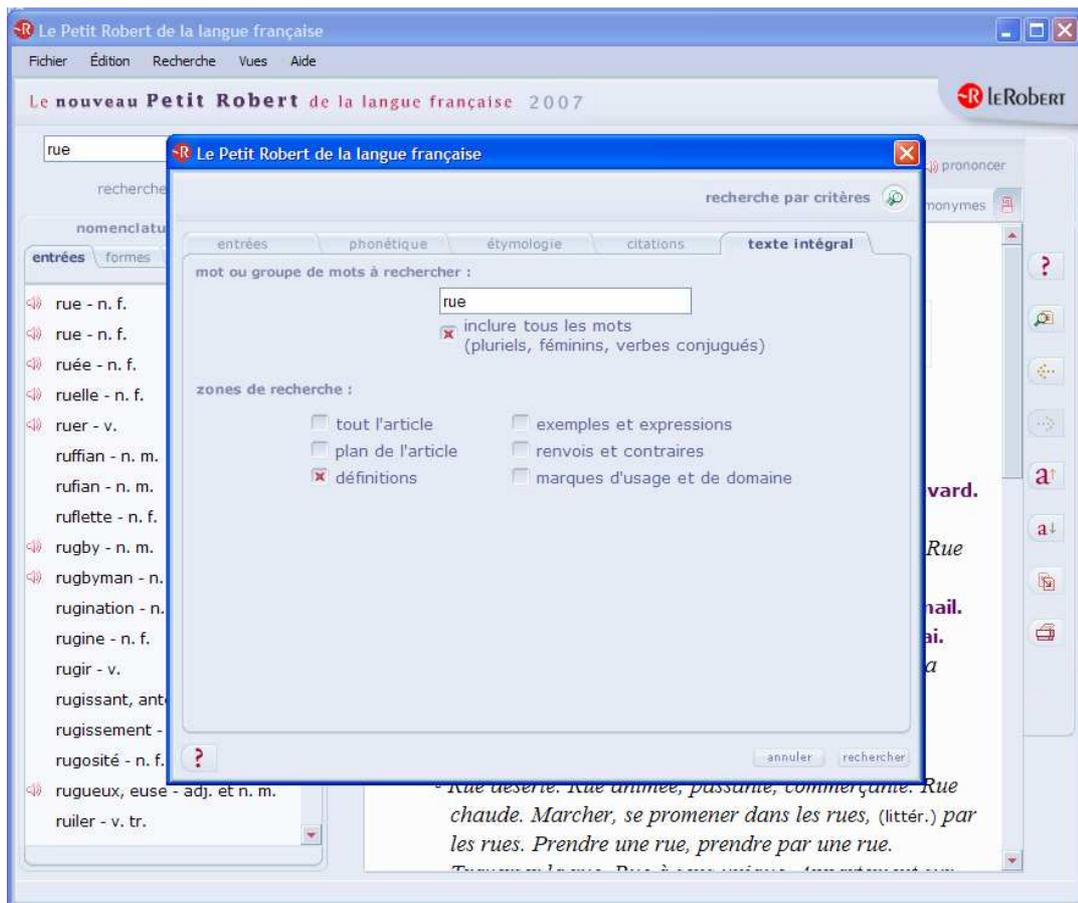
The screenshot shows the 'Le Petit Robert de la langue française' dictionary interface. The search term 'meilleur' is entered in the search bar. The main entry for 'meilleur, e' is displayed, including its phonetic transcription [mɛjœʁ] and classification as an adjective. The etymology section states: 'ÉTYM. début XIII^e; meillor 1080; latin melior, compar. de bonus « bon »'. The definition is divided into two numbered sections:

- 1** Compar. de supériorité de *bon*.
1 Qui l'emporte dans l'ordre de la **bonté**. *Les femmes « sont meilleures ou pires que les hommes » (La Bruyère). Vouloir rendre l'homme meilleur, l'humanité meilleure, l'améliorer. « Il faut toujours être meilleur Que l'homme que l'on voudrait être » (Verlaine).*
- 2** Plus cour. Qui l'emporte dans l'ordre de la qualité, de l'agrément. *Vous avez meilleure mine qu'hier. Meilleure santé ! Je ne connais rien de meilleur. → 2. dessus (au-dessus). C'est meilleur avec du beurre. Être de meilleure humeur. Article de meilleure qualité. → supérieur. Meilleur marché*. De meilleure heure : plus tôt. Il est meilleur peintre que dessinateur.*

The word 'bonté' in the first definition is highlighted with a red box. The interface also shows a sidebar with a list of other words and various navigation icons.

4) De même, nous chercherons ensuite, si les étapes précédentes n'ont pas abouti, les mots dont les définitions contiennent l'équivalent comme trait définitoire. Si, dans cette liste de mots, l'un est proche formellement du mot étranger, alors ce dernier fonctionnera comme mot relais et le mot étranger sera considéré comme transparent indirectement.

Par exemple, le nom EN *street* est considéré comme transparent indirectement car, en cherchant son équivalent, « rue », dans les autres définitions du dictionnaire :



on observe qu'il constitue un trait sémantique de la définition d'un mot qui est proche formellement de *street*, à savoir « trottoir » :

The screenshot shows the interface of the 'Le Petit Robert de la langue française' website. At the top, the title 'Le nouveau Petit Robert de la langue française 2007' is visible. A search bar contains the word 'rue'. Below the search bar, a list of search results is shown, with 'trottoir - n. m.' highlighted. The main content area displays the entry for 'trottoir' [TRÔTWAR] nom masculin. The etymology is given as 'ÉTYM. 1577; de trotter'. The entry includes four numbered definitions:

- 1 vx** Piste où l'on fait trotter les chevaux.
- 2 (XVIIe s.) Anciennt** Chemin élevé le long des quais et des ponts. → **allée**. « *Les trottoirs garnis d'arbres, qui courent le long du Rhône* » (Stendhal).
- 3 (1835) Mod.** Chemin surélevé réservé à la circulation des piétons (sur les côtés d'une **rue**) → **accotement, banquette**. *Se promener sur les trottoirs. « l'absence complète de trottoir étonnait; la chaussée [...] arrivait au ras des maisons »* (Robbe-Grillet). **Fam.** *Radio*-trottoir*.
◆ **Loc.** (1852) *Faire le trottoir* : se prostituer, racoler les passants.
- 4 (1900) Trottoir roulant** : plateforme qui roule sur des rails ou des catènes et sert à transporter des personnes ou des

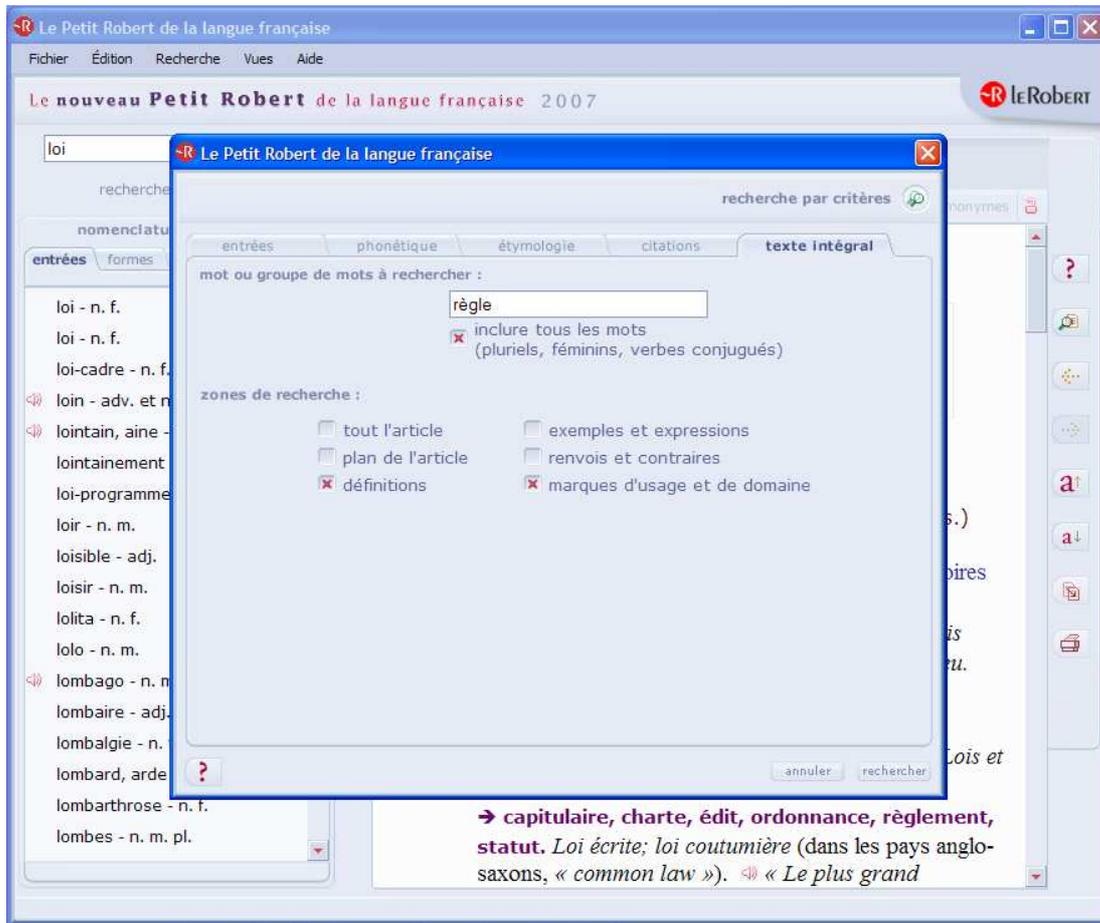
5) Si aucune des étapes citées ci-dessus ne s'avère concluante, nous lancerons des recherches par traits sémantiques. Nous considérons l'indication de domaine comme premier trait sémantique d'une définition (d'ailleurs tout à fait comparable au sème de domaine selon F. Rastier), si indication de domaine il y a. Ensuite, sont considérés comme traits certains mots lexicaux. Cependant, chacun des mots lexicaux ne constitue pas un trait. En effet, comme nous l'avons déjà souligné au début de ce chapitre, une définition de dictionnaire se différencie d'une analyse sémique en ce qu'elle est « une phrase incomplète mais rédigée, un syntagme conforme aux règles syntaxiques de la langue » (Picoche 1997 : 142). Les traits y sont donc reliés entre eux par des mots grammaticaux d'une part, et par des mots lexicaux de signification générale d'autre part. Le nombre de définitions du *Petit Robert* étant trop important pour que nous puissions relever un nombre défini de traits précis, permettant à eux seuls de définir tous les mots du dictionnaire (contrairement au nombre fini de traits permettant de définir les établissements commerciaux dans le dictionnaire spécialisé qu'est le *Dictionnaire encyclopédique bilingue de la distribution* analysé par M.-C. L'Homme et J. Dancette), nous déciderons nous-même, pour chaque définition, des traits la constituant. Nous sélectionnerons les traits qui ont une signification précise, spécifique, et non les traits qui ont une signification générale. Seront considérés comme trop généraux les traits qui entrent dans la définition d'un grand nombre de mots dans le dictionnaire, car cela signifie qu'ils servent à opposer des mots à un niveau de généralité trop important. Par exemple, nous considérons qu'il y a cinq traits dans la définition de « logement » (« Local à usage d'habitation; spécialt Partie de maison, d'immeuble où l'on réside habituellement »), à savoir « local », « habitation », « maison », « immeubles » et « réside ». Certains traits sont parfois nommés par un groupe de mots, et non un mot unique. Rappelons d'ailleurs qu'en sémantique structurale, la « dénomination [des sèmes] peut au besoin consister en une périphrase » (Rastier 1987 : 36). Cependant, la « Recherche sur critères » n'offre pas la possibilité de rechercher des groupes de mots précis (tel que le permet par exemple la recherche avec guillemets dans les moteurs de recherche sur Internet). Ainsi, si nous entrons plusieurs mots dans la « Recherche sur critères », nous obtenons la liste des entrées du dictionnaire qui contiennent ces mots dans leur définition, mais pas obligatoirement les uns à côté des autres. Lorsque nous aurons à rechercher un trait nommé par plusieurs mots, nous ne porterons notre attention que sur les entrées proposées qui contiennent les différents mots du trait recherché en tant que trait unique.

Pour trouver les mots qui ont une relation sémantique avec l'équivalent, nous lancerons des recherches successives trait par trait, jusqu'à trouver un mot proche formellement du mot étranger. Il aurait été intéressant d'exploiter ce que nous avons appelé le « gradient de proximité sémantique » (voir p. 164), c'est-à-dire de lancer une première recherche avec tous les traits sauf un, puis une autre recherche avec tous les traits sauf un autre, puis tous les traits sauf deux, etc., pour avoir des listes de mots qui sont le plus proche possible sémantiquement de l'équivalent. Malheureusement, le nombre de combinaisons possibles serait alors démesuré pour cette étude (trente par exemple pour une définition à cinq traits). Nous considérerons qu'un mot-relais, proche sémantiquement de l'équivalent par un trait, en l'occurrence un sème de transfert, est suffisant pour que le mot étranger soit considéré comme transparent indirectement.

Par exemple, le nom NL *wet* est considéré comme transparent indirectement. En effet, lorsqu'on cherche la définition de son équivalent, « loi », on trouve un premier trait sémantique, « règle » :

The screenshot shows the 'Le Petit Robert de la langue française' dictionary interface. The search term 'loi' is entered in the search bar. The main content area displays the entry for 'loi' (n. f.), including its etymology and a primary definition: '1. **RÈGLE IMPÉRATIVE IMPOSÉE À L'HOMME** (fin Xe s.)'. The word 'RÈGLE' is highlighted with a red box. Below this, a detailed definition is provided: '1 (début XIIe s. **Règle** ou ensemble de règles obligatoires établies par l'autorité souveraine d'une société et sanctionnées par la force publique. *Lois humaines; lois positives, civiles.* « *L'Esprit des lois* », de Montesquieu. *Étude des lois.* → **nomologie.** » Les lois d'un État, d'un pays, d'une nation. → **législation**; 3. **droit.** *Recueil de lois.* → **code.** *Lois et institutions. Les lois de l'Ancien Régime.* → **capitulaire, charte, édit, ordonnance, règlement, statut.** *Loi écrite; loi coutumière* (dans les pays anglo-saxons, « *common law* »). « *Le plus grand*

et si l'on cherche ce trait sémantique dans les autres définitions du dictionnaire :



on observe que l'un des résultats de la recherche, « vertu », est proche formellement de *wet* :

The screenshot shows the interface of the Le Petit Robert dictionary. At the top, the title is "Le Petit Robert de la langue française". Below it, there is a search bar containing the word "loi". To the left, a sidebar titled "résultats de la recherche" lists various words, with "vertu - n. f." highlighted. The main content area displays the entry for "vertu - n. f.", which includes several numbered definitions and examples. The word "règle" in the second definition is highlighted with a red box. The interface also features a menu bar with options like "Fichier", "Édition", "Recherche", "Vues", and "Aide", and a toolbar with icons for navigation and search.

Le Petit Robert de la langue française

Fichier Édition Recherche Vues Aide

Le nouveau Petit Robert de la langue française 2007

loi

recherche par critères

sym plus féminins et pluriels

vertu - n. f.

plan étymologie renvois et contraires citations exemples et expressions synonymes homonymes

4 Vieilli ou plaisant Chasteté ou fidélité sentimentale, conjugale (d'une femme). → **honnêteté**. « *Cet infidèle mari qui semblait l'engager à commettre des fautes en taxant sa vertu d'insensibilité* » (Balzac). **Loc.** *Femme de petite vertu*, de mœurs légères.
= *Dragon* de vertu*.

5 (Sens objectif) La **règle** morale, le principe qui pousse à la vertu (2°). « *Ô vertu, science sublime des âmes simples* » (Rousseau). *Suivre le chemin, le sentier de la vertu*.

B UNE, LES VERTUS.

1 Disposition constante à accomplir une sorte d'actes moraux par un effort de volonté; qualité portée à un haut degré. « *C'est une grande et rare vertu que la patience* » (Gide). « *les vertus bourgeoises, et particulièrement, le goût de la propriété et de l'épargne* » (Chardonnet).
= *Parer qqn de toutes les vertus*, lui attribuer toutes les qualités

nomenciatuure explorateur recherche historique

regaranser - v. tr.
régulateur, trice - adj. et n. m.
religion - n. f.
répondre - v. tr. dir. et ind.
salique - adj.
selon - prép.
sévère - adj.
souffler - v.
spécialité - n. f.
statutaire - adj.
té - n. m.
tolérance - n. f.
trappiste - n. m.
tricher - v. intr.
typomètre - n. m.
usage - n. m.
vagabond, onde - adj. et n.
vertu - n. f.
volant, ante - adj.

109 réponses

Présentons maintenant les tableaux des hyperlexiques anglais/français, allemand/français et néerlandais/français qui ont été constitués grâce à cette méthodologie, et qui contiennent les mots transparents directement et les mots transparents indirectement entre ces langues. Rappelons toutefois un point capital : nous n'affirmons en rien que les mots qui apparaissent dans ces tableaux sont infailliblement compréhensibles par des francophones. Cette remarque, que nous avons déjà faite au début du chapitre précédent, sur la détermination des interlexiques, est d'autant plus vraie ici puisqu'il faut pratiquer des ajustements divers. Il s'agit uniquement pour nous de dresser la liste des mots qui obéissent aux règles de la transparence directe et de la transparence indirecte. De même, même si certaines voies de passage entre le mot étranger et son équivalent par le biais de l'ajustement sémantique paraissent contre-intuitives, il faut garder à l'esprit qu'il s'agit de l'application scrupuleuse de la méthodologie. Enfin, puisque la recherche des mots-relais s'effectue grâce au *Petit Robert*, ce sont les entrées proposées par ce dictionnaire qui jouent, dans cette troisième partie de notre travail, le rôle de mots : on trouvera par exemple, outre les lexies primaires, dérivées et composées (qui correspondent à notre caractérisation du mot, c.f. p. 19-20), des lexies prépositionnelles telles que « à l'avenant », ou bien des emprunts tels que « talk-show », considérés par les lexicographes du *Petit Robert* comme intégrés au lexique du français.

Les tableaux, pour chaque langue, sont organisés en six colonnes : apparaissent successivement la fréquence du mot étranger considéré, sa forme lemmatisée, sa catégorie grammaticale, l'équivalent français transparent directement ou indirectement avec le mot étranger, puis, si le mot étranger est transparent indirectement, les manipulations formelles et/ou sémantiques pratiquées pour passer du mot étranger à son équivalent. Dans la colonne « Ajustement sémantique », plusieurs abréviations dont nous donnons la signification dans le tableau suivant décrivent la façon dont le mot proche formellement du mot étranger a été trouvé :

Abréviation	Qui représente	Type de découverte du mot proche formellement du mot étranger
« r. »	« renvoi »	Un renvoi dans la notice de l'équivalent français est proche formellement du mot étranger.
« s. »	« sème »	Un des sèmes de la définition de l'équivalent français est proche formellement du mot étranger.
« m.r. » « s.t. »	« mot-relais » « sème de transfert »	Une autre entrée du dictionnaire, ayant le rôle de mot-relais, est proche formellement du mot étranger, et est proche sémantiquement de l'équivalent français par le biais d'un sème de transfert.
« m.r. » « s. »	« mot-relais » « sème »	Une autre entrée du dictionnaire, ayant le rôle de mot-relais, est proche formellement du mot étranger, et est proche sémantiquement de l'équivalent français car ce dernier est un sème de la définition du mot-relais.
« m.r. » « r. »	« mot-relais » « renvoi »	Un sème de la définition de l'équivalent constitue un renvoi dans la notice du mot-relais, ou bien l'équivalent lui-même constitue un renvoi dans la notice du mot-relais (auquel cas c'est le même mot qui apparaît dans la colonne « Équivalent français » et après l'abréviation « r. »).

Les sept exemples suivants, extraits du tableau de la transparence hyperlexicale anglais/français, illustrent les différents cas de figure présents dans les tableaux.

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
24165	look	V	regarder		r. : loucher

Ces lignes doivent se comprendre ainsi : le verbe *look*, qui a une fréquence de 24165 (c'est-à-dire qu'il a été relevé 24165 fois dans le corpus anglais, constitué d'environ 18 millions de mots), est transparent indirectement par rapport son équivalent français « regarder » (qui est le premier équivalent proposé dans le dictionnaire bilingue), par le biais du mot-relais « loucher ». « Loucher » est en effet proche formellement directement du mot anglais (les graphèmes <k> et <ch> peuvent tous deux noter, en français, le phonème /k/ ; ils sont donc équivalents) et proche sémantiquement indirectement de l'équivalent « regarder ». La relation sémantique n'est pas visible ici par la découverte d'un sème de transfert, mais « loucher » est effectivement dans le champ sémantique de « regarder » parce qu'il constitue un renvoi dans la notice de ce dernier.

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
23651	way	N	chemin		s. : voie

Ces lignes doivent se comprendre ainsi : le nom *way*, qui a une fréquence de 23651, est transparent indirectement avec son équivalent français « chemin », par le biais du mot-relais « voie ». « Voie » est en effet proche formellement directement du mot anglais (les graphèmes <w> et <v> peuvent tous deux noter, en français, le phonème /v/) et proche sémantiquement indirectement de l'équivalent « chemin ». La relation sémantique entre les deux mots français apparaît grâce au fait que « voie » est un des sèmes de la définition de « chemin ».

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
11794	keep	V	garder		m.r. : accaparer s.t. : retenir

Ces lignes doivent se comprendre ainsi : le verbe *keep*, qui a une fréquence de 11794, est transparent indirectement avec son équivalent français « garder », par le biais du mot-relais « accaparer ». « Accaparer » est en effet proche formellement directement du mot anglais (les graphèmes <k> et <cc> notent tous deux le phonème /k/, il y a une consonne finale de plus en français, le /r/, mais elle est la troisième, et le dernier <r> en français ne se prononce pas) et proche sémantiquement indirectement de l'équivalent « garder ». « Garder » et « accaparer » appartiennent au même champ sémantique par le sème qu'ils ont en commun, à savoir 'retenir'.

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
12983	hand	N	main		m.r. : handball s. : main

Ces lignes doivent se comprendre ainsi : le nom *hand*, qui a une fréquence de 12983, est transparent indirectement avec son équivalent français « main », par le biais du mot-relais « handball ». « Handball » est en effet proche formellement directement du mot anglais (ils commencent par les quatre mêmes premières lettres) et proche sémantiquement indirectement de l'équivalent « main ». « Main » et « handball » entretiennent une relation sémantique parce que « main » est un sème dans la définition de « handball ».

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
15911	seem	V	sembler	/s/-/ʃ/	m.r. : chimère r. : illusion

Ces lignes doivent se comprendre ainsi : le verbe *seem*, qui a une fréquence de 15911, est transparent indirectement avec son équivalent français « sembler » (et non directement car il y a ajout en français de *deux* consonnes au contact du <m>), par le biais du mot-relais « chimère ». « Chimère » est en effet proche formellement indirectement du mot anglais (par l'ajustement formel /s/-/ʃ/, et il y a une consonne finale en plus en français, mais elle est la troisième) et proche sémantiquement indirectement de l'équivalent « sembler ». Les deux mots français entretiennent une relation sémantique parce que le sème 'illusion' de la définition de « sembler » constitue un renvoi dans la notice de « chimère ».

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
3089	top	N	sommet		m.r. : top niveau r. : sommet

Ces lignes doivent se comprendre ainsi : le nom *top*, qui a une fréquence de 3089, est transparent indirectement avec son équivalent français « sommet », par le biais du mot-relais « top niveau ». « Top niveau » est en effet proche formellement directement du mot anglais (l'un des mots a moins de quatre lettres, et le mot le plus long commence par toutes les lettres du mot le plus court) et proche sémantiquement indirectement de l'équivalent « sommet ». Les deux mots français entretiennent une relation sémantique parce que « sommet » constitue un renvoi dans la notice de « top niveau ».

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
24308	just	ADV	juste		

Enfin, ces lignes doivent se comprendre ainsi : l’adverbe *just*, qui a une fréquence de 24308, est transparent directement avec son équivalent « juste ». Il n’y a en effet recours ni à l’ajustement formel ni à l’ajustement sémantique : *just* est proche formellement directement d’un des équivalents proposés dans le dictionnaire bilingue. Les mots apparaissant dans les tableaux de l’hyperlexique et dont les colonnes « Ajustement formel » et « Ajustement sémantique » sont vides correspondent donc à la partie de l’hyperlexique qu’est l’interlexique.

Comme dans le chapitre précédent, nous allons extraire le début d’un des tableaux, celui de l’anglais, pour expliquer la façon dont nous l’avons constitué⁸⁹. La méthodologie est identique pour les autres langues. Nous renvoyons aux annexes pour une consultation des tableaux dans leur intégralité. Nous ne sélectionnons pas ici les dix premières lignes, mais les douze premières, pour pouvoir présenter des mots qui demandent un ajustement formel (seul ou accompagné d’un ajustement sémantique).

	Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
					Ajustement formel	Ajustement sémantique
1.	70930	can	V	pouvoir		m.r. : chance s.t. : possibilité
2.	69206	will	V	vouloir		
3.	60143	there	ADV	y		m.r. : terre s.t. : lieu
4.	41842	make	V	marquer		
5.	40954	so	ADV	si		

⁸⁹ Nous avons ajouté ici, comme dans le chapitre précédent, une numérotation des lignes pour les besoins de l’explication. Cette numérotation n’apparaît pas dans les tableaux des annexes.

	Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
					Ajustement formel	Ajustement sémantique
6.	38013	know	V	connaître		
7.	37195	one	N	un		
8.	35351	time	N	temps		m.r. : timing s. : temps
9.	35152	come	V	venir		m.r. : chemin s.t. : se déplacer
10.	34323	take	V	prendre		r. : toucher
11.	33749	then	ADV	donc	/ð/-/d/	
12.	29231	man	N	homme	/n/-/l/	r. : mâle

Parmi ces douze premiers mots relevant de l’hyperlexique anglais/français, certains relèvent également de l’interlexique : comme nous l’avons annoncé, les lignes pour lesquelles les colonnes « Ajustement formel » et « Ajustement sémantique » sont vides correspondent aux mots transparents directement.

1. Le *Robert & Collins* propose pour équivalents du verbe *can* les verbes « pouvoir » et « savoir ». Ces deux équivalents sont, par définition, proches sémantiquement directement du verbe anglais. Toutefois, aucun n’en est proche formellement, ni directement ni indirectement. Nous partons donc à la recherche d’un éventuel mot-relais, c’est-à-dire d’un mot français qui serait proche formellement du mot anglais et proche sémantiquement indirectement du premier équivalent.

Nous cherchons la définition du verbe « pouvoir » dans le *Petit Robert* électronique, et lançons en premier lieu les recherches qui s’appuient sur les renvois des notices. Nous cherchons donc d’abord s’il y a parmi les renvois de la notice de « pouvoir » un mot qui est proche formellement du mot anglais ; ce n’est pas le cas. Nous cherchons ensuite grâce à la fonctionnalité « Recherche sur critères » si « pouvoir » constitue un renvoi dans une autre notice du dictionnaire ; ce n’est pas le cas non plus. Nous lançons

dans un second temps les recherches qui s'appuient sur les traits sémantiques. Nous cherchons alors s'il y a parmi les traits sémantiques de la définition de « pouvoir » un mot qui serait proche formellement du mot étranger ; ce n'est pas le cas non plus. Nous cherchons ensuite si « pouvoir » lui-même est un trait sémantique dans une autre définition du dictionnaire. Puisque ce n'est toujours pas le cas, nous cherchons enfin si un trait de la définition de « pouvoir » est également un trait dans la définition d'un mot qui serait proche formellement du mot anglais. La recherche avec les premiers traits « possible », « hypothèse », « souhait » et « capacité » se révèle infructueuse. En revanche, en cherchant le trait suivant, à savoir « possibilité », on trouve parmi les 73 résultats de la recherche, c'est-à-dire parmi les notices qui contiennent « possibilité » dans leur définition, un mot qui est proche formellement du mot anglais : « chance ». *Can* et « chance » sont proches formellement directement : il y a deux consonnes de plus en français, <h> et <c>, mais qui sont au contact d'autres consonnes, qui, elles, sont équivalentes à celles de l'anglais ; les deux mots ont donc la même trame consonantique. Ces recherches successives nous conduisent à la conclusion suivante : il existe un mot français proche formellement directement du verbe anglais *can* ; ce mot français, « chance », est proche sémantiquement indirectement de l'équivalent de *can*, c'est-à-dire « pouvoir », par le biais d'un sème de transfert, c'est-à-dire « possibilité ». *Can* est donc considéré comme transparent indirectement.

2. Le verbe anglais *will* est transparent directement par rapport au français. Nous avons expliqué dans le chapitre 3.2. pourquoi nous l'avons considéré comme un mot transparent directement.

3. Il y a deux équivalents proposés dans le dictionnaire bilingue pour l'adverbe *there* : « y » et « là ». Aucun des deux n'en est proche formellement. Nous lançons donc les recherches pour découvrir s'il existe un mot-relais qui serait proche formellement de *there* et proche sémantiquement indirectement de « y ». L'étape de recherche qui a abouti est, comme dans l'exemple de *can*, la dernière étape, c'est-à-dire la recherche d'un trait sémantique de la définition de l'équivalent dans d'autres définitions du *Petit Robert* électronique. En effet, le trait « lieu » de la définition de « y » constitue aussi un trait dans la définition du mot « terre ». Étant donné que « terre » et *there* sont proches formellement (directement en l'occurrence : ils sont la même trame consonantique puisque le <h> anglais est au contact d'une autre consonne et que le <r> français est également au contact

d'une autre consonne), il existe un mot français proche formellement directement de l'adverbe anglais *there* ; ce mot français, « terre », est proche sémantiquement indirectement de l'équivalent de *there*, c'est-à-dire « y », par le biais d'un sème de transfert, c'est-à-dire « lieu ». *There* est donc considéré comme transparent indirectement.

4. 5. 6. 7. Les mots anglais *make*, *so*, *know* et *one* sont transparents directement par rapport au français. Nous renvoyons au chapitre 3.2. pour leur analyse.

8. Les équivalents proposés pour le nom *time* sont les suivants : « temps », « heure », « moment », « fois », « mesure ». Aucun n'en est proche formellement (*time* et « temps » ne sont pas proches formellement car il y a ajout de *deux* consonnes en français, <p> et <s>, au contact du <m>). Nous lançons donc les recherches à partir de la notice de « temps ». C'est une fois encore une recherche s'appuyant sur les traits qui s'avère concluante, mais dans ce cas précis, c'est l'équivalent « temps » qui constitue un trait sémantique dans la définition d'un autre mot du dictionnaire, à savoir « timing ». Ainsi, « timing » est proche sémantiquement indirectement de « temps », et proche formellement directement de *time* (les deux mots ont la même trame consonantique : le groupe de consonnes <ng> final en français est précédé de deux consonnes équivalentes entre les deux mots, à savoir <t> et <m>). Par conséquent, puisque *time* est proche formellement d'un mot français, « timing », qui, lui, est proche sémantiquement indirectement de son équivalent « temps », *time* est considéré comme transparent indirectement.

9. En face du verbe anglais *come* sont proposés les équivalents suivants : « venir », « arriver », « se trouver », « être placé », « se produire », « jouir ». Aucun n'en est proche formellement. Nous lançons donc les recherches à partir de la notice de « venir ». C'est ici encore l'étape de la recherche des traits de la définition de l'équivalent dans d'autres définitions du dictionnaire qui est concluante. En effet, le trait « se déplacer », présent dans la définition de « venir », est également présent, notamment, dans la définition de « chemin ». « Chemin » est proche formellement directement de *come* : ils ont la même trame consonantique, parce que le <h> français est au contact d'une autre consonne, le <c>, et que le <n> final français est précédé de deux consonnes identiques entre les deux mots, <c> et <m>. Ainsi, il existe un mot français proche formellement directement du verbe anglais *come* ; ce mot français, « chemin », est proche sémantiquement indirectement de l'équivalent de *come*, c'est-à-dire « venir », par le biais d'un sème de

transfert, c'est-à-dire « se déplacer ». *Come* est donc considéré comme transparent indirectement.

10. De nombreux équivalents sont proposés pour le verbe anglais *take* : « prendre », « saisir », « tirer », « enlever », « ôter », « s'emparer de », « capturer », « attraper », « obtenir », « remporter », « avoir », « s'asseoir sur », « louer », « suivre », « grimper », « sauter », « passer », « se présenter à », « faire », « faire cours à », « accepter », « supporter », « contenir », « avoir une capacité de », « supposer », « imaginer », « demander », « être suivi de », « apporter », « emporter », « emmener », « conduire », « accompagner ». Aucun de ces équivalents n'est proche formellement de *take*. Nous lançons donc des recherches à partir de la notice de « prendre ». Nous trouvons un mot proche formellement du mot anglais dès la première étape de recherche : un des renvois dans la notice de « prendre », « toucher », est proche formellement directement de *take*. « Toucher » et *take* ont la même trame consonantique car les graphèmes <ch> et <k> peuvent tous deux noter en français le phonème /k/. Ainsi, *take* est proche formellement d'un mot français, « toucher », qui, lui, est proche sémantiquement indirectement de l'équivalent du verbe anglais, « prendre ». *Take* est donc considéré comme transparent indirectement.

11. Les équivalents proposés pour l'adverbe anglais *then* sont : « alors », « à l'époque », « puis », « ensuite », « alors », « donc », « et puis », « d'ailleurs ». Un de ces équivalents est proche formellement indirectement du mot anglais : « donc ». Ils sont proches formellement indirectement parce que, après avoir pratiqué un ajustement formel, ils ont la même trame consonantique : les phonèmes /ð/ et /d/ appartiennent au même champ phonologique (ils sont donc considérés comme équivalents) ; la consonne de plus en français, <c>, est au contact d'une autre consonne. *Then* est proche formellement indirectement d'un de ses équivalents, dont il est donc proche sémantiquement directement. *Then* est donc considéré comme transparent indirectement.

12. Les équivalents proposés pour le nom anglais *man* sont : « homme », « domestique », « ouvrier », « employé », « joueur », « équipier », « soldat », « matelot », « pièce », « pion ». Aucun de ces équivalents n'est proche formellement de *man*. Nous lançons donc des recherches à partir de la notice de « homme ». Un des renvois de la

notice de « homme », « mâle », est proche formellement indirectement du mot anglais. « Mâle » et *man* ont la même trame consonantique après avoir pratiqué l’ajustement formel /l/-/n/. Ainsi, *man* est proche formellement d’un mot français, « mâle », qui, lui, est proche sémantiquement de l’équivalent « homme ». *Man* est donc considéré comme transparent indirectement.

Nous allons maintenant présenter, dans chacune des sections suivantes, les résultats obtenus après l’analyse des trois corpus.

3.3.1. Détermination de l’hyperlexique anglais/français

Nous allons présenter les résultats chiffrés pour l’anglais de la façon suivante : nous rappelons d’abord le tableau présentant le nombre de mots transparents directement, nous donnons ensuite le tableau présentant le nombre de mots transparents indirectement, et donnons enfin le tableau présentant le nombre global de mots transparents (directement et indirectement, c’est-à-dire ceux qui rendent compte de l’hyperlexique). Cette présentation nous permettra d’avoir une vue d’ensemble de l’extension de la transparence. Rappelons que le corpus anglais est constitué de 1 000 mots, répartis en 131 adjectifs, 115 adverbes, 496 noms et 258 verbes.

Mots transparents directement :

	Absolu	Relatif à l’interlexique	Relatif au corpus
Total	528	100 %	53 %
Adjectifs	77	15 %	8 %
Adverbes	25	5 %	3 %
Noms	310	59 %	31 %
Verbes	116	22 %	12 %

Mots transparents indirectement :

	Absolu	Relatif à la transp. indir.	Relatif au corpus
Total	320	100 %	32 %
Adjectifs	38	12 %	4 %
Adverbes	27	8 %	3 %
Noms	155	48 %	16 %
Verbes	100	31 %	10 %

Mots transparents globalement :

	Absolu	Relatif à l'hyperlexique	Relatif au corpus
Total	848	100 %	85 %
Adjectifs	115	14 %	12 %
Adverbes	52	6 %	5 %
Noms	465	55 %	46 %
Verbes	216	25 %	22 %

Le résultat le plus remarquable est sans doute celui du nombre de mots transparents après la prise en compte de la transparence indirecte : 85 % des mots du corpus obéissent aux règles des transparences directe et indirecte. Ainsi, sur les 1 000 mots lexicaux les plus fréquents de l'anglais, 848 sont des représentants d'hyperlexèmes partagés avec le français.

L'évolution du taux de transparence par catégorie grammaticale est présentée dans le tableau suivant :

	Transparence directe		Transparence indirecte		Transparence globale	
	Absolu	Relatif à la catégorie	Absolu	Relatif à la catégorie	Absolu	Relatif à la catégorie
Adjectifs	77	59 %	38	29 %	115	88 %
Adverbes	25	22 %	27	23 %	52	45 %
Noms	310	62 %	155	31 %	465	94 %
Verbes	116	45 %	100	39 %	216	84 %

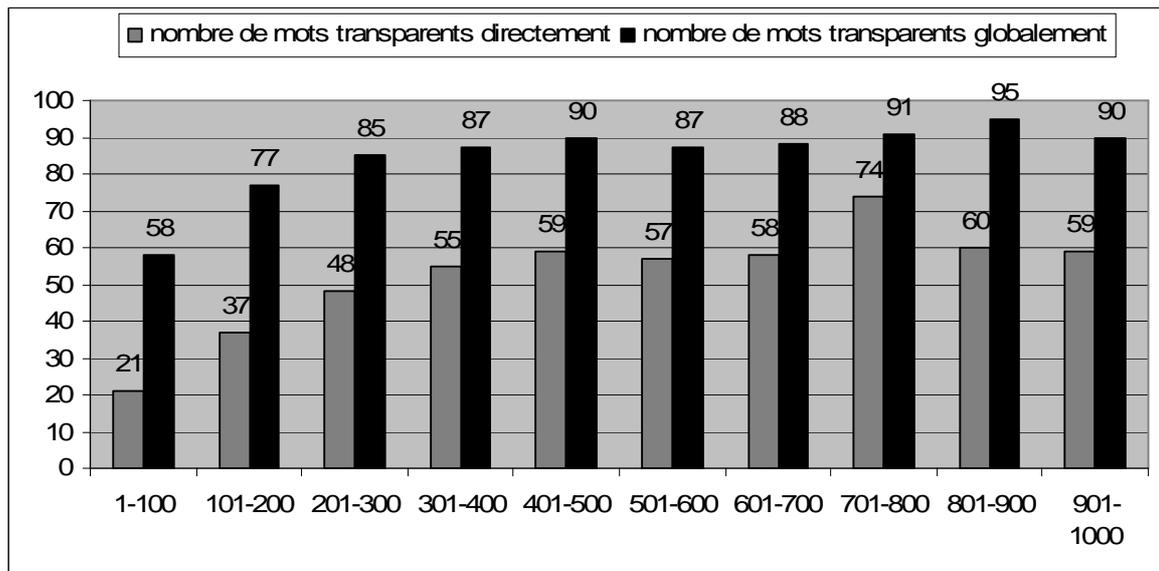
Les noms sont proportionnellement les plus transparents directement ; ce sont les verbes qui profitent le plus de la transparence indirecte ; mais, finalement, ce sont les noms qui sont globalement les plus transparents (plus de 9 noms sur 10 sont transparents). Ce chiffre ne doit pas voiler celui du taux de transparence globale des autres catégories : presque 9 adjectifs sur 10 sont transparents, plus de 8 verbes sur 10 sont transparents, et presque la moitié des adverbes sont transparents. Les adverbes ont néanmoins à peine bénéficié de l'extension de la définition de la transparence, puisque, même si près de la moitié d'entre eux sont transparents, au final, seuls 6 % des mots transparents sont des adverbes.

32 % supplémentaires du corpus anglais sont devenus transparents. Il nous a semblé intéressant de savoir par quels moyens ces mots étaient devenus transparents. Le tableau suivant présente combien de fois les différents types d'ajustements ont été utilisés. Ces chiffres sont toutefois à considérer avec prudence, puisque nous n'avons pas exploré toutes les voies permettant de rendre un mot transparent. Nous avons en effet annoncé que la découverte d'un mot-relais suffisait à considérer un mot étranger comme transparent indirectement. Les résultats qui suivent sont donc valables pour les mots-relais retenus.

	Absolu	Relatif à la transp. indir.	Relatif au corpus
Ajustements formels seuls	60	19 %	6 %
Ajustements sémantiques seuls	163	51 %	16 %
Recours aux deux ajustements	97	30 %	10 %
Total ajustements formels	157	49 %	16 %
Total ajustements sémantiques	260	81 %	26 %
Total ajustements (= mots transp. indirectement)	320	100 %	32 %

Il apparaît clairement que le gain de transparence s'est effectué très majoritairement grâce aux ajustements sémantiques, qui ont été utilisés 8 fois sur 10, relativement au nombre de mots transparents indirectement, contre 1 fois sur 2 pour les ajustements formels. Ainsi, sur les 849 mots transparents de l'anglais, seuls 60 requièrent uniquement un ajustement formel, alors que 163 requièrent uniquement un ajustement sémantique. Nous pouvons en conclure, en gardant à l'esprit les réserves que nous avons exposées avant le tableau, qu'une grande partie des mots anglais sont proches par la forme de mots français. Il convient ensuite de jouer sur la signification des mots ainsi suggérés en français pour aboutir aux équivalents des mots anglais considérés.

Nous avons enfin cherché à savoir si le taux de transparence était dépendant de la fréquence. Le graphique suivant rappelle l'évolution du taux de transparence directe, et présente l'évolution du taux de transparence globale par tranches de 100 mots.



La transparence directe et la transparence globale suivent proportionnellement le même type d'évolution. Il faut noter que la transparence globale peut constituer jusqu'à 95 % d'une tranche de mots, et que, si l'on fait exception des 200 mots les plus fréquents, le taux de transparence globale pour les mots lexicaux anglais est d'environ 90 %.

3.3.2. Détermination de l'hyperlexique allemand/français

Comme pour l'anglais, nous présentons les chiffres de la transparence globale en allemand en rappelant auparavant les chiffres de la transparence directe et en donnant ceux de la transparence indirecte. Rappelons que le corpus allemand est constitué de 1 000 mots, répartis en 173 adjectifs, 119 adverbes, 472 noms et 236 verbes.

Mots transparents directement :

	Absolu	Relatif à l'interlexique	Relatif au corpus
Total	222	100 %	22 %
Adjectifs	44	20 %	4 %
Adverbes	7	3 %	1 %
Noms	150	68 %	15 %
Verbes	21	9 %	2 %

Mots transparents indirectement :

	Absolu	Relatif à la transp. indir.	Relatif au corpus
Total	520	100 %	52 %
Adjectifs	85	16 %	8 %
Adverbes	41	8 %	4 %
Noms	251	48 %	25 %
Verbes	143	28 %	14 %

Mots transparents globalement :

	Absolu	Relatif à l'hyperlexique	Relatif au corpus
Total	742	100 %	74 %
Adjectifs	129	17 %	13 %
Adverbes	48	6 %	5 %
Noms	401	54 %	40 %
Verbes	164	22 %	16 %

La moitié du corpus allemand est transparente indirectement, si bien que les trois quarts des mots allemands sont transparents globalement. Parmi ceux-ci, les noms sont toujours les plus nombreux. Les verbes, qui étaient en nombre nettement moins important que les adjectifs au sein de la transparence directe, bénéficient de la transparence indirecte, puisqu'ils constituaient 9 % des mots transparents directement et constituent maintenant 22 % des mots transparents globalement (les adjectifs quant à eux passent de 20 % à 17 %).

Le tableau suivant présente l'évolution de la transparence au sein des catégories grammaticales :

	Transparence directe		Transparence indirecte		Transparence globale	
	Absolu	Relatif à la catégorie	Absolu	Relatif à la catégorie	Absolu	Relatif à la catégorie
Adjectifs	44	25 %	85	49 %	129	75 %
Adverbes	7	6 %	41	34 %	48	40 %
Noms	150	32 %	251	53 %	401	85 %
Verbes	21	9 %	143	61 %	164	69 %

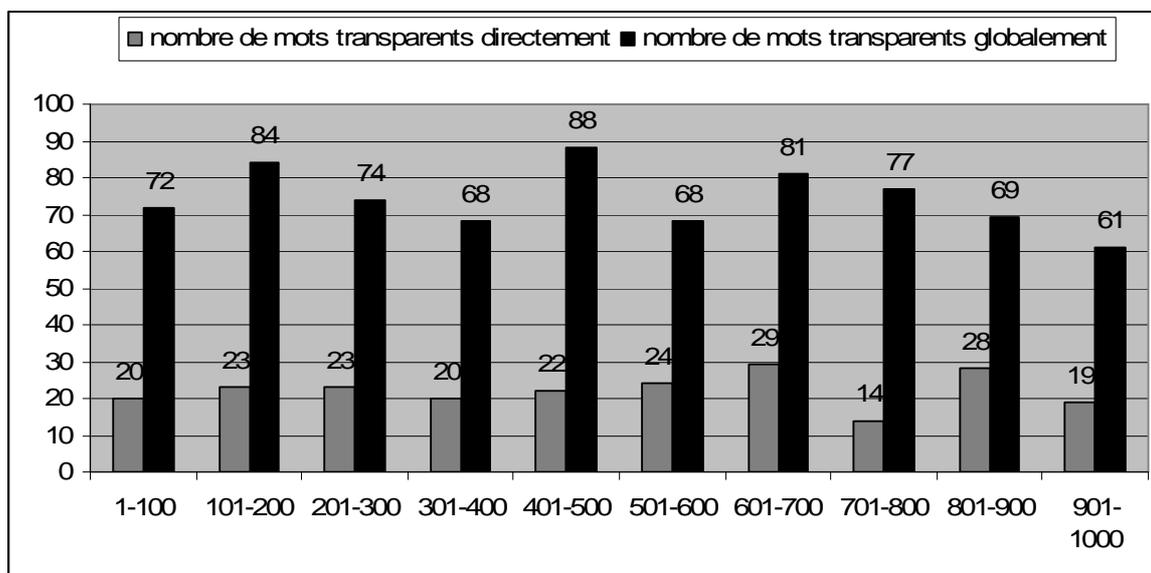
On voit ici pourquoi les verbes constituent une plus grande partie des mots transparents globalement que des mots transparents directement : en transparence indirecte, ce sont eux qui sont les plus transparents. Les adverbes connaissent également une amélioration considérable de leur transparence grâce aux proximités formelles et sémantiques indirectes. De façon générale, le corpus allemand bénéficie de la transparence indirecte.

Au sein de la transparence indirecte, l'utilisation des ajustements se répartit comme suit (les remarques faites dans la section consacrée à l'anglais, quant au fait que les résultats ne valent que pour les mots-relais retenus, est évidemment toujours valable) :

	Absolu	Relatif à la transp. indir.	Relatif au corpus
Ajustements formels seuls	118	23 %	12 %
Ajustements sémantiques seuls	144	28 %	14 %
Recours aux deux ajustements	258	50 %	26 %
Total ajustements formels	376	72 %	38 %
Total ajustements sémantiques	402	77 %	40 %
Total ajustements (= mots transp. indirectement)	520	100 %	52 %

Dans la transparence indirecte, le recours à un type d'ajustement ou à l'autre est pratiquement équivalent, et les deux ont été pratiqués ensemble la moitié des fois. Il ne se dégage donc pas de préférence en faveur de l'un ou de l'autre, et il semble que, pour rendre transparents les 52 % du corpus allemand qui le sont de façon indirecte, il faille savoir mettre en œuvre des manipulations formelles et sémantiques.

L'évolution du taux de transparence en fonction de la fréquence est présentée par le graphique suivant :



Non seulement les proportions ne suivent pas une évolution progressive en fonction de la fréquence en transparence directe et en transparence globale, mais elles ne sont pas

non plus proportionnelles : ce n'est donc pas là où l'interlexique est le plus important ou le plus bas que l'hyperlexique l'est aussi. Notons aussi une baisse constante du taux de transparence globale dans les quatre dernières tranches. L'hypothèse selon laquelle la transparence serait en relation avec la fréquence ne se vérifie donc pas pour l'allemand, ou du moins très peu, entre les deux premières tranches.

3.3.3. Détermination de l'hyperlexique néerlandais/français

Nous présentons les chiffres de la transparence globale en néerlandais en donnant auparavant les chiffres de la transparence directe et ceux de la transparence indirecte. Rappelons que le corpus néerlandais est constitué de 1 000 mots, répartis en 172 adjectifs, 117 adverbes, 398 noms et 313 verbes.

Mots transparents directement :

	Absolu	Relatif à l'interlexique	Relatif au corpus
Total	208	100 %	21 %
Adjectifs	40	19 %	4 %
Adverbes	9	4 %	1 %
Noms	133	64 %	13 %
Verbes	26	13 %	3 %

Mots transparents indirectement :

	Absolu	Relatif à la transp. indir.	Relatif au corpus
Total	509	100 %	51 %
Adjectifs	87	17 %	9 %
Adverbes	42	8 %	4 %
Noms	209	41 %	21 %
Verbes	171	34 %	17 %

Mots transparents globalement :

	Absolu	Relatif à l'hyperlexique	Relatif au corpus
Total	717	100 %	72 %
Adjectifs	127	18 %	13 %
Adverbes	51	7 %	5 %
Noms	342	48 %	34 %
Verbes	197	27 %	20 %

Le corpus néerlandais est constitué de deux fois et demie de plus de mots transparents indirectement que de mots transparents directement, si bien que près des trois quarts des 1 000 mots lexicaux les plus fréquents du néerlandais sont transparents par rapport au français. Parmi ces mots transparents, presque la moitié sont des noms ; arrivent ensuite les verbes, les adjectifs et les adverbes.

L'évolution de la proportion de mots transparents au sein de chaque catégorie grammaticale est visible dans le tableau suivant :

	Transparence directe		Transparence indirecte		Transparence globale	
	Absolu	Relatif à la catégorie	Absolu	Relatif à la catégorie	Absolu	Relatif à la catégorie
Adjectifs	40	23 %	87	51 %	127	74 %
Adverbes	9	8 %	42	36 %	51	44 %
Noms	133	33 %	209	52 %	342	86 %
Verbes	26	8 %	171	55 %	197	63 %

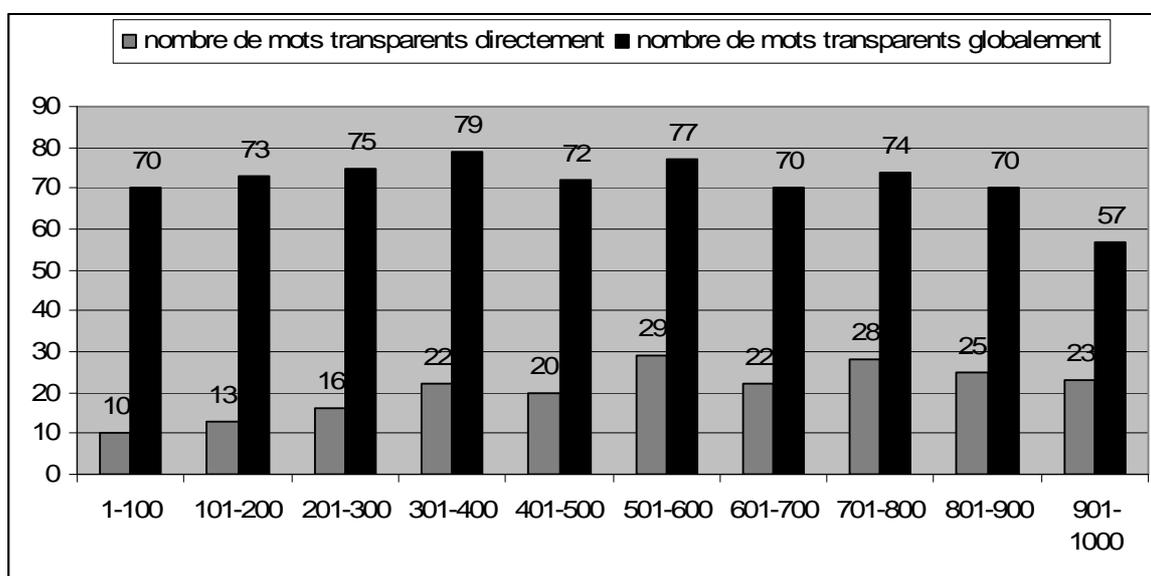
Environ la moitié des adjectifs, des noms et des verbes sont transparents indirectement ; ils connaissent donc proportionnellement la même augmentation entre l'interlexique et l'hyperlexique. Seuls les adverbes sont transparents pour moins de la moitié.

Pour les 509 mots transparents indirectement, les deux types d'ajustements ont été pratiqués de la sorte :

	Absolu	Relatif à la transp. indir.	Relatif au corpus
Ajustements formels seuls	93	18 %	9 %
Ajustements sémantiques seuls	186	37 %	19 %
Recours aux deux ajustements	230	45 %	23 %
Total ajustements formels	323	63 %	32 %
Total ajustements sémantiques	416	82 %	42 %
Total ajustements (= mots transp. indirectement)	509	100 %	51 %

Ce sont les ajustements sémantiques qui ont été les plus nombreux, même s'il y a eu également recours assez souvent à des ajustements formels. Cela tend à indiquer que le néerlandais possède un certain nombre de mots qui, même s'ils ne sont pas transparents directement par rapport au français, sont proches du point de vue de la forme de mots français, et qui peuvent donc être identifiés par des francophones.

Le graphique suivant présente l'évolution du taux de transparence directe et du taux de transparence globale par tranches de 100 mots, par fréquence décroissante.



On remarque que les nombres de mots transparents directement d'une part et de mots transparents globalement de l'autre suivent la même courbe. En d'autres termes, c'est quand les mots d'une tranche sont plus transparents directement que dans la tranche précédente que les mots de la même tranche sont également plus transparents globalement que dans la tranche précédente.

3.3.4. Bilan comparatif des hyperlexiques

Nous allons, dans cette section, présenter de façon synthétique et comparative les résultats chiffrés liés à l'identification des trois interlexiques et hyperlexiques.

Voici d'abord les nombres de mots transparents directement, indirectement puis de mots transparents globalement.

Mots transparents directement :

	ANGLAIS			ALLEMAND			NEERLANDAIS		
	Absolu	Relatif à l'inter- lexique	Relatif au corpus	Absolu	Relatif à l'inter- lexique	Relatif au corpus	Absolu	Relatif à l'inter- lexique	Relatif au corpus
Total	528	100 %	53 %	222	100 %	22 %	208	100 %	21 %
Adjectifs	77	15 %	8 %	44	20 %	4 %	40	19 %	4 %
Adverbes	25	5 %	3 %	7	3 %	1 %	9	4 %	1 %
Noms	310	59 %	31 %	150	68 %	15 %	133	64 %	13 %
Verbes	116	22 %	12 %	21	9 %	2 %	26	13 %	3 %

Mots transparents indirectement :

	ANGLAIS			ALLEMAND			NEERLANDAIS		
	Absolu	Relatif à la transp. indir.	Relatif au corpus	Absolu	Relatif à la transp. indir.	Relatif au corpus	Absolu	Relatif à la transp. indir.	Relatif au corpus
Total	320	100 %	32 %	520	100 %	52 %	509	100 %	51 %
Adjectifs	38	12 %	4 %	85	16 %	8 %	87	17 %	9 %
Adverbes	27	8 %	3 %	41	8 %	4 %	42	8 %	4 %
Noms	155	48 %	16 %	251	48 %	25 %	209	41 %	21 %
Verbes	100	31 %	10 %	143	28 %	14 %	171	34 %	17 %

Mots transparents globalement :

	ANGLAIS			ALLEMAND			NEERLANDAIS		
	Absolu	Relatif à l'hyper- lexique	Relatif au corpus	Absolu	Relatif à l'hyper- lexique	Relatif au corpus	Absolu	Relatif à l'hyper- lexique	Relatif au corpus
Total	848	100 %	85 %	742	100 %	74 %	717	100 %	72 %
Adjectifs	115	14 %	12 %	129	17 %	13 %	127	18 %	13 %
Adverbes	52	6 %	5 %	48	6 %	5 %	51	7 %	5 %
Noms	465	55 %	46 %	401	54 %	40 %	342	48 %	34 %
Verbes	216	25 %	22 %	164	22 %	16 %	197	27 %	20 %

La prise en compte des transparences directe et indirecte fait atteindre quasiment les 75 % en allemand et en néerlandais et 85 % en anglais la proportion de mots transparents. D'une façon générale, l'allemand et le néerlandais ont des comportements très similaires, et l'anglais se différencie. Même si les trois langues ont un taux de transparence très élevé, il est intéressant de constater que ces chiffres ne sont pas atteints de la même façon dans les trois langues. En anglais, 62 % des mots transparents sont transparents directement, alors qu'en allemand et en néerlandais, 70 % des mots transparents sont transparents indirectement. Le lexique transparent de l'anglais est donc plus proche de celui du français que ne le sont ceux de l'allemand et du néerlandais, c'est-à-dire qu'il demande proportionnellement la mise en œuvre d'un moins grand nombre d'ajustements.

Pour analyser les catégories grammaticales séparément, rappelons l'évolution du taux de transparence de chacune d'elles :

Mots transparents directement :

	ANGLAIS		ALLEMAND		NEERLANDAIS	
	Absolu	Relatif à la catégorie	Absolu	Relatif à la catégorie	Absolu	Relatif à la catégorie
Adjectifs	77	59 %	44	25 %	40	23 %
Adverbes	25	22 %	7	6 %	9	8 %
Noms	310	62 %	150	32 %	133	33 %
Verbes	116	45 %	21	9 %	26	8 %

Mots transparents indirectement :

	ANGLAIS		ALLEMAND		NEERLANDAIS	
	Absolu	Relatif à la catégorie	Absolu	Relatif à la catégorie	Absolu	Relatif à la catégorie
Adjectifs	38	29 %	85	49 %	87	51 %
Adverbes	27	23 %	41	34 %	42	36 %
Noms	155	31 %	251	53 %	209	52 %
Verbes	100	39 %	143	61 %	171	55 %

Mots transparents globalement :

	ANGLAIS		ALLEMAND		NEERLANDAIS	
	Absolu	Relatif à la catégorie	Absolu	Relatif à la catégorie	Absolu	Relatif à la catégorie
Adjectifs	115	88 %	129	75 %	127	74 %
Adverbes	52	45 %	48	40 %	51	44 %
Noms	465	94 %	401	85 %	342	86 %
Verbes	216	84 %	164	69 %	197	63 %

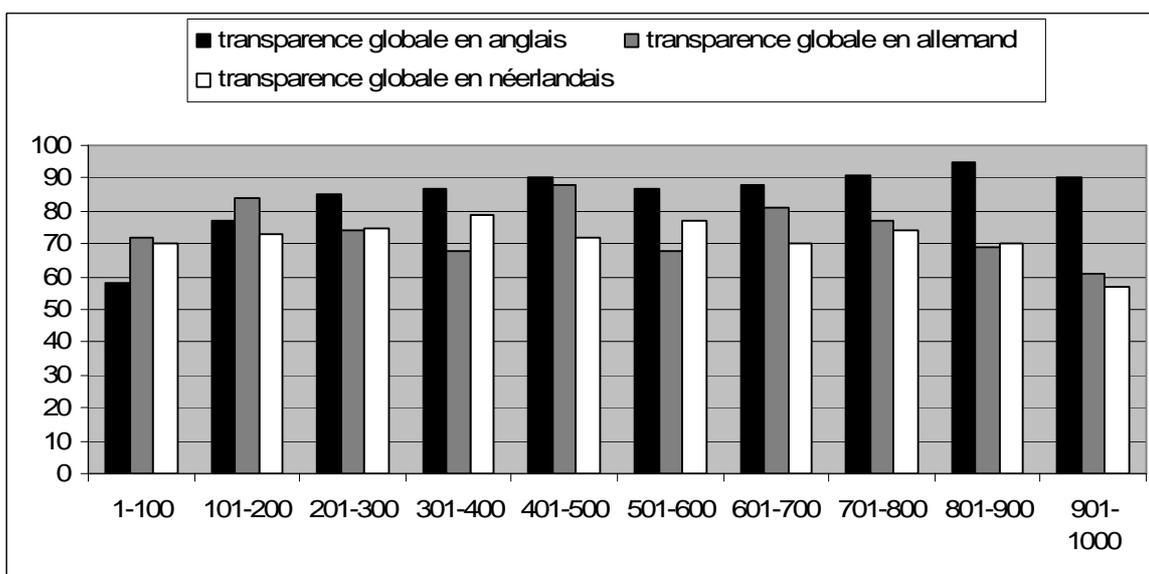
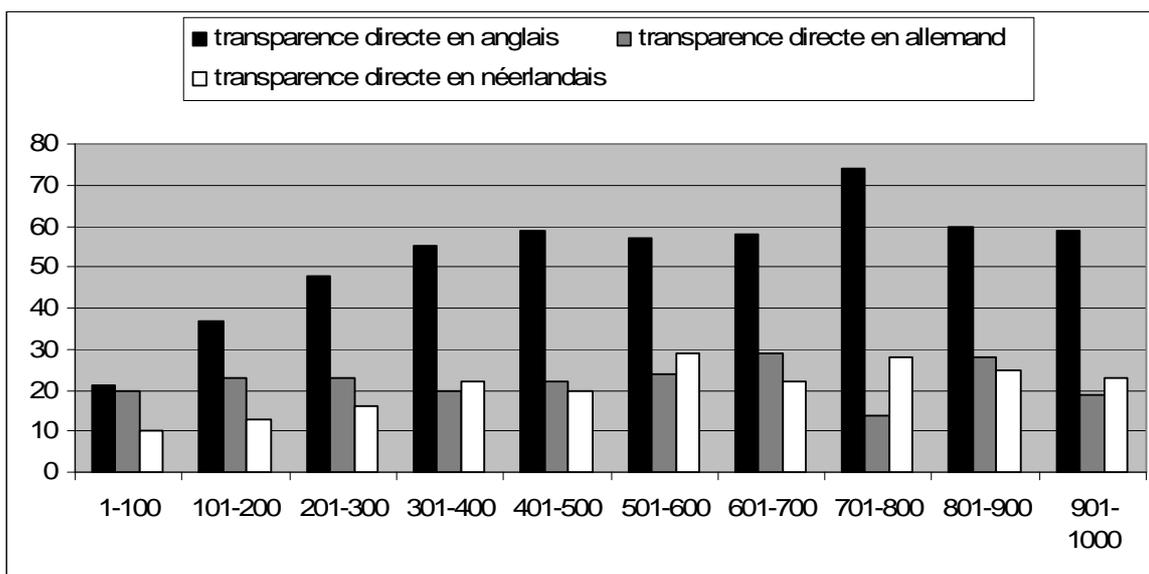
Tant en transparence directe qu'en transparence globale, les noms sont dans les trois langues la catégorie la plus transparente. De plus, on observe, comme pour les chiffres globaux, que les taux de transparence directe par catégorie qui sont faibles en allemand et en néerlandais connaissent une forte augmentation grâce à la transparence indirecte. Dans les trois langues, ce sont les verbes qui sont les plus transparents indirectement. Les adverbes sont manifestement une catégorie opaque, puisque, dans les trois langues, moins de la moitié d'entre eux sont transparents et que les adverbes transparents ne constituent que 5 % des corpus des trois langues.

Les six tableaux précédents font ressortir par les valeurs en pourcentages à quel point l'allemand et le néerlandais sont similaires, voire identiques. Le tableau suivant, qui présente de façon synthétique les chiffres des trois langues germaniques quant au nombre de fois où les deux types d'ajustements ont été utilisés en transparence indirecte, laisse voir une différence intéressante entre l'allemand et le néerlandais :

	ANGLAIS			ALLEMAND			NEERLANDAIS		
	Absolu	Relatif à la transp. indir.	Relatif au corpus	Absolu	Relatif à la transp. indir.	Relatif au corpus	Absolu	Relatif à la transp. indir.	Relatif au corpus
Ajustements formels seuls	60	19%	6%	118	23%	12%	93	18%	9%
Ajustements sémantiques seuls	163	51%	16%	144	28%	14%	186	37%	19%
Recours aux deux ajustements	97	30%	10%	258	50%	26%	230	45%	23%
Total ajustements formels	157	49%	16%	376	72%	38%	323	63%	32%
Total ajustements sémantiques	260	81%	26%	402	77%	40%	416	82%	42%
Total ajustements (= mots transp. indir.)	320	100%	32%	520	100%	52%	509	100%	51%

Dans les autres analyses, l'anglais s'opposait aux deux autres langues germaniques. Ici, on observe une progression qui va de l'anglais, au néerlandais et à l'allemand. En effet, en anglais, il y a eu nettement moins recours à l'ajustement formel qu'à l'ajustement sémantique (tant utilisés seuls que combinés à l'autre ajustement). En néerlandais, l'ajustement formel a également été moins pratiqué que l'ajustement sémantique, mais la différence entre les deux est plus petite qu'en anglais. Enfin, en allemand, l'ajustement formel a été à peine moins pratiqué que l'ajustement sémantique. Nous pouvons en conclure, en gardant à l'esprit que les mots transparents indirectement constituent 32 % du corpus anglais alors qu'ils constituent respectivement 52 % et 51 % des corpus allemand et néerlandais, que l'anglais et le néerlandais nécessitent pour leur compréhension plus de travail sur le sens que sur la forme, et l'allemand autant de travail sur le sens que sur la forme. Les mots étant identifiés par leur versant formel, on peut imaginer que l'anglais et le néerlandais sont d'une approche plus aisée que l'allemand.

Rappelons pour finir l'évolution du taux de transparence pour les trois langues en fonction de la fréquence :



L'hypothèse que nous avons posée, à savoir que moins les mots sont fréquents plus ils sont transparents, ne se vérifie que partiellement. D'une façon générale, elle se vérifie plus pour l'anglais que pour les deux autres langues, et se vérifie plus pour l'interlexique que pour l'hyperlexique.

3.4. Bilan

Cette troisième partie a été pour nous l'occasion d'appliquer les règles de la transparence que nous avons établies dans la deuxième partie à un corpus. Ce corpus, constitué grâce à la base de données CELEX Database, comprend les 1 000 mots lexicaux les plus fréquents de l'anglais, de l'allemand et du néerlandais. Pour pouvoir appliquer les règles formulées dans la deuxième partie, nous les avons adaptées à une analyse hors texte d'un corpus de mots. Cette méthodologie hors texte fait appel aux outils que sont les dictionnaires. Nous avons utilisé des dictionnaires bilingues pour établir la transparence directe et des dictionnaires bilingues ainsi qu'un dictionnaire monolingue pour établir la transparence indirecte. Les dictionnaires bilingues, en proposant des équivalents pour les mots du corpus, nous ont permis de vérifier si ces mots étaient transparents directement. Pour qu'un mot soit transparent directement, il faut qu'il soit proche formellement directement et proche sémantiquement directement d'un mot français. Puisque les équivalents donnés dans les dictionnaires sont proches sémantiquement directement du mot étranger considéré, si un de ces équivalents était, en outre, proche formellement directement du mot étranger, alors ce mot étranger était considéré comme transparent directement. Pour qu'un mot soit transparent indirectement, il faut qu'il y ait :

- i) proximité formelle directe et proximité sémantique indirecte, ou bien
- ii) proximité formelle indirecte et proximité sémantique directe, ou bien
- iii) proximité formelle indirecte et proximité sémantique indirecte.

Les seuls dictionnaires bilingues nous ont également permis d'établir la transparence indirecte selon le point ii). Puisque les équivalents donnés dans les dictionnaires sont proches sémantiquement directement du mot étranger considéré, si un de ces équivalents était, en outre, proche formellement indirectement du mot étranger, alors ce mot étranger était considéré comme transparent indirectement. En revanche, pour évaluer la proximité sémantique indirecte, nous avons également utilisé le dictionnaire monolingue du français qu'est le *Nouveau Petit Robert 2007* électronique. La proximité sémantique indirecte entre deux mots se définissant intralinguistiquement par le partage d'au moins un sème par les sèmes de ces mots, le dictionnaire monolingue nous a permis de chercher si d'autres mots du français étaient proches sémantiquement indirectement de l'équivalent considéré tout en étant proches formellement du mot étranger. Si cet autre mot du français, c'est-à-dire ce mot-relais, était proche formellement directement du mot étranger, alors le mot

étranger était considéré comme transparent indirectement selon le point i), et si le mots-relais était proche formellement indirectement du mot étranger, alors le mot étranger était considéré comme transparent indirectement selon le point iii).

La proximité formelle a été déterminée par le critère des premières lettres en commun ou par celui de la trame consonantique. Si la trame consonantique était identique entre les deux mots considérés, ces derniers étaient considérés comme proches formellement directement, et si la trame consonantique était identique après au moins un ajustement formel, alors les mots étaient considérés comme proches formellement indirectement. C'est l'application stricte de ces règles de proximité formelle, développées en détail dans les sections 2.2.1. et 2.3.1., qui nous ont permis de déclarer deux mots proches formellement.

Ce critère de la trame consonantique a par exemple permis de considérer comme proches formellement directement des mots tels que EN *fact*/FR « fait », DE *kurz*/FR « court » ou NL *spreken*/FR « prêcher », et comme proches formellement indirectement des mots tels que EN *house*/FR « hutte », DE *Gemeinde*/FR « commune » ou NL *proberen*/FR « éprouver ».

Cependant, puisque notre analyse a été réalisée en suivant scrupuleusement les règles, la qualification de mots comme proches formellement parfois a pu sembler contraire à l'intuition. Par exemple, on pourrait être tenté de qualifier les verbes EN *have*, DE *haben* et NL *hebben* de proches formellement du verbe français « avoir », mais le <h> initial, qui se prononce dans les trois langues germaniques, empêche l'identité de la trame consonantique. Rappelons toutefois que les critères de proximité formelle directe ont été établis par observation de paires de mots données dans un dictionnaire de faux amis, donc par observation de mots qui ont pour caractéristique principale d'être proches du point de vue de la forme.

De même, la proximité sémantique indirecte a été déterminée par l'identification de champs sémantiques grâce au *Petit Robert*. Ainsi, nous avons pu déclarer transparents indirectement des mots tels que EN *like*/FR « aimer » par le biais de « languir », DE *kalt*/FR « froid » par le biais de « gelé » ou NL *weten*/FR « savoir » par le biais d'« averti ». Cependant, parce que les mots-relais obtenus sont le résultat de l'application stricte des principes énoncés, des relations sémantiques entre mots français ont également parfois pu sembler étonnantes. Par exemple, nous ne pouvons affirmer qu'un francophone

comprenne le mot EN *game* en pensant au mot « gambit », proche sémantiquement indirectement de « jeu », ou qu'il mette en relation « marge » (proche formellement de DE *mögen*) et « possibilité », ou bien « nature » (proche formellement de NL *nodig*) et « nécessaire ».

Ces quelques précisions sur les tableaux des interlexiques et hyperlexiques, présents dans leur intégralité dans les annexes, ne doivent pas remettre en cause les règles de transparence établies dans la deuxième partie. En effet, les résultats obtenus dans cette troisième partie correspondent à l'application de principes bien précis à un corpus de mots hors texte, analysés grâce à des dictionnaires. Ils ne rendent donc pas compte du processus de compréhension tel qu'il est mis en œuvre par des apprenants. D'abord parce que les apprenants lisent des textes, et non une suite de mots isolés ; leur compréhension est donc éminemment guidée par le contexte. Ensuite parce que les dictionnaires ne sont pas le reflet de la connaissance lexicale des locuteurs ; ces derniers pourraient donc songer à d'autres voies que celles exposées dans les tableaux ou ne pas songer à certaines autres qui y figurent.

Cette troisième partie avait donc uniquement pour but de nous permettre d'observer les résultats obtenus après l'application des principes de la transparence, adaptés à une analyse hors texte, donc artificielle, à un corpus de 3 000 mots.

Les trois principaux résultats que nous retenons de cette analyse sont les suivants :

- l'anglais possède un taux de transparence très important par rapport au français ; un peu plus de la moitié du corpus anglais est transparente directement et 8,5 mots sur 10 sont transparents, directement et indirectement, par rapport au français.

- l'allemand et le néerlandais possèdent des taux de transparence très semblables par rapport au français ; les taux de transparence directe sont respectivement de 22 % et de 21 %, et les taux de transparence globale sont respectivement de 74 % et de 72 %.

- l'allemand et le néerlandais se différencient par le recours aux deux types d'ajustements dans la transparence indirecte ; si la moitié du corpus de l'allemand a pu être considérée comme transparente, c'est grâce aux ajustements formels tout autant qu'aux ajustements sémantiques, alors que la moitié du corpus du néerlandais qui a pu être considérée comme transparente l'a été majoritairement grâce aux ajustements sémantiques.

Il nous semble délicat d'avancer des hypothèses après examen de ces résultats, qui ne rendent compte que de la transparence des trois lexiques germaniques considérés à travers le filtre d'un certain nombre de restrictions (nous avons travaillé à partir d'un corpus préétabli, dont nous avons, pour chaque langue, sélectionné les 1 000 mots lexicaux les plus fréquents, et avons ensuite analysé ces mots selon une méthodologie donnée). Les hypothèses qui suivent sont donc à considérer avec prudence, car elles sont entièrement dépendantes du corpus et des règles avec lesquelles il a été analysé.

Le taux de transparence est tel en anglais qu'il conviendrait de repenser l'enseignement de cette langue. Le lexique passerait du statut d'objet à apprendre par cœur dans le but de comprendre des textes à celui d'aide au lecteur pour la construction du sens de textes. La pratique de l'intercompréhension se trouve ainsi encouragée parce que grandement facilitée par le nombre de mots transparents. Le taux de transparence, bien que moindre, pourrait également être mis à profit pour l'apprentissage de l'allemand et du néerlandais. De plus, nous avons vu pour l'anglais et le néerlandais que les mots les moins transparents étaient aussi les plus fréquents ; l'obstacle qu'ils constituent se trouve donc compensé par le nombre de fois où ils sont rencontrés dans des textes par les apprenants, c'est-à-dire par leur mémorisation sans doute rapide.

Nous avons annoncé au début de cette troisième partie que la langue la plus apprise par les Français parmi les trois langues germaniques sur lesquelles nous travaillons était l'anglais ; arrivait ensuite l'allemand et loin derrière le néerlandais. Notre analyse tend à permettre une réflexion sur ce choix d'apprentissage. En effet, nous avons établi un classement des trois langues vis-à-vis du français en fonction de leur facilité d'accès. L'anglais est la langue la plus transparente, et l'allemand et le néerlandais ont ensuite des valeurs similaires. Cependant, le néerlandais se distingue de l'allemand en ce qu'il semble plus proche du français du point de vue la forme. En effet, nous avons vu que, à proportion de mots transparents indirectement égale, nous avons eu recours plus souvent à l'ajustement formel en allemand qu'en néerlandais. Puisque les mots sont identifiés par les apprenants par leur versant formel, et que les mots du néerlandais sont plus proches formellement du français que ceux de l'allemand (c'est-à-dire qu'ils sont plus souvent proches formellement directement en néerlandais et plus souvent proches formellement indirectement en allemand), il semble que le néerlandais soit une langue d'un accès plus simple pour un francophone que ne l'est l'allemand. Cette constatation rejoint les observations des membres du programme ICE quant à la facilité des apprenants à comprendre les textes au début des séances de langues germaniques. Il a en effet été

remarqué que le néerlandais était d'un abord bien plus aisé pour les francophones que ne l'était l'allemand. On pourrait alors imaginer que le néerlandais soit promu au rang des langues proposées par les établissements de l'enseignement secondaire de toute la France, au même titre que l'anglais, l'allemand et l'espagnol.

Conclusion

Le but que nous avons fixé à notre étude était de proposer une définition objective de la transparence lexicale, conçue comme la qualité attribuable aux mots compréhensibles par des lecteurs face à un texte rédigé dans une langue autre que leur langue maternelle. Sont ainsi qualifiés de « transparents » non seulement les mots immédiatement compréhensibles par des lecteurs débutants, et qui ressortissent dans ce cas de ce que nous avons appelé la « transparence directe », mais aussi les mots qui demandent une capacité de lecture souple, approximative, pour être identifiés, et qui relèvent alors de la « transparence indirecte ». Nous avons défini chacun de ces deux types de transparence par leurs versants formel et sémantique : nous avons ainsi déterminé ce que nous entendions par « proximité formelle » et par « proximité sémantique », ces proximités pouvant en outre être qualifiées de « directes » ou « indirectes ». La transparence directe se définit donc par une proximité formelle directe et une proximité sémantique directe, et la transparence indirecte par une proximité formelle directe et une proximité sémantique indirecte, ou par une proximité formelle indirecte et une proximité sémantique directe, ou bien par une proximité formelle indirecte et une proximité sémantique indirecte.

Nous avons commencé cette étude en présentant les différents travaux qui se sont concentrés sur la transparence, tant pour utiliser la notion que pour en proposer une nouvelle définition. Les premiers travaux utilisaient ou cherchaient à définir la transparence conçue comme « identité » ou « quasi-identité » entre deux mots de langues différentes. Il en ressortait une vision binaire de la transparence : soit les mots étaient transparents, soit ils ne l'étaient pas. Nous nous sommes alors penchée sur des études proposées par des équipes travaillant sur l'intercompréhension. En développant ce concept d'intercompréhension, nous avons pu faire ressortir des points qui se sont avérés importants pour la suite de notre travail (tels que la notion d'approximation ou l'importance du contact des langues). Dans ces nouvelles approches, la transparence était considérée comme un phénomène plus complexe, puisque certains mots pouvaient, par des adaptations diverses, passer du statut de mot opaque à celui de mot transparent. Ces constatations, ainsi que notre propre expérience de la lecture « intercompréhensive », nous ont conduite à proposer, dans la deuxième partie de notre étude, une vision étendue de la

transparence, non limitée aux mots strictement identiques ou quasiment identiques. La transparence, telle que nous la concevons, englobe ainsi les mots traditionnellement considérés comme transparents et ceux qui demandent des manipulations sur leur forme et/ou leur sens pour être identifiés.

Nous avons fixé à notre objectif de définition de la transparence les trois critères suivants : l'objectivité, la considération de la seule synchronie actuelle et l'établissement d'un seuil de transparence, permettant de fixer une limite entre la transparence et l'opacité, et, à l'intérieur de la notion de transparence, entre la transparence directe et la transparence indirecte. Pour définir la transparence directe dans son versant formel, nous avons analysé un dictionnaire de faux amis, où sont précisément recensés des mots qui n'ont d'autre particularité que d'être proches du point de vue de la forme (le fait qu'ils aient des significations plus ou moins proches étant secondaire). Il est ressorti de cette analyse que la proximité formelle directe entre deux mots se définit par une identité des trames consonantiques ou par une identité des premières lettres. La transparence directe a été définie dans son versant sémantique par l'application d'un principe intralinguistique à un phénomène interlinguistique, à savoir la synonymie, ou parasynonymie ; elle est ainsi entendue comme une équivalence de mots en texte. Pour définir la transparence indirecte, nous avons développé les méthodes de l'ajustement formel et de l'ajustement sémantique, le premier consistant en une substitution réglée de phonèmes consonantiques et le second en une sélection de sèmes adéquats au contexte pour le choix d'un mot en langue maternelle différent de celui suggéré formellement par le mot étranger.

Après avoir défini la transparence en texte, nous avons ensuite appliqué les principes dégagés à une analyse de corpus de mots, donc hors texte. Les corpus de chaque langue, établis à partir de la base de données CELEX Database, sont constitués des 1 000 mots lexicaux les plus fréquents. La troisième partie de notre étude a ainsi consisté en la présentation de la méthodologie choisie pour analyser le corpus, c'est-à-dire de la méthodologie choisie pour déterminer si les mots anglais, allemands et néerlandais du corpus devaient être considérés comme transparents ou non, et en l'analyse quantitative et qualitative des résultats obtenus.

Il apparaît ainsi qu'un peu plus de la moitié du corpus anglais et environ 1/5 des corpus allemand et néerlandais relèvent de la transparence directe. Grâce à la prise en compte des ajustements formels et sémantiques, ce sont respectivement 85 %, 74 % et 72 % des corpus anglais, allemand et néerlandais qui deviennent transparents. Dans les trois langues, ce sont les noms qui représentent la plus grande partie des mots transparents.

Nos analyses nous ont permis d'ébaucher un classement des trois langues germaniques par rapport au français en fonction de leur facilité d'accès. L'anglais serait la langue la plus facile d'accès pour un francophone, suivie du néerlandais et en enfin de l'allemand. Ce classement tient au nombre d'ajustements formels pratiqués sur les corpus des trois langues. En effet, moins les mots d'une langue demandent à être modifiés du point de vue de la forme pour être mis en relation avec des mots français, plus ils sont facilement identifiables. L'identification d'un mot passant d'abord par sa forme, nous faisons l'hypothèse que les taux de proximité formelle directe de l'anglais, du néerlandais et de l'allemand puissent correspondre à leur facilité d'accès par des apprenants francophones.

Les listes de mots transparents que nous avons dressées dans la troisième partie ne doivent pas être considérées comme l'ensemble des mots dont nous disons qu'ils sont compréhensibles par des francophones. D'abord parce que, comme nous l'avons rappelé, les analyses menées ici ne le sont qu'entre deux langues ; les apprenants, quant à eux, peuvent mettre à profit leurs connaissances dans d'autres langues que leur langue maternelle. Par exemple, l'adjectif NL *alleen* a été considéré par nous comme opaque. Un francophone ayant des notions d'anglais pourrait le mettre en relation avec l'adjectif EN *alone*, et ainsi proposer une interprétation pour le mot néerlandais. En outre, les listes que nous avons données sont uniquement constituées des mots qui respectent les règles de proximités formelle et sémantique établies dans la deuxième partie de notre étude et adaptées pour une méthodologie d'analyse. Ainsi, il est possible que certains mots respectent ces règles et ne soient pour autant pas compréhensibles par des lecteurs. Par exemple, le nom DE *Grenze* a été considéré comme transparent directement par rapport au nom FR « frontière ». Ils sont proches sémantiquement car « frontière » est proposé comme équivalent possible du mot allemand, et ils sont proches formellement car ils ont la même trame consonantique : il y a remplacement de consonnes (**G**renze/fran**t**ière, Gren**z**e/fran**t**ière), mais qui sont au contact d'autres consonnes identiques (**G**renze/fran**t**ière et Gren**z**e/fran**t**ière), et il y a suppression d'une consonne finale en allemand (Grenze/fran**t**ière), mais il y a déjà deux consonnes identiques entre les deux mots (le <r> et le <n>). D'autres exemples comme celui-là illustrent ce que nous disions plus haut : tous les mots étrangers que nous donnons dans la troisième partie ne pourront pas nécessairement être mis en relation avec leur équivalent par des lecteurs. Que ce constat ne voile pas les bénéfices de l'application au corpus des règles dégagées : les tableaux pourront constituer des listes d'exemples réutilisables à des fins pédagogiques ;

un enseignant ou un modérateur de séance de formation à l'intercompréhension pourront ainsi y puiser des voies d'accès à la compréhension de mots des lexiques considérés et les proposer aux apprenants pour les guider. Que ce constat ne voile pas non plus les règles de proximité formelle et de proximité sémantique que nous avons formulées dans la deuxième partie de notre travail : ces règles ont été établies pour rendre compte de la compréhension de textes ; elles sont donc valables pour une lecture de mots situés en texte, c'est-à-dire lorsqu'il y a un contexte comme force unificatrice empêchant la multiplication d'hypothèses.

Nous avons, dans la dernière partie de notre travail, déterminé les interlexiques puis les hyperlexiques anglais-français, allemand-français et néerlandais-français, c'est-à-dire par paires de langues. Toutefois, la méthodologie que nous avons établie peut être appliquée en l'état aux autres paires de langues, pour peu que l'on possède des dictionnaires bilingues (papiers ou électroniques) et des dictionnaires monolingues électroniques qui permettent de lancer des recherches dans l'ensemble des notices. Il serait fort intéressant de mener une telle étude pour mettre au jour l'interlexique et l'hyperlexique des quatre langues et ainsi déterminer les points de passage qui les relient. Il serait également envisageable d'étendre l'aire linguistique considérée, et de prendre également en compte l'ensemble des langues romanes pour déterminer l'hyperlexique de toute l'Europe occidentale.

Nous avons tenté de proposer dans notre travail une approche globale de la transparence, en ce qu'elle permet de rapprocher les langues de toute une aire linguistique. Nous devons ce choix à l'approche simultanée des langues telle qu'elle est proposée dans le cadre du programme ICE. Les langues y sont vues comme des systèmes interagissant, dans lesquels les contacts géographiques jouent un rôle au moins aussi important que la proximité génétique. Les langues de l'Europe nord-occidentale que sont le français, l'anglais, l'allemand et le néerlandais constituent ainsi une zone au sein de laquelle, comme le montre l'expérience ICE, un certain degré d'intercompréhension est possible. La prise en compte du français dans le groupe de langues cité limite nécessairement le nombre d'éléments concernés par cette intercompréhension, mais il est justement intéressant de voir qu'elle autorise tout de même une comparaison conduisant à dégager les éléments communs aux quatre langues. Nous avons à cet effet forgé le concept d'« hyperlangue », qui représente une langue (en tant que système théorique) qui trouverait ses réalisations

dans les langues effectives⁹⁰. Il s'agit de développer l'opposition langue/discours au niveau interlinguistique. De la sorte, tout comme [islam] et [izlam]⁹¹ sont deux réalisations en français du même mot, transcrit à l'écrit <islam>, les noms EN *house*, DE *Haus*, NL *huis* et FR « hutte » seraient les réalisations d'un même « hyperlexème ». L'« hyperlexique » de langues données est ainsi constitué de l'ensemble des hyperlexèmes, unités théoriques qui trouvent leurs réalisations dans les mots effectifs des langues. Les mots qui peuvent être reliés à un même hyperlexème sont transparents. S'ils sont transparents directement, ils sont les représentants d'un « interlexème » (et de ce fait d'un hyperlexème, le niveau hyper- englobant le niveau inter-), et s'ils sont transparents indirectement, ils sont uniquement les représentants d'un hyperlexème. Dans ce dernier cas, il faut alors avoir recours aux notions d'hyperphonème et/ou d'hypersémème, c'est-à-dire pratiquer un ajustement formel et/ou un ajustement sémantique pour accroître leur proximité.

Cette notion d'hyperlangue nous paraît capitale dans le cadre d'une réflexion sur l'intercompréhension. En effet, les langues d'une zone géographique donnée ne sont plus considérées comme des entités indépendantes les unes des autres, mais comme les différentes réalisations possibles d'une hyperlangue fictive. Elles ne sont plus considérées comme des langues étrangères à apprendre, mais comme des langues proches qui présentent suffisamment de ressemblances pour qu'un minimum d'intercompréhension soit possible. Elles sont considérées comme un ensemble de traits communs, incarnés plus ou moins semblablement dans les langues précises. Ainsi, tout comme, à l'intérieur d'une langue, deux mots prononcés avec des réalisations peu différentes ([islam] et [izlam] par exemple) sont considérés comme des réalisations du même mot, deux mots appartenant à des langues différentes peuvent être considérés, si les différences ne sont pas trop importantes, comme des réalisations d'un même mot, un hyperlexème (*house* et « hutte » par exemple). On voit là l'intérêt de l'hyperlexème, qui permet de mettre en relation non seulement des mots très proches les uns des autres, mais aussi des mots qui peuvent être identifiés grâce à des habitudes de lecture en langues étrangères (entre *house* et « hutte » en l'occurrence, grâce à un ajustement formel /s/-/t/). Si l'on part du principe que l'on peut trouver dans les langues de nos voisins les réalisations des « mêmes mots » que ceux qui existent dans notre langue, le concept d'intercompréhension prend toute sa valeur. On voit alors l'importance de former les apprenants à la compréhension *simultanée* de plusieurs langues, ces langues constituant tout un réseau de points de comparaison.

⁹⁰ Il s'agit donc d'un concept bien différent de celui forgé par S. Auroux (voir p. 89).

⁹¹ Nous empruntons cet exemple à B. Laks (2007 : 192).

Nous voulons terminer en proposant deux pistes de recherche ouvertes par notre étude.

La proximité formelle a été définie en considérant les mots comme une suite de lettres et de phonèmes. Il pourrait également être intéressant de mener une étude sur la transparence en prenant en compte des données d'ordre morphologique. En posant l'hypothèse que certains affixes peuvent être identifiés, un mot comme le nom NL *gevoelen* (« sentiment ») pourrait être considéré comme transparent par rapport au français, la mise entre parenthèses du préfixe *ge-* permettant de passer par le mot-relais « feeling ». Notons que ce type de procédure est mis en œuvre par les apprenants qui suivent la méthode ICE.

Il nous semble également que les principes de proximité formelle que nous avons établis pourraient faire l'objet d'une exploitation dans le domaine du TAL (Traitement Automatique des Langues). Nous avons déploré, dans la troisième partie de notre travail, le fait de ne pas posséder de programme informatique permettant de comparer des mots par leur forme. Peut-être les critères de la trame consonantique et des quatre premières lettres identiques (ou moins de quatre si l'un des mots est constitué de moins de quatre lettres) pourraient-ils être mis à profit dans les techniques d'alignement, qui consistent à faire concorder des segments de textes dont les uns sont les traductions des autres.

Malgré sa perfectibilité, cette recherche a été l'occasion de proposer une définition fine de la transparence, libérée d'une évaluation subjective pour la détermination de la proximité formelle et du recours à l'étymologie pour la détermination de la proximité sémantique. Nous l'avons définie en deux temps, selon les degrés de proximité reliant des mots appartenant à des lexiques divers. Certains points de ce travail tels que la mise en évidence de la trame consonantique et la caractérisation des ajustements formel et sémantique, notamment par l'exposé du processus de sélection du ou des sèmes de transfert en fonction des isotopies, participent de cette définition. Nous espérons que ces notions sauront trouver un écho et seront utiles à d'autres dans leurs recherches théoriques et leurs développements appliqués.

Bibliographie

Les références suivent la présentation donnée par le « Guide pour la rédaction et la présentation des thèses à l'usage des doctorants (2007) », publié par le Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche. Il est disponible à l'adresse suivante : www.sup.adc.education.fr/bib/acti/These/guidoct.rtf

AGNÈS Yves, *Manuel de journalisme*, Paris : La Découverte, 2002, 448 p. (Grands Repères)

ALLINNE Frédéric, *Les faux amis de l'anglais*, Paris : Belin, 1999, 376 p. (Le français retrouvé)

APOTHÉLOZ Denis, *La construction du lexique français*, Paris : Ophrys, 2002, 164 p. (L'essentiel français)

AUROUX Sylvain, *La raison, le langage et les normes*, Paris : PUF, 1998, 337 p. (Sciences, modernités, philosophies)

AUROUX Sylvain, DESCHAMPS Jacques, KOULOUGHLI Djamel, *La philosophie du langage*, Paris : PUF, 2004, 412 p. (Quadrige)

BAAYEN R. Harald, PIEPENBROCK Richard, GULIKERS Leon, *The CELEX Lexical Database (CD-ROM)*, Linguistic Data Consortium, Philadelphia : University of Pennsylvania, PA, 1995

BADER Françoise (dir.), *Langues indo-européennes*, Paris : CNRS Éditions, 1994, 330 p. (Sciences du langage)

BALLARD Michel

- *Les faux amis*, Paris : Ellipses, 1999, 284 p. (Universités)

- Éléments pour la structuration de l'équivalence : point de vue traductologique, *Actes des journées scientifiques 2003 « La traduction : questions d'équivalences. Le sujet syntaxique »*, Reims : PUR, 2005, p. 135-179

BANKS David, *Writing the Sounds of English, A précis of English phonetics, with exercises in phonemic transcription*, 2^e éd., Paris, Budapest, Torino : l'Harmattan, 2005, 162 p.

BARZILAY Regina, MCKEOWN Kathleen R., Extracting Paraphrases from a Parallel Corpus, *Proc. of the ACL/EACL*, 2001, p. 50-57

BEACCO Jean-Claude, Langues et répertoire de langues : le plurilinguisme comme « manière d'être » en Europe, Strasbourg : Conseil de l'Europe, 2005, 24 p.

BENVENISTE Émile

- *Problèmes de linguistique générale*, Paris : Gallimard, 2002, vol. 1, 356 p. (Tel)

- *Problèmes de linguistique générale*, Paris : Gallimard, 2004, vol. 2, 288 p. (Tel)

BERENT Iris, PERFETTI Charles A., A rose is a REEZ: the two-cycles model of phonology assembly in reading English, *Psychological Review*, 1995, n°102(1), p. 146-184

BERTONCINI Josiane, DE BOYSSON-BARDIES Bénédicte, La perception et la production de la parole avant deux ans, *L'acquisition du langage. Le langage en émergence, de la naissance à trois ans*, éd. par Michèle Kail et Michel Fayol, Paris : PUF, 2000, p. 95-136 (Psychologie et sciences de la pensée)

BERTRAND Claude-Jean, LÉVY Claude, *L'anglais de base*, Paris : Éditions du temps, 1999, 412 p.

BESSE Henri, *Méthodes et pratiques des manuels de langue*, Paris : Didier, 1985, 183 p. (Essais)

BLANCHE-BENVENISTE Claire

- Aménagements progressifs de la syntaxe, *Actes du colloque international EuroSem2003 « Premières journées internationales sur l'InterCompréhension Européenne : intercompréhension et inférence »*, Reims : PUR, 2004, p. 41-75 (ICE)

- Formes de compréhension approximative, *Les enjeux de l'intercompréhension*, Reims : Epure, 2007, p. 167-179 (ICE)

BOELHOUWER Bob, *From letter strings to phonemes: the role of orthographic context in phonological recoding*, Nijmegen : Nici, 1998, 175 p.

BOGAARDS Paul, *Le vocabulaire dans l'apprentissage des langues étrangères*, Paris : Hatier/Didier, 1994, 256 p. (Langues et apprentissage des langues)

BOOIJ Geert, *The phonology of Dutch*, Oxford : Clarendon Press, 1995, 205 p.

BÖRESTAM UHLMANN Ulla, Interscandinavian comprehension and scandinavian language community: real or ideal ?, *Pour une modélisation de l'apprentissage simultané de plusieurs langues apparentées à partir de la méthode EuRom4*, éd. par E. Caduc et E. Castagne, Nice : Publications de la Faculté des LASH, 2002

Disponible sur : http://ancilla.unice.fr/~brunet/pub/ulla.html#_ftnref2 (consultée le 19/07/09)

BOUGÉ Patrick, CAILLIES Stéphanie, Compréhension de textes inter-langues et activité inférentielle : approche psychologique, *Actes du colloque international EuroSem2003 « Premières journées internationales sur l'InterCompréhension Européenne : intercompréhension et inférence »*, Reims : PUR, 2004, p. 77-90 (ICE)

BOUQUET Simon, *Introduction à la lecture de Saussure*, Paris : Payot et Rivages, 1997, 396 p. (Bibliothèque scientifique Payot)

BOUSCAREN Christian, RIVIÈRE Claude, *Anglais : les 1 800 mots de base*, Paris : Ophrys, 1993, 154 p.

BRAUN Peter, Internationalismen, *Lexikologie. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen*, éd. par D.A. Cruse, F. Hundsnurscher, M. Job et P.R. Lutzeier, Berlin, New York : De Gruyter, 2002, p. 1380-1384

BRÉAL Michel, *Essai de sémantique, Science des significations*, Genève : Slatkine Reprints, 1976, 372 p.

BRETON Roland, *Atlas des langues du monde*, Paris : Autrement, 2003, 80 p. (Atlas/Monde)

CAID Leila, Les cognates français/anglais, *Ela*, 2008, n°149, p. 65-76

CALVET Louis-Jean

- *La guerre des langues et les politiques linguistiques*, Paris : Hachettes littératures, 1999, 294 p. (Pluriel)

- *Essais de linguistique : La langue est-elle une invention des linguistes ?*, Paris : Plon, 2004, 250 p.

CAMPBELL George L., *Concise compendium of the world's languages*, London, New York : Routledge, 1998, 670 p.

CARTON Fernand, *Introduction à la phonétique du français*, Paris : Bordas, 1974, 250 p.

CASTAGNE Éric

- Comment accéder à l'intercompréhension européenne : quelques pistes inspirées de l'expérience EuRom4, *Ein Kopf – viele Sprache: Koexistenz, Interaktion und Vermittlung*, éd. par J. Müller-Lancé et C.L. Riehl, Aachen : Shaker-Verlag, 2002, p. 99-107

Disponible sur : <http://logatome.eu/publicat/Munich2001.pdf> (consultée le 28/07/09)

- Le programme « Inter Compréhension Européenne » (ICE) ou comment utiliser la linguistique contrastive pour mieux se comprendre en Europe, *Actes du colloque Beiträge zum romanisch-deutschen und innerromanischen Sprachvergleich*, éd. par Schmitt & Wotjak, Bonn : Romanistischer Verlag, 2003, p. 13-36

Disponible sur : <http://logatome.eu/publicat/Leipzig2003.pdf> (consultée le 28/07/09)

- Le Programme InterCompréhension Européenne, *Actes du colloque international organisé le 3 juillet 2003 à l'Université de Reims Champagne-Ardenne « L'avenir du patrimoine linguistique et culturel de l'Europe »*, 2004a, p. 9-12

- Inférences sémantiques et construction de la compréhension en langues étrangères européennes, *Actes du colloque international EuroSem2003 « Premières journées internationales sur l'InterCompréhension Européenne : intercompréhension et inférence »*, Reims : PUR, 2004b, p. 91-115 (ICE)

Disponible sur : <http://logatome.eu/publicat/Reims2003.pdf> (consultée le 06/05/2008)

- Intercompréhension européenne et plurilinguisme : propositions pour quelques aménagements linguistiques favorisant la communication plurilingue, *Neuere Forschungen zur europäischen Interkomprehension*, éd. par Horst G. Klein et Dorothea Rutke, Aachen : Shaker Verlag, 2004c, p. 95-108 (Editiones EuroCom)

- Transparences lexicales entre langues voisines, *Les enjeux de l'intercompréhension*, Reims : Epure, 2007a, p. 155-166 (ICE)

- Modélisation de la formation d'éducateurs à l'intercompréhension de plusieurs langues : réflexions et pistes, *Le français dans le monde*, 2007b, p. 66-75

- Les langues anglaise et française : amies ou ennemies ?, *Ela*, 2008, n°149, p. 31-42

CASTAGNE Éric, CAURE Mélisandre, CHANTEGRAIL Aude, Les langues européennes et l'intercompréhension, *Actes de l'université européenne d'été « Des identités nationales à l'identité européenne »*, vol. I : Sciences du langage, Ecole doctorale 139 : *Connaissance, langage, modélisation*, Publications de l'université de Paris X Nanterre, 2007, p. 21-40

CASTAGNE Éric, CHARTIER Jean-Paul, Modélisation de la formation d'éducateurs à l'intercompréhension de plusieurs langues : réflexions et pistes, *Le français dans le monde*, 2007, p. 66-75

Disponible sur : <http://logatome.eu/publicat/Paris2007.pdf> (consultée le 28/07/09)

CATACH Nina

- *L'orthographe française, Traité théorique et pratique avec des travaux d'application et leurs corrigés*, Paris : Nathan Université, 1986, 334 p.

- *L'orthographe*, Paris : PUF, 1992, 127 p. (Que sais-je ?)

- *Dictionnaire historique de l'orthographe française*, Paris : Larousse, 1995, 1327 p.

CAURE Mélisandre, CHAZAL Tilman, TYVAERT Jean-Emmanuel, Le Programme Sapir, *Les langues modernes* [en ligne], 2008, n°1/2008

CAVALLI-SFORZA Luca, *Gènes, peuples et langues*, Paris : Odile Jacob, 1996, 322 p. (Travaux du Collège de France)

CHABIR Ayadi, Fondements du dictionnaire arabe : la racine, *Dictionnaires bilingues : méthodes et contenus*, dir. par Thomas Szende, Paris : Honoré Champion, 2000, p. 117-133 (Bibliothèque de l'INaLF : études de lexicologie, lexicographie et dictionnaire)

CHADELAT Jean-Marc, Du signe au sens : l'adaptation traductive du lexique dans quelques traductions de Shakespeare, *Palimpsestes. De la lettre à l'esprit : traduction ou adaptation ?*, éd. par Christine Raguét, Presses de la Sorbonne Nouvelle, 2004, n°16, p. 85-104

COLLINS Beverley, MEES Inger M., *The Phonetics of English and Dutch*, 5^e éd., Leiden : Brill, 2003, 363 p.

Commission des communautés européennes (cité comme COM)

- Un nouveau cadre stratégique pour le multilinguisme, Bruxelles, 2005, 33 p.

Disponible sur :

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2005:0596:FIN:FR:PDF>

(consultée le 22/07/09)

- Multilinguisme : un atout pour l'Europe et un engagement commun, Bruxelles, 2008, 16 p.

Disponible sur :

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=SEC:2008:2445:FIN:FR:PDF>

(consultée le 22/07/09)

Conseil de l'Europe, *Un cadre européen commun de référence pour les langues – apprendre, enseigner, évaluer*, Paris : Didier, 2001, 192 p.

Disponible sur : http://www.coe.int/t/dg4/linguistic/Source/Framework_FR.pdf (consultée le 23/07/09)

CORBIN Danielle, *Morphologie dérivationnelle et structuration du lexique*, Lille : Presses universitaires de Lille, 1991, 2 vol., 937 p. (Sens et structures)

CORDIER Françoise, De l'intérêt d'une technique d'amorçage sémantique dans l'étude des relations lexicales et conceptuelles, *Actes d'Eurosem 1994 « Questions de méthode et de délimitation en sémantique lexicale »*, pub. par Hiltraud Dupuy-Engelhardt, Reims : PUR, 1996, p. 13-23

COSERIU Eugenio, Vers une typologie des champs lexicaux, *Cahiers de lexicologie*, Paris : Honoré Champion, 1975, n°27, p. 30-51

CRÉPIN André, *Deux mille ans de langue anglaise*, Paris : Nathan Université, 1994, 191 p. (Fac langues étrangères)

CRUSE David Alan, *Lexical semantics*, Cambridge ; New York : Cambridge University Press, 1986, 310 p. (Cambridge textbooks in linguistics)

DEGACHE Christian, *Didactique du plurilinguisme, Travaux sur l'intercompréhension et l'utilisation des technologies pour l'apprentissage des langues*, Mémoire d'habilitation à diriger des recherches, Université Stendhal – Grenoble 3, 2006, 232 p.

Disponible sur : http://www.galanet.eu/publication/fichiers/HDR2006_DegacheC.pdf (consultée le 04/10/09)

DEGACHE Christian, MASPERI Monica, La communication en toile de fond de l'entraînement à la compréhension des langues romanes, *De la didactique des langues à la didactique du plurilinguisme, hommage à Louise Dabène*, éd. par J. Billiez, Lidilem, Grenoble, 1998, p. 361-376

Disponible sur : www.galanet.eu/publication/fichiers/dc-mm1998.pdf (consultée le 28/07/09)

DELESALLE Simone, GARY-PRIEUR Marie-Noëlle, Le lexique, entre la lexicologie et l'hypothèse lexicaliste, *Langue française*, 1976, n°30, p. 4-33

Disponible sur : http://www.persee.fr/web/revues/home/prescript/article/lfr_0023-8368_1976_num_30_1_6109 (consultée le 09/03/09)

DGLFLF (Délégation générale à la langue française et aux langues de France), L'intercompréhension entre langues apparentées [en ligne], 2006, 12 p.

Disponible sur :

<http://www.culture.gouv.fr/culture/dglf/publications/intercomprehension.pdf> (consultée le 19/07/09)

DOYÉ Peter, L'intercompréhension, Strasbourg : Conseil de l'Europe, 2005, 23 p.

Disponible sur : <http://www.coe.int/t/dg4/linguistic/Source/DoyeFR.pdf> (consultée le 23/07/09)

DUBOIS Jean, GIACOMO Mathée, GUESPIN Louis, MARCELLESI Christiane, MARCELLESI Jean-Baptiste, MÉVEL Jean-Pierre, *Dictionnaire de linguistique*, Paris : Larousse, 2002, 514 p.

DUVAL Alain, Le rôle de l'exemple dans le dictionnaire bilingue français-anglais, *Approches contrastives en lexicographie bilingue*, dir. par Thomas Szende, Paris : Honoré Champion, 2000, p. 79-87 (Bibliothèque de l'INALF. Études de lexicologie, lexicographie et dictionnaire)

ELOY Jean-Michel, Langues proches : que signifie de les enseigner ?, *Ela*, 2004, n°136, p. 393-402

FERRAND Ludovic, *Psychologie cognitive de la lecture, reconnaissance des mots écrits chez l'adulte*, Bruxelles : De Boeck, 2007, 537 p. (Ouvertures psychologiques)

FEUILLET Jack, *Grammaire structurale de l'allemand*, Paris : Peter Lang, 1993, 944 p. (Contacts)

FOURQUET Jean, *Grammaire de l'allemand*, Paris : Hachette, 1952, 284 p. (Classiques)

FRANÇOIS Jacques, KEROMNÈS Yvon, De la fidélité en matière de traduction littéraire : esquisse d'une méthode d'évaluation comparative, *Lectures, Hommage à Geneviève Hily-Mane*, Reims : PUR, 1995, p. 41-68

FRADIN Bernard, L'identification des unités lexicales, *Sémiotiques*, 1996, n°11, p. 55-93
Disponible sur : http://www.revue-texto.net/Parutions/Semiotiques/SEM_n11_5.pdf
(consultée le 18/06/09)

FROMNOT Jacqueline, LEGUY Isabelle, FONTANE Gilbert, *Vocabulaire de l'anglais contemporain*, Paris : Nathan, 2009, 432 p. (Le Robert & Nathan)

FRUYT Michèle, Nature et limites de la polysémie, *La polysémie*, éd. par Olivier Soutet, Paris : Presses universitaires de Paris-Sorbonne, 2005, p. 23-36

GABILAN Jean-Pierre, *Grammaire expliquée de l'anglais*, Paris : Ellipses, 2006, 415 p. (Optimum)

GAK Vladimir Grigorevich, VILDÉ-LOT Irène, *L'orthographe du français : essai de description théorique et pratique*, Paris : Société d'études linguistiques et anthropologiques de France, 1976, 318 p.

GARDES-TAMINE Joëlle

- *La grammaire, 1. Phonologie, morphologie, lexicologie*, 3^e éd., Paris : Armand Colin, 2002, 174 p. (Cursus)

- *La grammaire, 2. Syntaxe*, 3^e éd., Paris : Armand Colin, 2004, 192 p. (Cursus)

GARNIER Frédéric R., Introduction : Enseigner et apprendre le lexique, *Les langues modernes*, 2009, n°1/2009, p. 12-15

GAONAC'H Daniel, *Théories d'apprentissage et acquisition d'une langue étrangère*, Paris : Hatier, 1991, 239 p. (Langues et apprentissage des langues)

GENETTE Gérard, *Seuils*, Paris : Éditions du Seuil, 1987, 389 p. (Poétique)

GERMAIN Claude, *Évolution de l'enseignement des langues : 5 000 ans d'histoire*, Paris : CLE international, 1993, 351 p. (Didactique des langues étrangères)

GEYSEN Raymond, *Dictionnaire des formes analogues en 7 langues, Latin, Italien, Espagnol, Français, Anglais, Néerlandais, Allemand, avec résumé de grammaire comparée*, 2^e éd., Paris : Duculot, 1990, 831 p.

GIASSON Jocelyne, *La compréhension en lecture*, Bruxelles : De Boeck, 2005, 272 p.

GINÉSY Michel, *Mémento de phonétique anglaise avec exercices corrigés*, Paris : Armand Colin, 2004, 288 p. (Fac Anglais)

GOFFIN Roger, Le traitement des faux amis dans le dictionnaire bilingue allemand-français de Sachs et Villatte : un piège insidieux en lexicographie contrastive, *Approches contrastives en lexicographie bilingue*, dir. par Thomas Szende, Paris : Honoré Champion, 2000, p. 71-78 (Bibliothèque de l'INALF. Études de lexicologie, lexicographie et dictionnaire)

GOGOLIN Ingrid, *Diversité linguistique et nouvelles minorités en Europe*, Strasbourg : Conseil de l'Europe, 2002, 22 p.

Disponible sur : <http://www.coe.int/t/dg4/linguistic/Source/GogolinFR.pdf> (consultée le 01/08/09)

GOOSKENS Charlotte, HEERINGA Wilbert, Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data, *Language variation and Change*, 2004, n°16(3), p. 189-207

Disponible sur : <http://www.let.rug.nl/~heeringa/dialectology/papers/lvc04.pdf> (consultée le 26/05/08)

GOOSKENS Charlotte, VAN BEZOOIJEN Renée, Linguistic distance between Flemish and Norwegian around 1900: subjective views and objective measurements, *Symposium on Ethnolinguistic Nationalism in the Dutch-speaking Region and Scandinavia*, Groningen, 2005.

Disponible sur : <http://www.let.rug.nl/%7Egooskens/pdf/Ethnolinguistics.pdf> (consultée le 26/05/08)

GRADDOL David, *English next*, London : British Council, 2006, 128 p.

Disponible sur : <http://www.britishcouncil.org/learning-research-english-next.pdf> (consultée le 11/10/09)

GRAINGER Jonathan, Visual word recognition in bilinguals, *The bilingual lexicon*, éd. par R. Schreuder and B. Weltens, Amsterdam : John Benjamins, 1993, p. 11-25

GREIMAS Algirdas Julien, COURTÉS Joseph, *Sémiotique : dictionnaire raisonné de la théorie du langage*, Paris : Hachette, 1979, 424 p. (Hachette Université ; Langue, linguistique, communication)

HAEUSSER Christiane, HIRST Daniel, Transparence/opacité : contribution à une étude de la contrastivité lexicale français-anglais, dans une perspective didactique, *Contrastes*, 1987, n°14-15, p. 279-283

HAGÈGE Claude

- *L'enfant aux deux langues*, Paris : Odile Jacob, 1996a, 298 p.

- *L'homme de paroles : contribution linguistique aux sciences humaines*, Paris : Fayard, 1996b, 316 p. (Le temps des sciences)

- *Le souffle de la langue*, Paris : Odile Jacob, 2000, 288 p.

- Le plurilinguisme européen, *Actes du colloque international organisé le 3 juillet 2003 à l'Université de Reims Champagne-Ardenne « L'avenir du patrimoine linguistique et culturel de l'Europe »*, 2004, p. 35-38

- *La structure des langues*, 6^e éd., Paris : PUF, 2007, 127 p. (Que sais-je ?)

HARM Michael W, SEIDENBERG Mark S., Phonology, reading acquisition, and dyslexia: Insights from connectionist models, *Psychological Review*, 1999, n°106(3), p. 491-528

HAUSMANN Franz Josef, La transparence et l'obstacle. Essai de chrestolexicographie, *Ela*, 2002, n°128, p. 447-454

Disponible sur : http://www.cairn.info/article.php?ID_ARTICLE=ELA_128_0447
(consultée le 18/06/08)

HÉDIARD Marie, Quand la transparence risque d'être trompeuse : *Pas tout à fait d'accord ou pas d'accord du tout*, *Lidil*, 2003, n°28 [En ligne]

Disponible sur : <http://lidil.revues.org/index1943.html> (consultée le 15/06/09)

HEEMSKERK Josée, ZONNEVELD Wim, *Uitspraakwoordenboek*, Utrecht : Het Spectrum, 2000, 858 p.

HÉLOT Christine, *Du bilinguisme en famille au plurilinguisme à l'école*, Paris : L'Harmattan, 2007, 282 p.

HENRARD Roger, *Grammaire du néerlandais*, 6^e éd., Anvers : De Sikkel, 1985, 157 p.

HERBILLON Michel, Rapport d'information déposé par la délégation de l'assemblée nationale pour l'union européenne sur la diversité linguistique dans l'union européenne [en ligne], 2003

Disponible sur : <http://www.assemblee-nationale.fr/12/europe/rap-info/i0902.asp>
(consultée le 22/07/09)

ISABELLE Pierre, La bi-textualité : vers une nouvelle génération d'aides à la traduction et la terminologie, *Meta*, XXXVII, 1992, n°4, p.721-731

JAKOBSON Roman, *Essais de linguistique générale*, Paris : Les Éditions de Minuit, 1963, 259 p.

JONES Daniel, ROACH Peter, HARTMAN James, *English Pronouncing Dictionary*, Cambridge, New York : Cambridge University Press, 1997, 559 p.

JOUVE Vincent, *Poétique du roman*, 2^e éd., Paris : Armand Colin, 2007, 238 p. (Cursus Lettres)

KAHANE Sylvain, Les unités minimales de la syntaxe et de la sémantique : le cas du français [en ligne], *Actes du Congrès Mondial de Linguistique Française*, éd. par Jacques Durand, Benoît Habert et Bernard Laks, 2008

Disponible sur :

<http://www.linguistiquefrancaise.org/index.php?option=article&access=standard&Itemid=129&url=/articles/cmlf/pdf/2008/01/cmlf08106.pdf> (consultée le 11/10/09)

KAPPLER Claire, MÉJEAN-THIOLIER Suzanne, *Le plurilinguisme au Moyen-Âge : de Babel à la langue une : Orient-Occident*, Paris : L'Harmattan, 2009, 372 p.

KERBRAT-ORECCHIONI Catherine, *La connotation*, 3^e éd., Lyon : Presses universitaires de Lyon, 1984, 256 p.

KESSLER Brett, Computational dialectology in Irish Gaelic, *Proceedings of the European Association for Computational Linguistics*, Dublin : ACL, 1995, p. 60-67

Disponible sur : http://arxiv.org/PS_cache/cmp-lg/pdf/9503/9503002v1.pdf (consultée le 11/10/09)

KICHENASSAMY Satyanad, La compréhension inter-linguistique en Inde, *Actes du colloque international organisé le 3 juillet 2003 à l'Université de Reims Champagne-Ardenne « L'avenir du patrimoine linguistique et culturel de l'Europe »*, 2004, p. 45-49

KLEIBER Georges

- Quelques réflexions sur le vague dans les langues naturelles, *Etudes de linguistique générale et de linguistique latine, offertes en hommage à Guy Serbat*, Paris : Société pour l'information grammaticale, 1987, p. 157-172

- *La sémantique du prototype, Catégories et sens lexical*, Paris : PUF, 1990, 199 p. (Linguistique nouvelle)

- Sens, référence et existence : que faire de l'extra-linguistique ?, *Langages*, 1997, n°127, p. 9-37

- *Problèmes de sémantique : la polysémie en questions*, Villeneuve d'Ascq : Presses universitaires du Septentrion, 1999, 220 p. (Sens et structures)

- Petit essai pour montrer que la polysémie n'est pas un sens interdit, *Congrès Mondial de Linguistique Française*, Paris, 2008

Disponible sur : <http://www.linguistiquefrancaise.org/index.php?option=article&access=standard&Itemid=129&url=/articles/cmlf/pdf/2008/01/cmlf08341.pdf> (consultée le 16/06/09)

KLEIN Horst G., RUTKE Dorothea, *EuroComRom : pour un plurilinguisme européen*, *Sociolinguistica*, 1997, n°11, p. 178-183

Disponible sur : <http://www.eurocomresearch.net/lit/socioling.htm> (consultée le 15/06/09)

KLEIN Horst G., STEGMANN Tilbert D., *EuroComRom – Die sieben Siebe : Romanische Sprachen sofort lesen können*, Aachen : Shaker Verlag, 2000, 288 p. (Editiones EuroCom)

KOHLER Klaus J., *Einführung in die Phonetik des Deutschen*, Berlin : E. Schmidt, 1995, 249 p. (Grundlagen der Germanistik)

KONDRAK Grzegorz, Identifying Cognates by phonetic and Semantic Similarity, *NAACL*, Pittsburgh, Carnegie Mellon University, 2001, p. 103-110

Disponible sur : <http://www.aclweb.org/anthology-new/N/N01/N01-1014.pdf> (consultée le 30/05/08)

KÖNIG Ekkehard, VAN DER AUWERA Johan, *The Germanic Languages*, London, New York : Routledge Reference, 1994, 631 p.

KRAIF Olivier

- Identification des cognats et alignement bi-textuel : une étude empirique, *Actes de la 6ème conférence annuelle sur le Traitement Automatique des Langues Naturelles, TALN, Cargèse, 12-17 juillet 1999*, 1999, p. 205-214

Disponible sur : http://www.atala.org/doc/actes_taln/AC_0008.pdf (consultée le 06/05/08)

- Repérage de traduction et commutation interlingue : Intérêt et méthodes, *Actes de TALN 2003, Batz-sur-Mer, 11-14 juin 2003*, 2003, tome 2, p. 127-138

Disponible sur : <http://w3.u-grenoble3.fr/lidilem/labo/file/Kraif2003TALN.pdf> (consultée le 11/01/09)

LAKS Bernard

- De la variation et des variantes : à propos du relâchement, *Linx*, 2000, n°42, p. 21-29

- Continu et discontinu : l'articulation phonétique/phonologie, *Cahiers de proxématique*, 2004, n°42, p. 145-175

- De l'approximation dans la relation phonétique/phonologie, *Les enjeux de l'intercompréhension*, Reims : Epure, 2007, p. 181-200 (ICE)

LARREYA Paul, RIVIÈRE Claude, *Grammaire explicative de l'anglais*, 3^e éd., Paris : Pearson, 2007, 445 p.

LARREYA Paul, WATBLED Jean-Philippe, *Linguistique générale et langue anglaise*, Paris : Nathan Université, 2003, 127 p. (128)

LEDEGEN Gudrun, ROSSI Nathalie (dir.), *Les grammaires du français et les « mots outils »*, Caen : Presses universitaires de Caen, 2002, 150 p.

LEDERER Marianne, *La traduction aujourd'hui, le modèle interprétatif*, Caen : Lettres modernes minard, 2006, 196 p. (Cahiers Champollion)

LEHMANN Alise, MARTIN-BERTHET Françoise, *Introduction à la lexicologie, Sémantique et morphologie*, 2^e éd., Paris : Armand Colin, 2005, 214 p. (Lettres supérieures)

LE NY Jean-François, *Comment l'esprit produit du sens*, Paris : Odile Jacob, 2005, 416 p.

LÉON Monique, LÉON Pierre, *La prononciation du français*, Paris : Armand Colin, 2004, 128 p. (128)

LÉON Pierre, *Phonétisme et prononciations du français avec des travaux pratiques d'application et leurs corrigés*, 2^e éd., Paris : Nathan université, 1996, 192 p. (Fac linguistique)

Le Robert

- *Le Nouveau Petit Robert, Dictionnaire de la langue française*, Paris : Dictionnaires Le Robert, 2002, 2843 p.

- *Le Nouveau Petit Robert de la langue française*, version électronique, 2007

Le Robert & Collins, *Dictionnaire français-anglais et anglais-français*, 8^e éd., Londres : Collins, Paris : Dictionnaires Le Robert, 2007, 2285 p.

Le Robert & Van Dale, *Dictionnaire français-néerlandais et néerlandais-français*, 4^e éd., dir. par Paul Bogaards, Utrecht/Antwerpen : Van Dale Lexicografie, Paris : Dictionnaires Le Robert, 2007, 875 p.

LEROT Jacques, *Précis de linguistique générale*, Paris : Éditions de Minuit, 1993, 446 p.

L'HOMME Marie-Claude, DANCETTE Jeanne

- Modélisation des relations sémantiques dans un dictionnaire spécialisé bilingue, *L'éloge de la différence : la voix de l'autre. Actes des 6^e Journées du Réseau LTT*, éd. par A. Clas *et al*, 2001, Paris : AUF, p. 385-400 (Actualité scientifique)

Disponible sur : <http://www.ling.umontreal.ca/lhomme/docs/beyroutl.pdf> (consultée le 27/09/08)

- Acquisition de liens conceptuels entre termes à partir de leur définition, *Cahiers de lexicologie*, 2003, n°83, p. 25-48

LOFFLER-LAURIAN Anne-Marie, Les apports de la méthodologie contrastive à la lexicographie bilingue, *Approches contrastives en lexicographie bilingue*, dir. par Thomas

Szende, Paris : Honoré Champion, 2000, p. 135-146 (Bibliothèque de l'INALF. Études de lexicologie, lexicographie et dictionnaire)

LUKATELA Georgije, EATON Thomas, LEE Chang, TURVEY Michael T., Does visual word identification involve a sub-phonemic level?, *Cognition*, 2000, n°78, p. 41-52

LUPU Mihaela, Concepts vagues et catégorisation, *Cahiers de Linguistique Française*, 2003, n°25, p. 291-304

Disponible sur : <http://clf.unige.ch/display.php?idFichier=70> (consultée le 06/05/08)

LYONS John, *Linguistique générale, introduction à la linguistique théorique*, Paris : Larousse, 1970, 384 p. (Langue et langage)

MACKEY William Francis, *Bilinguisme et contact des langues*, Paris : Klincksieck, 1976, 539 p. (Initiation à la linguistique)

MALAVIEILLE Michèle, ROTGÉ Wilfrid, *Anglais : la grammaire*, Paris : Hatier, 2008, 432 p. (Bescherelle)

MALHERBE Michel, *Les langages de l'humanité, Une encyclopédie des 3000 langues parlées dans le monde*, Paris : Seghers, 1983, 444 p.

MARTIN Robert, *Pour une logique du sens*, Paris : PUF, 1983, 268 p. (Linguistique nouvelle)

MARTINET André, *Éléments de linguistique générale*, 5^e éd., Paris : Armand Colin, 2008, 223 p. (Cursus linguistique)

MARTINEZ Pierre, *La didactique des langues étrangères*, 4^e éd., Paris : PUF, 2004, 128 p. (Que sais-je ?)

MCKOON Gail, RATCLIFF Roger, Inference during reading, *Psychological review*, 1992, n°99, p. 440-466

Disponible sur :

<http://star.psy.ohio-state.edu/coglab/People/roger/pdf/Papers/psychrev92a.pdf> (consultée le 25/07/09)

MÉDIONI Maria-Alice, L'acquisition du vocabulaire : encore une question d'activité, *Les langues modernes*, 2009, n°1/2009, p. 16-25

MEISSNER Franz-Joseph, Que peut apporter la didactique de l'intercompréhension aux systèmes éducatifs européens ?, *Les langues modernes*, 2008, n°1/2008, p. 15-24

MEISSNER Franz-Joseph, MEISSNER Claude, KLEIN Horst G., STEGMANN Tilbert, *Introduction à la didactique de l'eurocompréhension, EuroComRom. Les sept tamis. Lire les langues romanes dès le début*, Aachen : Shaker-Verlag, 2004, 70 p.

Disponible sur : <http://www.uni-giessen.de/meissner/plurling/esquisse.pdf> (consultée le 06/05/08)

MELAMED I. Dan, Bitext maps and alignment via pattern recognition, *Computational Linguistics*, 1999, n°25(1), p. 107-130

MICHELOT Marius, *Les systèmes sténographiques*, Paris : PUF, 1959, 128 p. (Que sais-je ?)

MONSELL Stephen, The nature and locus of word frequency effects in reading, *Basic processes in reading : Visual word recognition*, éd. par D. Besner et G. Humphreys, Hillsdale (Nj), Lawrence Erlbaum, 1991, p. 148-197

MORTUREUX Marie-Françoise, *La lexicologie entre langue et discours*, Paris : Armand Colin, 2001, 192 p. (Campus Linguistique)

MOUNIN Georges, *Les problèmes théoriques de la traduction*, Paris : Gallimard, 1976, 296 p. (Tel)

MULLER François, *Grammaire de l'allemand*, Paris : Nathan, 2001, 408 p. (Le Robert et Nathan Langues actuelles)

MURILLO PUYAL Julio, PICCALUGA Myriam, HARMEGNIES Bernard, Pour un français partenaire dans l'aire romane : aspects psycholinguistiques et didactiques de l'intercompréhension entre locuteurs de langues latines, *S'entendre entre langues voisines : vers l'intercompréhension*, éd. par Virginie Conti & François Grin, Genève : Georg, 2008, p.341-370

NAMER Fiammetta, Productivité morphologique, représentativité et complexité de la base : le système MoQuête, *Langue française*, 2003, n°140, p. 79-101

NERBONNE John, HEERINGA Wilbert, KLEIWEG Peter, Edit distance and dialect proximity, *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*, éd. par David Sankoff et Joseph Kruskal, Stanford : CSLI, 1999, p. 5-15

Disponible sur : <http://www.let.rug.nl/~heeringa/dialectology/papers/eddp99.pdf> (consultée le 11/10/09)

NÈVE François-Xavier, *Essai de grammaire de la langue des signes française*, Bibliothèque de la Faculté de Philosophie et Lettres, Université de Liège, 1996, 472 p.

NEVEU Franck, *Dictionnaire des sciences du langage*, Paris : Armand Colin, 2004, 317 p.

NIKLAS-SALMINEN Aïno, *La lexicologie*, Paris : Armand Colin, 1997, 188 p. (Cursus)

NOIZET Georges, VION Monique, « Les stratégies de compréhension dans le traitement des relations fonctionnelles de base », *Psycholinguistique de l'enfant*, éd. par J.P. Bronckart, M. Kail et G. Noizet, Neuchâtel : Delachaux et Niestlé, 1983, p. 51-72

NYCKEES Vincent

- *La sémantique*, Paris : Belin, 1998, 365 p. (Sujets)

- Les mots, les choses... et nous, *Sciences humaines, Hors-série : Le langage, origine, nature, diversité*, 1999-2000, n°27, p.24-27

ONIONS Charles Talbut, *Oxford dictionary of English etymology*, Oxford : Clarendon press, 1996, 1025 p.

OUSTINOFF Michaël, *La traduction*, Paris : PUF, 2003, 127 p. (Que sais-je ?)

PETTON André, *Les faux-amis anglais en contexte*, Rennes : Presses universitaires de Rennes, 1995, 361 p. (Didact anglais)

PICOCHÉ Jacqueline

- *Précis de lexicologie française, L'étude et l'enseignement du vocabulaire*, Paris : Nathan, 1997, 191 p. (Fac. linguistique)

- Le dictionnaire du français usuel, principes et finalités, *Approches contrastives en lexicographie bilingue*, dir. par Thomas Szende, Paris : Honoré Champion, 2000, p. 199-212 (Bibliothèque de l'INALF. Études de lexicologie, lexicographie et dictionnaire)

PICOCHÉ Jacqueline, MARCHELLO-NIZIA Christiane, *Histoire de la langue française*, 5^e éd., Paris : Nathan, 1998, 399 p. (Fac linguistique)

PHILIPP Marthe, *Phonologie de l'allemand*, Paris : PUF, 1970, 206 p. (SUP Le Linguiste)

POLLOCK Joseph J., ZAMORA Antonio, Automatic spelling correction in scientific and scholarly text, *Communications of the ACM*, 1984, n°27(4), p. 358-368

PORTER Edward H., WINKLER William E., Approximate string comparison and its effect in an advanced record linkage system, *Record Linkage Techniques*, éd. par Alvey and Jamerson, 1999, National Research Council, Washington, D.C : National Academy Press, p. 190-199

Disponible sur : <http://www.fcsn.gov/working-papers/porter-winkler.pdf> (consultée le 06/05/08)

POTTIER Bernard

- Vers une sémantique moderne, *Travaux de linguistique et de littérature*, 1964, n°2, p. 107-137

- *Linguistique générale, Théorie et description*, Paris : Klincksieck, 1974, 340 p.

- *Sémantique générale*, Paris : PUF, 1992, 237 p. (Linguistique nouvelle)

PY Bernard, Quelques remarques sur l'apprentissage du lexique d'une langue étrangère, *Cahiers de lexicologie*, 1997, n°70, p. 175-184

RASTIER François

- *Sémantique interprétative*, Paris : PUF, 1987, 277 p.

- Le problème épistémologique du contexte et le statut de l'interprétation dans les sciences du langage, *Langages*, 1998, n°129, p. 97-111

- La macrosémantique, *Texto !* [en ligne], 2002a

Disponible sur : <http://www.revue-texto.net/index.php?id=548> (consultée le 01/10/08)

- Glossaire-index de notions, *Texto !* [en ligne], 2002b

Disponible sur : <http://www.revue-texto.net/1996-2007/Reperes/Reperes.html> (consultée le 16/02/09)

- De la signification au sens. Pour une sémiotique sans ontologie, *Texto !* [en ligne], juin-sept. 2003

Disponible sur :

http://www.revue-texto.net/Inedits/Rastier/Rastier_Semiotique-ontologie.html (consultée le 05/01/09)

- La microsémantique, *Texto !* [en ligne], juin 2005, vol. X, n°2

Disponible sur : http://www.revue-texto.net/Inedits/Rastier/Rastier_Microsemantique.html (consultée le 06/05/08)

- Du réalisme au postulat référentiel, *Texto !* [en ligne], 2007

Disponible sur : <http://www.revue-texto.net/Inedits/Inedits.html> (consultée le 22/06/08)

RAYNAUD Franziska, *Histoire de la langue allemande*, Paris : PUF, 1993, 128 p. (Que sais-je ?)

REY Alain, TOMI Marianne, HORDÉ Tristan, TANET Chantal, *Dictionnaire historique de la langue française*, Paris : Dictionnaires Le Robert, 2006, 3 vol., 4 304 p.

REY Jean, BOUSCAREN Christian, MOUNOLOU Alain, *Le mot et l'idée : anglais, 2*, Paris : Ophrys, 2003, 391 p.

REY-DEBOVE Josette

- La définition lexicographique : recherches sur l'équation sémique, *Cahiers de lexicologie*, 1966, n°1, p. 71-94

- La synonymie ou les échanges de signes comme fondement de la sémantique, *Langages*, 1997, n°128, p. 91-104

Disponible sur : http://www.persee.fr/web/revues/home/prescript/article/lgge_0458-726x_1997_num_31_128_2135 (consultée le 09/10/09)

REYNÉS Philippe, Approche de l'apprentissage lexical et du rôle du lexique dans plusieurs méthodes et manuels de catalan, *Ela*, 2004, n°136, p. 449-463

RIEGEL Martin, PELLAT Jean-Christophe, RIOUL René, *Grammaire méthodique du français*, 2^e éd., Paris : PUF, 1998, 646 p. (Linguistique nouvelle)

RINGBOM Håkan, *Cross-linguistic similarity in foreign language learning*, Clevedon : Multilingual Matters Ltd., 2007, 144 p. (Second Language Acquisition)

RIVENC Paul, *Apprentissage d'une langue étrangère/seconde : la méthodologie*, Bruxelles : De Boeck, 2003, 384 p.

RIVIÈRE Claude, *Un air de famille, les faux amis français-anglais*, Paris : Ophrys, 2003, 141 p. (Gramvoc anglais)

ROBERT Jean-Michel

- Proximité linguistique et pédagogie des langues non maternelles, *Ela*, 2004a, n°136, p. 499-511

- Langues voisines, langues proches et langues lointaines : implications didactiques [en ligne], Communication présentée au Colloque du LESCLaP : *Langues voisines et langues étrangères : Approches didactiques et sociolinguistiques*, Alliance française de Paris, 2004b

Disponible sur : http://www.u-picardie.fr/LESCLaP/IMG/pdf/artic_robert_2.pdf (consultée le 11/10/09)

ROSENTHAL Victor, VISETTI Yves-Marie, Sens et temps de la Gestalt, *Texto !* [En ligne], 2001

Disponible sur : <http://www.revue-texto.net/index.php?id=625> (consultée le 27/02/09)

ROULET Eddy, *Théories grammaticales : descriptions et enseignement des langues*, Paris : Nathan, 1978, 125 p. (Langues et culture)

RUHLEN Merritt, *L'origine des langues : sur les traces de la langue mère*, Paris : Belin, 1996, 287 p. (Débats)

SACHS Karl, VILLATTE Césaire

- *Langenscheidt Großwörterbuch Deutsch-Französisch*, 18^e éd., Berlin, München : Langenscheidt, 1990, 1080 p.

- *Langenscheidt Großwörterbuch Französisch-Deutsch*, 13^e éd., Berlin, München : Langenscheidt, 1990, 1047 p.

SAUSSURE Ferdinand de

- *Cours de linguistique générale*, Paris : Payot, 1995, 520 p. (Grande Bibliothèque Payot)

- *Écrits de linguistique générale*, éd. par Simon Bouquet et Rudolf Engler, Paris : Gallimard, 2002, 353 p. (Bibliothèque de philosophie)

SCHANEN François, CONFAIS Jean-Paul, *Grammaire de l'allemand, formes et fonctions*, Paris : Nathan Université, 1999, 608 p. (Réf.)

SCHMIDT Johannes, *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*, Weimar : Böhlau, 1872

SÉGUIN Hubert, Les congénères anglais-français. Étude de leur similarité graphique et suggestions d'exploitation pédagogique, *L'écrit en français langue étrangère, réflexions et propositions*, éd. par Claude Buridant et Jean-Christophe Pellat, Strasbourg : Presses universitaires de Strasbourg, 1994, p. 193-199

SIMARD Michel, FOSTER George F., ISABELLE Pierre, Using cognates to align sentences in bilingual corpora, *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation*, Montreal, Canada, 1992, p. 67-81

SOMMERFELT Alf, *Diachronic and synchronic aspects of language*, 2^{ème} éd., 's-Gravenhage : Mouton, 1971, 421 p.

SOUTET Olivier, *La polysémie* (éd.), Paris : Presses universitaires de Paris-Sorbonne, 2005, 456 p.

SZENDE Thomas, L'information sémantique en lexicographie bilingue (hongrois-français), *Dictionnaires bilingues : méthodes et contenus*, dir. par Thomas Szende, Paris : Honoré Champion, 2000, p. 69-81 (Bibliothèque de l'INaLF : études de lexicologie, lexicographie et dictionnaire)

SZULMAJSTER-CELNIKIER Anne, Eloge de la prudence méthodologique, *Recherche*, 1998, n°306, p. 76-81

TESNIÈRE Lucien, *Éléments de syntaxe structurale*, Paris : Klincksieck, 1959, 670 p.

TOURNIER Jean

- *Précis de lexicologie anglaise*, Paris : Ellipses, 2004, 240 p.

- *Introduction descriptive à la lexicogénétiq ue de l'anglais contemporain*, Paris, Genève : Champion-Slatkine, 2007, 517 p.

TRÉVILLE Marie-Claude, *Rôle des congénères interlinguaux dans le développement du vocabulaire réceptif en français langue seconde*, Université de Laval, Québec, 1993, 240 p.

Disponible sur :

http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/13/0a/18.pdf (consultée le 17/10/09)

TRIER Jost, Das sprachliche Feld. Eine Auseinandersetzung, *Neue Jahrbücher für Wissenschaft und Jugendbild*, 1934, n°10, p. 428-449

TROUBETZKOY Nicolas Sergueevitch, *Principes de phonologie*, Paris : Klincksieck, 1970, 396 p. (Tradition de l'humanisme)

TRUCHOT Claude, *Le plurilinguisme européen, Théories et pratiques en politique*, Paris : Honoré Champion, 1994, 423 p. (Politique linguistique)

TYVAERT Jean-Emmanuel

- Pour une Europe des langues et des cultures, *Actes du colloque international organisé le 3 juillet 2003 à l'Université de Reims Champagne-Ardenne « L'avenir du patrimoine linguistique et culturel de l'Europe »*, 2004a, p. 59-65

- Formes linguistiques et inférences dans le contrôle de l'élaboration du sens, *Actes du colloque international EuroSem2003 « Premières journées internationales sur l'InterCompréhension Européenne : intercompréhension et inférence »*, Reims : PUR, 2004b, p. 255-276 (ICE)

- La raison des langues et la double articulation sémantique, *Actes des journées scientifiques 2002 « La sémantique des textes : de la théorie à la pratique »*, Reims : PUR, 2005, p. 105-129

- Pour une refondation de la didactique des langues sur la base de l'intercompréhension, *S'entendre entre langues voisines : vers l'intercompréhension*, éd. par Virginie Conti & François Grin, Genève : Georg, 2008, p.251-276

- La résolution du « problème de Saussure » et l'élaboration des connaissances par la pratique des langues, *Actes du colloque « Res per nomen »*, Reims, 2009, p. 250-263

VALLI André, Une expérience de l'enseignement de l'intercompréhension des langues romanes, *Actes du troisième colloque de l'European Center for Modern Languages « Living together in Europe in the 21st century: the challenge of plurilingual and multicultural communication and dialogue »*, Council of Europe Publishing, 2001, 16 p.

Van Dale, *Elfalig Woordenboek, Praktische woordenschat in 11 talen*, Utrecht, Antwerpen : Van Dale Lexicografie, 2007, 1135 p.

VANDERPERREN François, *Dictionnaire des / Wörterbuch der faux amis deutsch-français / français-deutsch*, 2^e éd., Bruxelles : De Boeck, Duculot, 2001, 875 p.

VANDEVYVERE Ghislain, *Grammaire pratique du néerlandais, avec exercices et corrigé*, 5^e éd., Bruxelles : De Boeck, 2005, 128 p.

VAN ROEY Jacques, GRANGER Sylviane, SWALLOW Helen, *Dictionnaire des faux amis : français-anglais, Dictionary of faux amis : English-French*, 3^e éd., Paris : Duculot, 1998, 790 p.

VENDRYES Joseph, *Le langage, introduction linguistique à l'histoire*, Paris : Albin Michel, 1979, 444 p. (L'évolution de l'humanité)

VÉRONIS Jean, LANGLAIS Philippe, Evaluation of parallel text alignment systems : The Arcade project, *Parallel Text Processing*, éd. par Jean Véronis, Kluwer Academic Publishers, 2000, p. 369-388

Disponible sur : <http://sites.univ-provence.fr/~veronis/pdf/2000-ptp-arcade.pdf> (consultée le 25/01/09)

VICTORRI Bernard, FUCHS Catherine, *La polysémie : construction dynamique du sens*, Paris : Hermès, 1996, 220 p. (Langue, raisonnement, calcul)

WALTER Henriette

- *L'aventure des langues en occident : leur origine, leur histoire, leur géographie*, Paris : Robert Laffont, 1994, 498 p.

- *Honni soit qui mal y pense : l'incroyable histoire d'amour entre le français et l'anglais*, Paris : Robert Laffont, 2001a, 364 p.

- Les « faux amis » anglais et l'autre côté du miroir, *La linguistique*, 2001b, n°37, p. 101-112

Disponible sur : http://www.cairn.info/article.php?ID_ARTICLE=LING_372_0101 (consultée le 17/10/09)

WATBLED Jean-Philippe, *La prononciation de l'anglais*, Paris : Armand Colin, 2006, 128 p. (128)

WAYNE Harbert, *The Germanic languages*, Cambridge : University Press, 2006, 522 p.

WINKLER William E., String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage, *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 1990, p. 354-359

Disponible sur : http://www.amstat.org/sections/SRMS/proceedings/papers/1990_056.pdf
(consultée le 26/05/08)

YAGUELLO Marina, *Alice au pays du langage, pour comprendre la linguistique*, Paris : Seuil, 1981, 207 p.

Annexes

Ces annexes sont constituées des tableaux présentant notre analyse du corpus, dont nous avons exposé les résultats sous forme de statistiques dans la troisième partie de notre travail. Les différents tableaux donnés ici sont les suivants :

- Tableau de l'interlexique anglais/français,
- Tableau de l'interlexique allemand/français,
- Tableau de l'interlexique néerlandais/français,

- Tableau de l'hyperlexique anglais/français,
- Tableau de l'hyperlexique allemand/français,
- Tableau de l'hyperlexique néerlandais/français,

- Tableau des mots anglais opaques,
- Tableau des mots allemands opaques,
- Tableau des mots néerlandais opaques.

Tableau de l'interlexique anglais/français

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
69206	will	V	vouloir
41842	make	V	marquer
40954	so	ADV	si
38013	know	V	connaître
37195	one	N	un
26543	people	N	peuple
24308	just	ADV	juste
19158	new	A	nouveau
18686	too	ADV	trop
16496	also	ADV	aussi
15433	great	A	grand
14307	use	V	user de

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
13078	course	N	cours
13011	no	ADV	non
12514	work	N	œuvre
11491	place	N	place
10617	really	ADV	réellement
10372	turn	V	tourner
10251	fact	N	fait
10237	part	N	partie
9368	long	A	long
9198	school ¹	N	école
9053	problem	N	problème
8969	sit	V	installer
8871	case ¹	N	cas
8871	case ²	N	caisse
8487	long	ADV	longtemps
8462	group	N	groupe
8248	face	N	face
8239	change	V	changer
8092	start	V	sursauter
8069	party	N	parti
7834	large	A	lourd
7785	point	N	point
7780	book	N	bouquin
7693	government	N	gouvernement
7653	move	V	émouvoir
7528	family	N	famille
7323	question	N	question
7272	number	N	nombre
7226	money	N	monnaie
7192	social	A	social
7170	different	A	différent
7132	idea	N	idée
6760	in	ADV	à l'intérieur
6748	hour	N	heure
6684	system	N	système
6600	important	A	important
6557	British	A	britannique
6520	body	N	body
6389	lot ²	N	lot
6268	pay	V	payer
6267	boy	N	boy

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
6066	society	N	société
6059	sort	N	sorte
6034	possible	A	possible
6012	name	N	nom
5965	moment	N	moment
5873	reason	N	raison
5794	state	N	état
5679	parent	N	père
5611	form	N	forme
5597	other	N	autre
5453	student	N	étudiant
5436	member	N	membre
5374	political	A	politique
5354	real	A	réel
5333	black	A	black
5305	follow	V	filer
5226	appear	V	apparaître
5206	age	N	âge
5182	order	N	ordre
5155	cent	N	cent
5154	line	N	ligne
5151	probably	ADV	probablement
5150	interest	N	intérêt
5113	human	A	humain
5095	expect	V	espérer
5075	union	N	union
5073	minute	N	minute
5034	office	N	office
5000	matter	N	matière
4994	example	N	exemple
4913	class	N	classe
4909	sense	N	sens
4860	service	N	service
4744	allow	V	allouer
4742	bit ²	N	bout
4725	sure	A	sûr
4705	person	N	personne
4702	continue	V	continuer
4687	pass	V	passer
4664	provide	V	pourvoir
4658	animal	N	animal

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
4652	view	N	vue
4624	general	A	général
4620	baby	N	bébé
4595	public	A	public
4502	air	N	air
4492	develop	V	développer
4487	stay ²	V	étayer
4484	true	A	droit
4483	movement	N	mouvement
4479	decide	V	décider
4477	company	N	compagnie
4450	local	A	local
4397	produce	V	produire
4358	national	A	national
4358	university	N	université
4339	wear	V	avoir
4208	table	N	table
4190	add	V	additionner
4179	education	N	éducation
4161	simply	ADV	simplement
4145	clear	A	clair
4145	action	N	action
4140	increase	V	croître
4129	American	A	américain
4108	suggest	V	suggérer
4102	break	V	briser
4090	story ¹	N	histoire
4071	suppose	V	supposer
4065	certainly	ADV	certainement
4063	policy ¹	N	politique
4063	policy ²	N	police (d'assurance)
4039	paper	N	papier
4021	law	N	loi
4003	century	N	centurie
3991	position	N	position
3973	return	V	retourner
3957	result	N	résultat
3916	experience	N	expérience
3907	drive	V	driver
3828	difficult	A	difficile
3791	rate	N	rythme

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
3788	change	N	changement
3763	effect	N	effet
3756	kill	V	couler
3730	consider	V	considérer
3726	poor	A	pauvre
3716	period	N	période
3712	energy	N	énergie
3712	accept	V	accepter
3707	development	N	développement
3700	easy	A	aisé
3700	letter	N	lettre
3694	police	N	police
3684	centre	N	centre
3672	offer	V	offrir
3628	realize	V	réaliser
3621	situation	N	situation
3606	explain	V	expliquer
3591	control	N	contrôle
3572	front	N	front
3517	economic	A	économique
3510	particular	A	particulier
3424	force	N	force
3423	particularly	ADV	particulièrement
3420	figure	N	figure
3415	industry	N	industrie
3414	history	N	histoire
3391	require	V	requérir
3387	private	A	privé
3383	nature	N	nature
3376	common	A	commun
3366	suddenly	ADV	soudain
3361	special	A	spécial
3358	united	A	uni
3304	doctor	N	docteur
3263	cover	V	couvrir
3258	subject	N	sujet
3242	cost	N	coût
3211	choose	V	choisir
3194	process	N	processus
3113	picture	N	peinture
3112	especially	ADV	spécialement

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
3109	simple	A	simple
3102	meeting	N	meeting
3102	sea	N	océan
3101	programme	N	programme
3097	value	N	valeur
3089	bank ¹	N	banc
3089	bank ²	N	banque
3089	chance	N	chance
3078	wrong	A	erroné
3078	piece	N	pièce
3075	mile	N	mile
3074	cause	V	causer
3073	arrive	V	arriver
3013	plan	V	planifier
3006	art	N	art
2992	modern	A	moderne
2968	information	N	information
2958	authority	N	autorité
2956	study	N	étude
2950	describe	V	décrire
2949	village	N	village
2934	finally	ADV	finalement
2933	stage	N	étage
2923	major	A	majeur
2899	necessary	A	nécessaire
2888	organization	N	organisation
2875	price	N	prix
2870	create	V	créer
2858	throw	V	tourner
2848	activity	N	activité
2835	receive	V	recevoir
2814	natural	A	naturel
2795	trouble	N	troubles
2787	language	N	langage
2784	decision	N	décision
2773	amount	N	montant
2767	nuclear	A	nucléaire
2761	act	N	acte
2752	doubt	N	doute
2725	sound ¹	N	son
2725	sound ²	N	sonde

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
2725	terms	N	termes
2697	difference	N	différence
2686	present	A	présent
2674	prove	V	prouver
2669	serve	V	servir
2639	court	N	cour
2629	community	N	communauté
2625	material	N	matière
2622	report	N	rapport
2613	personal	A	personnel
2596	sell	V	écouler
2595	glass	N	glace
2594	join	V	se joindre à
2593	industrial	A	industriel
2582	method	N	méthode
2571	leader	N	leader
2561	clearly	ADV	clairement
2560	hotel	N	hôtel
2549	degree	N	degré
2548	market	N	marché
2543	Soviet	A	soviétique
2535	exactly	ADV	exactement
2535	force	V	forcer
2517	garden	N	jardin
2514	officer	N	officier
2512	known	A	connu
2512	theory	N	théorie
2503	committee	N	comité
2499	discover	V	découvrir
2497	sign	N	signe
2480	future	N	futur
2469	president	N	président
2469	type	N	type
2467	recognize	V	reconnaître
2450	effort	N	effort
2446	role	N	rôle
2441	chair	N	chaire
2440	serious	A	sérieux
2438	pressure	N	pression
2438	face	V	faire face à
2437	attention	N	attention

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
2416	event	N	événement
2416	plan	N	plan
2407	space	N	espace
2397	support	N	support
2391	music	N	musique
2391	sex	N	sexe
2387	hall	N	hall
2369	quality	N	qualité
2364	account	N	compte
2364	relationship	N	relations
2361	department	N	département
2361	exist	V	exister
2360	point	V	pointer
2359	shake	V	secouer
2339	argument	N	argument
2336	tax	N	taxe
2336	tend ²	V	avoir tendance
2312	minister	N	ministre
2309	enter	V	entrer
2306	discuss	V	discuter
2303	couple	N	couple
2301	will	N	volonté
2300	hospital	N	hôpital
2299	production	N	production
2294	source	N	source
2291	fine	A	fin
2289	manage	V	être le manager de
2288	oil	N	huile
2288	river	N	rivière
2274	include	V	inclure
2256	form	V	former
2249	central	A	central
2249	chapter	N	chapitre
2248	finish	V	finir
2246	attitude	N	attitude
2245	marry	V	marier
2245	sound ¹	V	sonner
2245	sound ²	V	sonder
2244	contain	V	contenir
2235	push	V	pousser
2233	benefit	N	bienfait

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
2227	establish	V	établir
2224	rule	N	règle
2221	imagine	V	imaginer
2220	obviously	ADV	visiblement
2216	use	V	user de
2209	security	N	sécurité
2208	English	A	anglais
2202	machine	N	machine
2188	save	V	sauver
2184	club	N	club
2175	immediately	ADV	immédiatement
2169	plant	N	plante
2163	nation	N	nation
2158	cry	V	crier
2157	close	V	clure
2156	suffer	V	souffrir
2155	western	A	ouest
2151	demand	N	demande
2151	section	N	section
2150	wind ¹	N	vent
2144	council	N	conseil
2143	international	A	international
2136	operation	N	opération
2130	argue	V	argumenter
2127	army	N	armée
2127	reduce	V	réduire
2126	defence	N	défense
2120	pattern	N	patron
2114	support	V	supporter
2111	apply	V	appliquer
2103	stone	N	stèle
2103	study	V	étudier
2093	station	N	station
2085	population	N	population
2083	film	N	film
2081	rock ¹	N	rock
2081	rock ²	N	roc
2079	bottle	N	bouteille
2078	similar	A	similaire
2075	beautiful	A	beau
2073	choice	N	choix

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
2073	note	N	note
2069	right	N	droite
2053	corner	N	corner
2051	unit	N	unité
2047	completely	ADV	complètement
2043	television	N	télévision
2043	refuse	V	refuser
2039	character	N	caractère
2036	rich	A	riche
2023	mention	V	mentionner
2010	accord	V	accorder
1997	total	A	total
1997	practice	N	pratique
1996	ball ¹	N	balle
1996	ball ²	N	bal
1995	scene	N	scène
1992	colour	N	couleur
1991	attack	N	attaque
1975	physical	A	physique
1975	mass ¹	N	masse
1975	mass ²	N	messe
1975	success	N	succès
1975	visit	V	visiter
1970	attempt	N	attentat
1968	assume	V	assumer
1967	touch	V	toucher
1962	apart	ADV	à part
1954	green	A	jeune
1944	difficulty	N	difficulté
1944	science	N	science
1944	seat	N	centre
1940	direction	N	direction
1931	basic	A	basique
1926	final	A	final
1925	encourage	V	encourager
1918	detail	N	détail
1904	strange	A	étrange
1891	surface	N	surface
1882	report	V	rapporter
1881	certain	A	certain
1880	image	N	image

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
1879	structure	N	structure
1877	marriage	N	mariage
1877	occasion	N	occasion
1876	style	N	style
1876	telephone	N	téléphone
1874	express	V	exprimer
1871	past	N	passé
1871	deal	V	dealer
1866	set	N	set
1865	impossible	A	impossible
1862	turn	N	tour
1860	destroy	V	détruire
1859	object	N	objet
1852	term	N	terme
1850	usual	A	usuel
1849	entirely	ADV	entièrement
1846	standard	N	standard
1844	board	N	bord
1842	college	N	collège
1841	distance	N	distance
1841	step	N	step
1836	box ¹	N	box
1832	cause	N	cause
1832	skin	N	stencil
1830	interesting	A	intéressant
1829	represent	V	représenter
1826	blue	A	bleu
1823	interested	A	intéressé
1823	technology	N	technologie
1818	sight	N	spectacle
1814	recent	A	récent
1811	condition	N	condition
1804	star	N	star
1802	exercise	N	exercice
1802	place	V	placer
1800	press	N	presse
1792	individual	N	individu
1792	public	N	public
1792	hit	V	heurter
1775	individual	A	individuel
1773	advantage	N	avantage

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
1773	range	N	rangée
1773	present	V	présenter
1768	aware	A	averti
1764	treat	V	traiter
1763	secretary	N	secrétaire
1761	page ¹	N	page
1761	page ²	N	page
1752	dinner	N	dîner
1751	capital	N	capital
1748	training	N	entraînement
1747	admit	V	admettre
1742	military	A	militaire
1718	campaign	N	campagne
1717	ordinary	A	ordinaire
1716	second ¹	N	second
1716	second ²	N	seconde
1716	test	N	test
1715	fly	V	filer
1713	maintain	V	maintenir
1709	charge	N	charge
1701	forest	N	forêt
1698	scheme	N	combine
1695	danger	N	danger
1693	affair	N	affaire
1686	opinion	N	opinion
1685	resource	N	ressource
1683	stick	V	se coincer
1679	labour	N	labour
1678	branch	N	branche
1676	European	A	européen
1676	recently	ADV	récemment
1676	adult	N	adulte
1676	prevent	V	prévenir
1675	record	N	record
1674	flower	N	fleur
1674	pleasure	N	plaisir
1669	expression	N	expression
1664	conversation	N	conversation
1663	generally	ADV	généralement
1662	series	N	série
1658	coffee	N	café

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
1650	responsibility	N	responsabilité
1649	normal	A	normal
1649	bar	N	barre
1648	colonel	N	colonel
1644	factor	N	facteur
1635	return	N	retour
1627	tea	N	thé
1626	due	A	dû
1626	introduce	V	introduire
1622	economy	N	économie
1622	matter	V	importer
1620	election	N	élection
1620	silence	N	silence
1614	possibility	N	possibilité
1611	theatre	N	théâtre
1606	French	A	français
1603	survive	V	survivre
1597	approach	N	approche
1596	message	N	message
1596	revolution	N	révolution
1587	arrange	V	arranger
1584	memory	N	mémoire
1582	radio	N	radio
1581	duty	N	droit
1581	muscle	N	muscle
1579	farmer	N	fermier
1577	burn	V	brûler
1574	species	N	espèce
1567	gain	V	gagner
1567	press	V	presser
1564	race ²	N	race
1563	yard ¹	N	yard
1561	traditional	A	traditionnel
1561	design	V	dessiner
1559	reality	N	réalité
1553	complete	A	complet
1549	control	V	contrôler
1545	professor	N	professeur
1545	publish	V	publier
1544	key	N	clé
1544	beat	V	battre

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
1543	politics	N	politique
1541	order	V	ordonner
1535	post ¹	N	poteau
1535	post ²	N	poste
1535	post ³	N	poste
1534	protect	V	protéger
1530	farm	N	ferme
1527	aspect	N	aspect
1523	original	A	original
1522	bright	A	brillant
1517	reply	V	répliquer
1516	product	N	produit
1516	property	N	propriété
1511	compare	V	comparer
1509	conference	N	conférence
1508	extent	N	étendue
1507	cancer	N	cancer
1507	director	N	directeur
1505	pain	N	peine
1502	mountain	N	montagne
1499	basis	N	base
1498	wild	A	violent
1498	apparently	ADV	apparemment
1498	actor	N	acteur

Tableau de l'interlexique allemand/français

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
104092	sein	V	signifier
20116	so	ADV	ainsi
19041	noch	ADV	encore
12159	sagen	V	signifier
10540	groß	A	grand
10446	wollen	V	vouloir
6771	stehen	V	tenir
6267	weit	A	vaste
5577	lassen	V	laisser
5063	Land	N	land
4395	Uhr	N	heure
4362	liegen	V	être allongé
3971	halten	V	faire halte
3518	politisch	A	politique
2962	Staat	N	état
2884	fahren	V	faire
2726	amerikanisch	A	américain
2709	lang	A	long
2695	denn	ADV	donc
2686	Partei	N	parti
2579	gleich	A	égal
2341	Million	N	million
2289	natürlich	A	naturel
2269	kurz	A	court
2236	jung	A	jeune
2208	setzen	V	installer
2147	Januar	N	janvier
2065	Mai	N	mai
2062	sowjetisch	A	soviétique
2052	fest	A	fort
2038	Kampf	N	combat
2035	Politik	N	politique
1936	einfach	A	facile
1914	Problem	N	problème
1894	Juli	N	juillet
1870	international	A	international
1783	heißen ²	V	hisser
1764	nein	ADV	non

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
1723	Republik	N	république
1716	einzel	A	seul
1711	März	N	mars
1702	Sinn	N	sens
1699	französisch	A	français
1664	Präsident	N	président
1653	Name	N	nom
1643	Kreis	N	cercle
1594	Punkt	N	point
1586	Preis	N	prix
1579	Interesse	N	intérêt
1574	Juni	N	juin
1558	Februar	N	février
1554	Recht	N	droit
1550	Platz	N	place
1512	September	N	septembre
1510	klar	A	clair
1503	Werk	N	œuvre
1498	rund	A	rond
1449	kennen	V	connaître
1431	April	N	avril
1428	Minute	N	minute
1377	Kopf	N	champignon
1356	Form	N	forme
1337	modern	A	moderne
1282	Oktober	N	octobre
1257	Meter	N	mètre
1234	erkennen	V	reconnaître
1222	Plan	N	plan
1195	Gruppe	N	groupe
1189	technisch	A	technique
1187	britisch	A	britannique
1168	europäisch	A	européen
1165	Korrespondent	N	correspondant
1152	später	A	postérieur
1142	Telefon	N	téléphone
1136	Grenze	N	frontière
1113	Tendenz	N	tendance
1112	Situation	N	situation
1104	offen	A	franc
1091	Buch	N	bouquin
1068	Minister	N	ministre

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
1058	Polizei	N	police
1052	richten	V	braquer
1034	Einheit	N	unité
1030	recht	A	droit
1022	Mark ²	N	marche
1022	Mark ³	N	mark
1012	Diskussion	N	discussion
1010	Rolle	N	rôle
983	Professor	N	professeur
979	Satz	N	set
977	Industrie	N	industrie
971	demokratisch	A	démocrate
968	November	N	novembre
956	bald	ADV	bientôt
956	Nacht	N	nuit
943	Kollege	N	collègue
927	Foto	N	photo
916	Reihe	N	rang
904	Programm	N	programme
898	Person	N	personne
892	persönlich	A	personnel
890	Konferenz	N	conférence
883	Sicherheit	N	sécurité
876	Sprache	N	parole
866	praktisch	A	pratique
866	Produktion	N	production
862	Maschine	N	machine
838	Ausdruck	N	expression
831	wahr	A	vrai
829	Bank ¹	N	banc
829	Bank ²	N	banque
824	Universität	N	université
813	Dame	N	dame
811	Technik	N	technique
806	Film	N	film
806	Thema	N	thème
801	Kilometer	N	kilomètre
770	Pol	N	pôle
766	Grad	N	grade
765	System	N	système
753	Firma	N	firme
753	Wahrheit	N	vérité

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
750	General	N	général
742	kommunistisch	A	communiste
742	Ordnung	N	ordre
725	Dollar	N	dollar
725	erzielen	V	réaliser
725	militärisch	A	militaire
721	Student	N	étudiant
718	Milliarde	N	milliard
717	Musik	N	musique
717	Zug	N	section
714	Familie	N	famille
710	langsam	A	long
710	Natur	N	nature
706	direkt	A	direct
705	Chance	N	chance
705	national	A	national
703	Klasse	N	classe
700	Hotel	N	hôtel
700	Rahmen	N	ramette
699	Bürger	N	bourgeois
699	Soldat	N	soldat
691	Organisation	N	organisation
688	Auto	N	auto
688	einsetzen	V	instituer
684	Einfluß	N	influence
682	Wochenende	N	week-end
677	Element	N	élément
677	polnisch	A	polonais
670	englisch	A	anglais
669	Amerikaner	N	Américain
669	Delegation	N	délégation
664	Opfer	N	offrande
662	Nation	N	nation
659	Linie	N	ligne
653	Idee	N	idée
636	ökonomisch	A	économique
636	Theater	N	théâtre
635	erneuern	V	rénover
628	inner	A	intérieur
623	historisch	A	historique
621	Licht	N	clarté
619	kämpfen	V	combattre

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
619	Markt	N	marché
616	interessant	A	intéressant
608	Angst	N	angoisse
597	privat	A	privé
595	Institut	N	institut
590	Kosten ²	N	coût
585	Prinzip	N	principe
583	Demokratie	N	démocratie
578	Urteil	N	arrêt
575	Position	N	position
562	Motor	N	moteur
553	Presse	N	presse
547	Meister	N	maître
541	Kommission	N	commission
541	lang	ADV	longuement
534	Fenster	N	fenêtre
531	lange	ADV	longuement
531	wachsen ¹	V	avancer
530	Artikel	N	article
529	Osten	N	est
528	treiben	V	rabattre
526	Kurs	N	cours
520	Methode	N	méthode
520	Qualität	N	qualité
520	West	N	ouest
515	Doktor	N	docteur
515	italienisch	A	italien
515	Kommunist	N	communiste
513	konkret	A	concret
513	stetig	A	continu
507	Kritik	N	critique
506	Front	N	front
506	Titel	N	titre
504	Regel	N	règle
502	Aktie	N	action
497	Aktion	N	action
495	Parlament	N	parlement
494	aktiv	A	actif
494	Berliner	N	Berlinois
491	Politiker	N	politique
490	Sekretär	N	secrétaire
488	Mama	N	maman

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
487	Verband	N	formation
482	Zone	N	zone
479	Sprecher	N	speaker
474	Bischof	N	bischof
468	Stand	N	stand
467	Sport	N	sport
467	Wille	N	volonté
466	dauern ¹	V	durer
460	Kultur	N	culture
459	Post	N	poste
457	spüren	V	éprouver
452	Vorgang	N	affaire
450	Persönlichkeit	N	personnalité
448	Papier	N	papier
445	Größe	N	grosueur
445	Tonne	N	tonne

Tableau de l'interlexique néerlandais/français

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
123613	zo	ADV	ainsi
76697	willen	V	vouloir
58078	mens ¹	N	monde
51661	groot	A	grand
42525	lang	A	long
40957	liggen	V	être allongé
33096	altijd	ADV	toujours
32729	nieuw	A	nouveau
27996	plaats	N	place
27468	zijn	N	le sien
23062	hoofd	N	chef
22157	zeker	A	en sécurité
21514	kennen	V	connaître
21418	natuurlijk	A	naturel
20811	paar	N	paire
18351	noemen	V	nommer
18025	uur	N	heure
17857	net	ADV	net
17790	naam	N	nom
17652	zetten	V	installer
16928	juist	A	juste
15877	manier	N	manière
15474	kamer	N	chambre
15011	feit	N	fait
14811	zin	N	sens
14756	kort	A	court
14392	probleem	N	problème
14050	soort	N	sorte
13752	slaan ¹	V	cogner sur
13752	slaan ²	V	sonner
13676	groep	N	groupe
13261	jong	A	jeune
12617	moment	N	moment
12353	kant	N	côté
12305	staat	N	état
11257	nacht	N	nuit
10772	steken ²	V	être coincé
10356	politiek	A	politique

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
10292	school	N	école
9837	sociaal	A	social
9829	recht	N	droit
9812	zwaar	A	sévère
9735	wijn	N	vin
9705	precies	A	précis
9677	situatie	N	situation
9649	opnieuw	ADV	de nouveau
9600	stuk	N	article
9381	merken	V	marquer
8926	idee	N	idée
8892	rol ¹	N	rôle
8798	auto	N	auto
8688	meneer	N	monsieur
8562	kans	N	chance
8557	waarschijnlijk	A	vraisemblable
8517	plan	N	plan
8422	richting	N	orientation
8351	sluiten ¹	V	clôturer
8289	richten ¹	V	orienter
8243	persoon	N	personne
8091	klinken ²	V	enchaîner
8070	invloed	N	influence
7920	gang ²	N	gang
7609	waar	A	vrai
7555	kiezen ¹	V	choisir
7555	kiezen ²	V	choisir
7490	angst	N	angoisse
7359	minuut	N	minute
7291	punt ¹	N	point
7291	punt ³	N	pointe
7224	partij	N	partie
7155	direct	A	direct
7105	hond	N	chien
7076	relatie	N	relation
7011	systeem	N	système
6995	stil	A	silencieux
6929	proces	N	procès
6760	contact	N	contact
6753	orde	N	ordre
6715	functie	N	fonction
6689	ongeveer	ADV	environ

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
6642	schieten ³	V	shooter
6570	kleur	N	couleur
6505	pijn	N	peine
6459	artikel	N	article
6389	persoonlijk	A	personnel
6327	taak	N	tâche
6229	muur	N	mur
6194	passen ¹	V	passer
6147	duren	V	durer
6140	dokter	N	docteur
5978	kennis ¹	N	connaissance
5978	kennis ²	N	connaissance
5880	regel	N	règle
5832	min	A	moins
5809	gelijk	A	égal
5792	economisch	A	économique
5721	stand ¹	N	attitude
5721	stand ²	N	stand
5695	familie	N	famille
5674	periode	N	période
5593	resultaat	N	résultat
5578	normaal	A	normal
5554	rond	A	rond
5483	organisatie	N	organisation
5456	speciaal	A	spécial
5439	patiënt	N	patient
5425	broer	N	frère
5412	leveren	V	livrer
5408	daarvoor	ADV	devant
5398	totaal	A	total
5281	blauw	A	bleu
5270	klaar	A	clair
5215	karakter	N	caractère
5191	Amerikaans	A	américain
5164	hoogte	N	hauteur
5159	theorie	N	théorie
5139	Frans	A	français
5058	schrijver	N	écrivain
5033	meter ¹	N	mètre
4952	krant	N	quotidien
4915	centraal	A	central
4889	rest	N	reste

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
4886	muziek	N	musique
4875	scherp	A	coupant
4834	bank	N	banc
4830	blad	N	plateau
4802	zak	N	sac
4789	soldaat	N	soldat
4786	papier	N	papier
4759	fles	N	flacon
4752	waarheid	N	vérité
4723	grens	N	frontière
4706	minister	N	ministre
4702	wind	N	vent
4699	koffie	N	café
4691	stoppen ¹	V	stopper
4668	oorzaak	N	raison
4618	informatie	N	information
4598	methode	N	méthode
4594	positie	N	position
4580	huilen	V	hurler
4557	enorm	A	énorme
4522	foto	N	photo
4507	film	N	film
4500	activiteit	N	activité
4494	modern	A	moderne
4477	trachten	V	tâcher
4442	basis	N	base
4438	niveau	N	niveau
4416	lijn	N	ligne
4406	tante	N	tante
4386	prijs	N	prix
4344	reactie	N	réaction
4340	telkens	ADV	à tous les coups
4318	praktijk	N	pratique
4297	neus	N	nez
4283	stilte	N	silence
4261	inzicht	N	intention
4259	terrein	N	terrain
4247	model	N	modèle
4200	cultuur	N	culture
4197	reageren	V	réagir
4188	kapitein	N	capitaine
4155	grijpen	V	attraper

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
4088	element	N	élément
4045	stof ²	N	étoffe
4038	politie ¹	N	police
4037	Parijs	A	parisien
3987	voort	ADV	en avant
3951	factor	N	facteur
3951	succes	N	succès
3941	kost	N	coût
3940	natuur	N	nature
3913	hotel	N	hôtel
3890	fijn	A	fin
3883	studie	N	étude
3869	aspect	N	aspect
3851	figuur	N	figure
3841	nationaal	A	national
3835	toon	N	ton
3833	plek	N	place
3779	dame	N	dame
3774	stukje	N	article
3745	schenken	V	concéder
3742	kaart	N	carte
3740	term	N	terme
3738	bureau	N	bureau
3722	opdracht	N	ordre
3690	Engels	A	anglais
3685	effect	N	effet
3685	rand	N	rand
3660	althans	ADV	en tout cas
3652	plant	N	plante
3633	ervaren	V	éprouver
3587	tekst	N	texte
3567	telefoon	N	téléphone
3566	kosten	V	coûter
3558	bos ²	N	bois
3534	individueel	A	individuel
3532	structuur	N	structure
3529	roman	N	roman
3515	burger	N	bourgeois
3510	kring	N	cerne
3508	grijs	A	gris
3482	kwestie	N	question
3478	eventueel	A	éventuel

Fréquence	Lemme	Catégorie grammaticale	Équivalent français
3435	trein	N	train
3432	praktisch	A	pratique

Tableau de l'hyperlexique anglais/français

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
70930	can	V	pouvoir		m.r. : chance s.t. : possibilité
69206	will	V	vouloir		
60143	there	ADV	y		m.r. : terre s.t. : lieu
41842	make	V	marquer		
40954	so	ADV	si		
38013	know	V	connaître		
37195	one	N	un		
35351	time	N	temps		m.r. : timing s. : temps
35152	come	V	venir		m.r. : chemin s.t. : se déplacer
34323	take	V	prendre		r. : toucher
33749	then	ADV	donc	/ð/-/d/	
29231	man	N	homme	/n/-/l/	s. : mâle
26543	people	N	peuple		
26038	good	A	gentil	/d/-/t/	
25476	year	N	an	/j/-/ʒ/	m.r. : jour s.t. : durée
24308	just	ADV	juste		
24165	look	V	regarder		r. : loucher
23651	way	N	chemin		s. : voie
22921	give	V	donner	/g/-/k/, /v/- /f/	s. : confier
19857	as	ADV	comme		r. : ainsi
19525	find	V	trouver	/f/-/v/, /d/-/t/	r. : inventer
19389	child	N	enfant		r. : chiard
19158	new	A	nouveau		
19040	tell	V	dire	/t/-/d/, /l/-/n/	r. : dénoter
18686	too	ADV	trop		
16520	must	V	devoir		m.r. : impôt s.t. : obligation
16496	also	ADV	aussi		
15911	seem	V	sembler	/s/-/ʃ/	m.r. : chimère r. : illusion
15489	feel	V	palper	/f/-/p/	
15433	great	A	grand		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
15279	old	A	vieux	/l/-/n/	r. : aîné
15241	woman	N	femme		m.r. : vamp s. : femme
14426	put	V	mettre	/t/-/s/	s. : passer
14307	use	V	user de		
14042	leave	V	quitter		s. : enlever
13237	try	V	essayer	/t/-/s/	
13078	course	N	cours		
13011	no	ADV	non		
12983	hand	N	main		m.r. : handball s. : main
12514	work	N	œuvre		
12429	call	V	appeler	/l/-/n/	s. : sonner
12254	begin	V	commencer		m.r. : bégaïement r. : commencement
12155	away	ADV	loin		r. : vieux
11850	always	ADV	toujours	/l/-/n/	m.r. : univers s.t. : totalité
11794	keep	V	garder		m.r. : accaparer s.t. : retenir
11491	place	N	place		
11396	far	ADV	loin		r. : infra
11284	late	A	tardif		s. : lentement
10864	house	N	maison	/s/-/t/	r. : hutte
10766	small	A	petit	/l/-/n/	r. : menu
10617	really	ADV	réellement		
10372	turn	V	tourner		
10251	fact	N	fait		
10237	part	N	partie		
10015	country	N	pays		r. : contrée
9925	work	V	travailler		r. : œuvrer
9371	live	V	vivre	/l/-/n/, /v/-/b/	m.r. : anabiose s.t. : vie
9368	long	A	long		
9198	school ¹	N	école		
9198	school ²	N	banc		m.r. : écaille s.t. : poissons
9187	head	N	tête	/d/-/t/	s. : hauteur
9176	bring	V	apporter	/b/-/p/	
9141	talk	V	parler		m.r. : talk-show s.t. :

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
					conversation
9053	problem	N	problème		
9053	hear	V	entendre		m.r. : harmonie s.t. : oreille
8969	sit	V	installer		
8882	kind	N	genre	/k/-/g/	
8871	case ¹	N	cas		
8871	case ²	N	caisse		
8828	word	N	mot		r. : verbe
8751	young	A	jeune	/j/-/ʒ/	
8506	mother	N	mère		r. : mater
8487	long	ADV	longtemps		
8462	group	N	groupe		
8376	stand	V	mettre		r. : étendre
8317	write	V	rédiger	/t/-/d/	
8308	night	N	nuit		r. : nocturne
8260	water	N	eau		r. : aquatique
8248	face	N	face		
8239	change	V	changer		
8143	side	N	côté	/d/-/t/	
8092	start	V	sursauter		
8069	party	N	parti		
7903	run	V	courir	/r/-/l/	s. : élans
7856	girl	N	fille		r. : greluche
7834	large	A	lourd		
7785	point	N	point		
7782	end	N	bout		m.r. : nouer s. : bout
7780	book	N	bouquin		
7728	let	V	laisser	/t/-/s/	
7712	week	N	semaine		r. : week-end
7693	government	N	gouvernement		
7653	move	V	émouvoir		
7649	believe	V	croire	/b/-/v/	r. : avaler
7528	family	N	famille		
7323	question	N	question		
7272	number	N	nombre		
7245	play	V	jouer	/p/-/f/	r. : flamber
7226	money	N	monnaie		
7192	social	A	social		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
7170	different	A	différent		
7132	idea	N	idée		
7094	right	A	bien	/t/-/s/	m.r. : réussi s. : bien
7059	big	A	grand		r. : bringue
6957	first	ADV	plutôt	/f/-/p/	
6924	door	N	porte	/d/-/t/	m.r. : trou s.t. : ouverture
6760	in	ADV	à l'intérieur		
6748	hour	N	heure		
6684	system	N	système		
6655	speak	V	parler	/k/-/s/	m.r. : expression s.t. : s'exprimer
6600	important	A	important		
6557	British	A	britannique		
6538	remember	V	se souvenir de		r. : se remémorer
6520	body	N	body		
6484	war	N	guerre	/w/-/g/	
6389	lot ²	N	lot		
6386	friend	N	ami	/f/-/p/, /d/-/t/	r. : pote
6342	car	N	voiture		r. : char
6268	pay	V	payer		
6267	boy	N	boy		
6199	white	A	blanc	/t/-/d/	m.r. : vide r. : blanc
6159	power	N	force	/p/-/f/	
6152	area	N	superficie		r. : aire
6112	grow	V	s'accroître	/g/-/k/	
6096	home	N	maison		r. : home
6086	lose	V	perdre		m.r. : loser r. : perdant
6066	society	N	société		
6059	sort	N	sorte		
6034	possible	A	possible		
6012	name	N	nom		
6008	mind	N	esprit	/d/-/t/	r. : mental
5966	job	N	travail		r. : job
5965	moment	N	moment		
5955	stop	V	cesser	/t/-/s/	
5887	able	A	capable		r. : habile

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
5873	reason	N	raison		
5861	over	ADV	dessus	/r/-/l/	m.r. : voler s. : dessus
5857	foot	N	patte	/f/-/p/	
5855	like	V	aimer	/k/-/g/	m.r. : languir s.t. : désirer
5794	state	N	état		
5762	street	N	rue		m.r. : trottoir s. : rue
5735	whole	A	tout		m.r. : holo- s.t. : entier
5720	wait	V	attendre	/w/-/g/	r. : guetter
5679	parent	N	père		
5663	month	N	mois		r. : mensuel
5630	set	V	mettre	/s/-/t/, /t/-/d/	r. : étendre
5616	right	ADV	droit		s. : droite
5611	form	N	forme		
5598	morning	N	matin		s. : minuit
5597	other	N	autre		
5582	carry	V	porter		m.r. : charrier s.t. : transporter
5552	walk	V	marcher		s. : avancer
5524	learn	V	apprendre		m.r. : lire s.t. : s'instruire
5453	student	N	étudiant		
5436	member	N	membre		
5386	father	N	père	/f/-/p/	r. : pater
5374	political	A	politique		
5362	lie ¹	V	s'allonger		s. : long
5362	lie ²	V	mentir	/l/-/n/	s. : nier
5354	real	A	réel		
5333	black	A	black		
5317	food	N	pâtée	/f/-/p/, /d/-/t/	
5305	follow	V	filer		
5292	lead	V	conduire		m.r. : leader s.t. : tête
5276	fall	V	tomber		m.r. : s'affaler s. : tomber
5226	appear	V	apparaître		
5219	light	N	lumière		r. : clarté
5206	age	N	âge		
5182	order	N	ordre		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
5155	cent	N	cent		
5154	line	N	ligne		
5151	probably	ADV	probablement		
5150	interest	N	intérêt		
5113	human	A	humain		
5095	expect	V	espérer		
5075	union	N	union		
5073	minute	N	minute		
5034	office	N	office		
5000	matter	N	matière		
4994	example	N	exemple		
4913	class	N	class		
4909	sense	N	sens		
4860	service	N	service		
4831	bed	N	bâti	/d/-/t/	
4823	voice	N	voix		r. : vocal
4822	send	V	envoyer		m.r. : conduire s.t. : faire aller
4754	full	A	plein	/f/-/p/	
4744	allow	V	allouer		
4742	bit ¹	N	mors	/t/-/d/	m.r. : bride s. : mors
4742	bit ²	N	bout		
4725	sure	A	sûr		
4714	day	N	temps	/d/-/t/	
4705	person	N	personne		
4702	continue	V	continuer		
4696	remain	V	rester		m.r. : rémanent s.t. : subsister
4691	pound ¹	N	livre	/p/-/f/	m.r. : franc s.t. : unité monétaire
4691	pound ²	N	fourrière		m.r. : ponton s.t. : dépôt
4687	pass	V	passer		
4664	provide	V	pourvoir		
4658	animal	N	animal		
4652	view	N	vue		
4628	land	N	terre		m.r. : lande s. : terre
4624	general	A	général		
4620	baby	N	bébé		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
4619	city	N	ville		r. : cité
4595	public	A	public		
4594	buy	V	acheter	/b/-/p/	r. : payer
4577	reach	V	atteindre	/tʃ/-/s/	m.r. : réussir s. : atteindre
4502	air	N	air		
4492	develop	V	développer		
4487	stay ¹	V	rester		m.r. : stagnant s. : rester
4487	stay ²	V	étayer		
4484	true	A	droit		
4483	movement	N	mouvement		
4479	decide	V	décider		
4477	company	N	compagnie		
4458	road	N	route	/d/-/t/	
4450	local	A	local		
4442	wife	N	femme	/f/-/p/	r. : vamp
4409	business	N	affaires		r. : business
4397	produce	V	produire		
4358	national	A	national		
4358	university	N	université		
4339	wear	V	avoir		
4336	build	V	bâtir	/d/-/t/	
4279	die	V	mourir	/d/-/t/	m.r. : tuer s. : mourir
4277	love	V	aimer		m.r. : alcôve s.t. : amoureux
4236	level	N	niveau	/l/-/n/	
4208	table	N	table		
4204	actually	ADV	en fait		m.r. : actuellement s.t. : effectivement
4190	add	V	additionner		
4179	education	N	éducation		
4161	simply	ADV	simplement		
4151	use	N	emploi	/s/-/z/	m.r. : user s.t. : employer
4145	clear	A	clair		
4145	action	N	action		
4140	increase	V	croître		
4129	American	A	américain		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
4121	open	V	ouvrir		m.r. : épanoui s.t. : ouvert
4108	suggest	V	suggérer		
4102	break	V	briser		
4090	story ¹	N	histoire		
4090	story ²	N	étage	/r/-/l/	r. : escalier
4071	suppose	V	supposer		
4065	certainly	ADV	certainement		
4063	policy ¹	N	politique		
4063	policy ²	N	police (d'assurance)		
4052	love	N	amour	/v/-/b/	m.r. : libido s.t. : sexuel
4039	paper	N	papier		
4021	law	N	loi		
4005	town	N	ville		m.r. : township s. : ville
4003	century	N	centurie		
3991	position	N	position		
3973	return	V	retourner		
3957	result	N	résultat		
3916	experience	N	expérience		
3907	drive	V	driver		
3853	draw	V	tirer	/d/-/t/	
3828	difficult	A	difficile		
3808	open	A	ouvert		m.r. : épanoui s. : ouvert
3791	rate	N	rythme		
3788	change	N	changement		
3779	involve	V	impliquer	/v/-/b/	m.r. : embourber s.t. : engager
3776	wall	N	mur	/w/-/g/, /l/- /n/	m.r. : gaine s.t. : paroi
3766	arm	N	bras		m.r. : armé s. : bras
3763	effect	N	effet		
3756	kill	V	couler		
3730	consider	V	considérer		
3726	poor	A	pauvre		
3716	period	N	période		
3712	energy	N	énergie		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
3712	accept	V	accepter		
3707	development	N	développement		
3706	hard	A	dur		r. : acharné
3700	easy	A	aisé		
3700	letter	N	lettre		
3694	police	N	police		
3684	centre	N	centre		
3680	rest	N	repos		r. : arrêt
3672	offer	V	offrir		
3668	worker	N	travailleur		r. : ouvrier
3628	realize	V	réaliser		
3621	situation	N	situation		
3619	evening	N	soir	/v/-/f/	s. : fin
3613	short	A	court	/ʃ/-/k/	
3606	explain	V	expliquer		
3605	around	ADV	autour	/r/-/l/, /d/-/t/	r. : alentour
3605	game	N	jeu		m.r. : gambit s. : jeu
3591	control	N	contrôle		
3584	window	N	fenêtre	/d/-/t/	s. : ouverture
3576	hair	N	cheveux		m.r. : haire s.t. : poil
3572	front	N	front		
3517	economic	A	économique		
3514	back	N	dos	/k/-/s/	r. : bosse
3510	particular	A	particulier		
3486	rise	V	se dresser		
3459	catch	V	attraper		m.r. : accrocher s. : attraper
3434	thought	N	intention	/θ/-/t/, /t/-/s/	
3432	tree	N	arbre		s. : tronc
3424	force	N	force		
3424	agree	V	consentir	/g/-/k/	m.r. : accord s. : consentir
3423	particularly	ADV	particulièrement		
3421	field	N	champ		m.r. : fond r. : champ
3420	figure	N	figure		
3415	industry	N	industrie		
3415	cut	V	couper		m.r. : cutter s. : couper

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
3414	history	N	histoire		
3391	require	V	requérir		
3387	private	A	privé		
3386	pick	V	choisir	/p/-/b/, /k/- /s/	r. : embrasser
3383	nature	N	nature		
3376	common	A	commun		
3366	suddenly	ADV	soudain		
3361	special	A	spécial		
3358	united	A	uni		
3330	pull	V	tirer	/p/-/f/	r. : flinguer
3327	free	A	libre		r. : franc
3304	doctor	N	docteur		
3287	church	N	église		m.r. : clocher s. : église
3284	likely	A	probable	/k/-/ʒ/	m.r. : logique r. : vrai
3263	cover	V	couvrir		
3258	subject	N	sujet		
3242	cost	N	coût		
3219	thus	ADV	par conséquent	/ð/-/d/	r. : donc
3211	choose	V	choisir		
3194	process	N	processus		
3181	ground	N	terre	/g/-/k/, /d/-/t/	m.r. : croûte s. : terre
3167	building	N	bâtiment	/b/-/p/, /d/-/t/	m.r. : plâtrerie r. : bâtiment
3152	trade	N	commerce	/d/-/t/	m.r. : traite s. : commerce
3151	floor	N	plancher	/f/-/p/	
3147	raise	V	lever		r. : dresser
3140	leg	N	jambe		m.r. : ligoter s. : jambe
3113	picture	N	peinture		
3112	especially	ADV	spécialement		
3109	simple	A	simple		
3102	meeting	N	meeting		
3102	sea	N	océan		
3101	programme	N	programme		
3097	value	N	valeur		
3089	bank ¹	N	banc		
3089	bank ²	N	banque		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
3089	chance	N	chance		
3089	top	N	sommet		m.r. : top niveau r. : sommet
3078	wrong	A	erroné		
3078	piece	N	pièce		
3075	mile	N	mile		
3074	cause	V	causer		
3073	arrive	V	arriver		
3058	laugh	V	rire	/l/-/n/	m.r. : ineffable s. : rire
3056	forget	V	oublier	/f/-/p/, /t/-/d/	r. : perdre
3048	answer	V	répondre	/s/-/t/	m.r. : interroger s.t. : réponse
3013	plan	V	planifier		
3006	art	N	art		
2992	modern	A	moderne		
2968	information	N	information		
2966	cold	A	froid	/k/-/ʒ/	r. : gelé
2958	authority	N	autorité		
2956	study	N	étude		
2950	describe	V	décrire		
2949	village	N	village		
2937	heart	N	cœur		r. : charité
2934	finally	ADV	finalement		
2933	stage	N	étage		
2927	fish	N	poisson	/f/-/p/	
2923	major	A	majeur		
2910	teacher	N	professeur	/r/-/n/	s. : technique
2905	fire	N	feu	/f/-/p/	r. : pyro-
2899	necessary	A	nécessaire		
2892	smile	V	sourire		m.r. : aimable s. : sourire
2888	organization	N	organisation		
2878	dark	A	sombre	d/-/t/, /k/-/s/	r. : triste
2875	price	N	prix		
2870	create	V	créer		
2858	throw	V	tourner		
2848	activity	N	activité		
2835	receive	V	recevoir		
2826	home	ADV	chez soi		r. : home

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
2814	natural	A	naturel		
2795	trouble	N	troubles		
2787	language	N	langage		
2784	decision	N	décision		
2773	amount	N	montant		
2767	nuclear	A	nucléaire		
2762	husband	N	mari	/b/-/p/	m.r. : harpagon s.t. : homme
2762	win	V	gagner	/w/-/g/	
2761	act	N	acte		
2752	doubt	N	doute		
2728	sun	N	soleil	/n/-/l/	
2725	sound ¹	N	son		
2725	sound ²	N	sonde		
2725	sound ³	N	détroit	/s/-/t/	s. : étendue
2725	terms	N	termes		
2717	happy	A	heureux		m.r. : happy end s. : heureux
2704	evidence	N	témoignage		m.r. : évidence s.t. : vérité
2698	nice	A	sympathique	/n/-/l/	m.r. : plaisant s.t. : agréable
2697	difference	N	différence		
2686	present	A	présent		
2677	quickly	ADV	vite	/k/-/g/	m.r. : galop s. : vite
2674	prove	V	prouver		
2669	serve	V	servir		
2656	enjoy	V	aimer		m.r. : joie s.t. : agréable
2639	court	N	cour		
2629	community	N	communauté		
2628	all	ADV	tout		m.r. : plein s.t. : entièrement
2625	material	N	matière		
2622	red	A	rouge	/d/-/t/	m.r. : crête s. : rouge
2622	fear	N	peur	/f/-/p/	
2622	report	N	rapport		
2613	personal	A	personnel		
2598	hot	A	chaud		m.r. : hot-dog

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
					s. : chaud
2596	sell	V	écouler		
2595	glass	N	glace		
2594	join	V	se joindre à		
2593	industrial	A	industriel		
2583	issue	N	question		m.r. : essai s.t. : sujet
2582	method	N	méthode		
2571	leader	N	leader		
2561	clearly	ADV	clairement		
2560	hotel	N	hôtel		
2558	teach	V	apprendre		m.r. : technicien s.t. : connaître
2551	drop	V	tomber	/d/-/t/, /p/-/b/	
2549	degree	N	degré		
2549	fight	V	se battre		m.r. : s'affronter s.t. : combattre
2548	market	N	marché		
2543	Soviet	A	soviétique		
2538	blood	N	sang		m.r. : boudin s. : sang
2535	exactly	ADV	exactement		
2535	force	V	forcer		
2524	feeling	N	sensation		m.r. : feeling s.t. : sentiment
2518	only	A	seul		r. : un
2518	single	A	seul		s. : singulier
2517	garden	N	jardin		
2514	officer	N	officier		
2512	known	A	connu		
2512	theory	N	théorie		
2503	committee	N	comité		
2499	discover	V	découvrir		
2497	sign	N	signe		
2482	deal ¹	N	donne	/l/-/n/	
2480	future	N	futur		
2478	brother	N	frère	/b/-/p/	r. : pote
2469	president	N	président		
2469	type	N	type		
2467	recognize	V	reconnaître		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
2463	fill	V	plomber	/f/-/p/	
2461	play	N	jeu	/p/-/b/	r. : boules
2450	effort	N	effort		
2446	role	N	rôle		
2442	concerned	A	inquiet		r. : soucieux
2441	chair	N	chaire		
2440	serious	A	sérieux		
2438	pressure	N	pression		
2438	face	V	faire face à		
2437	attention	N	attention		
2421	knowledge	N	connaissance		m.r. : nullité s. : connaissance
2416	event	N	événement		
2416	plan	N	plan		
2407	space	N	espace		
2405	truth	N	vérité		r. : truisme
2397	support	N	support		
2391	music	N	musique		
2391	sex	N	sexe		
2387	answer	N	réponse	/s/-/t/	m.r. : interroger s. : réponse
2387	hall	N	hall		
2379	feed	V	paître	/f/-/p/, /d/-/t/	
2372	horse	N	cheval		r. : haras
2371	look	N	regard		r. : loucher
2369	quality	N	qualité		
2364	account	N	compte		
2364	relationship	N	relations		
2363	slowly	ADV	lentement		m.r. : slow s.t. : lent
2361	department	N	département		
2361	health	N	santé		m.r. : chaleur s.t. : organisme
2361	exist	V	exister		
2360	point	V	pointer		
2359	shake	V	secouer		
2355	purpose	N	but		r. : propos
2339	argument	N	argument		
2336	tax	N	taxe		
2336	tend ²	V	avoir tendance		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
2323	notice	V	s'apercevoir de		r. : noter
2321	drink	V	boire		m.r. : trinquer s. : boire
2312	minister	N	ministre		
2309	enter	V	entrer		
2306	discuss	V	discuter		
2306	sleep	V	dormir		r. : s'assoupir
2305	daughter	N	fille		m.r. : adopter s. : fille
2303	couple	N	couple		
2301	will	N	volonté		
2300	hospital	N	hôpital		
2299	production	N	production		
2294	source	N	source		
2291	fine	A	fin		
2291	various	A	divers		r. : varié
2289	shoulder	N	épaule		r. : aisselle
2289	manage	V	être le manager de		
2288	oil	N	huile		
2288	river	N	rivière		
2281	shop	N	magasin		r. : échoppe
2274	include	V	inclure		
2267	deep	A	profond	/d/-/t/, /p/-/b/	m.r. : tomber s. : profond
2256	form	V	former		
2249	central	A	central		
2249	chapter	N	chapitre		
2248	finish	V	finir		
2246	attitude	N	attitude		
2245	marry	V	marier		
2245	sound ¹	V	sonner		
2245	sound ²	V	sonder		
2244	contain	V	contenir		
2235	push	V	pousser		
2233	benefit	N	bienfait		
2232	occur	V	se produire		m.r. : échoir r. : advenir
2227	establish	V	établir		
2225	clothes	N	vêtements		m.r. : collant s.t. : vêtement
2224	rule	N	règle		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
2221	South	N	sud	/θ/-/d/	
2221	imagine	V	imaginer		
2220	obviously	ADV	visiblement		
2216	use	V	user de		
2212	finger	N	doigt	/f/-/p/	m.r. : panaris s. : doigt
2212	act	V	agir		s. : actes
2209	security	N	sécurité		
2208	English	A	anglais		
2205	research	N	recherche	/s/-/ʃ/	
2202	machine	N	machine		
2198	size ¹	N	colle	/s/-/ʃ/, /z/-/s/	m.r. : chassie s.t. : gluant
2198	size ²	N	taille	/s/-/k/	m.r. : quinze s. : taille
2188	save	V	sauver		
2184	club	N	club		
2175	immediately	ADV	immédiatement		
2175	care	V	aimer		r. : chérir
2174	forward	ADV	en avant	/f/-/v/, /w/- /g/	m.r. : avant- garde s. : en avant
2169	plant	N	plante		
2167	lay	V	mettre		s. : lieu
2163	nation	N	nation		
2158	cry	V	crier		
2157	close	V	clore		
2156	suffer	V	souffrir		
2155	western	A	ouest		
2154	staff ¹	N	personnel		m.r. : staff s.t. : entreprise
2154	staff ²	N	bâton		r. : stick
2154	staff ³	N	portée		m.r. : octave s.t. : musique
2151	demand	N	demande		
2151	section	N	section		
2150	wind ¹	N	vent		
2150	wind ²	N	tournant	/n/-/l/	m.r. : ovale s.t. : courbe
2147	seek	V	chercher	/s/-/ʃ/	
2145	growth	N	croissance	/g/-/k/	s. : croître
2144	council	N	conseil		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
2143	international	A	international		
2137	share	V	partager		r. : écarteler
2136	operation	N	opération		
2132	instance	N	exemple		m.r. : instar s. : exemple
2130	argue	V	argumenter		
2128	wide	A	vaste	/d/-/t/	
2127	army	N	armée		
2127	hill	N	colline	/l/-/n/	m.r. : hune s.t. : arrondie
2127	reduce	V	réduire		
2126	defence	N	défense		
2121	achieve	V	accomplir		r. : achever
2120	pattern	N	patron		
2114	support	V	supporter		
2111	apply	V	appliquer		
2108	reading	N	lecture	/r/-/l/, /d/-/t/	
2103	stone	N	stèle		
2103	study	V	étudier		
2093	station	N	station		
2085	population	N	population		
2083	film	N	film		
2082	news	N	nouvelle		m.r. : annonce r. : nouvelle
2081	rock ¹	N	rock		
2081	rock ²	N	roc		
2079	bottle	N	bouteille		
2078	similar	A	similaire		
2075	beautiful	A	beau		
2073	choice	N	choix		
2073	note	N	note		
2069	right	N	droite		
2061	dog	N	chien		m.r. : dogue s. : chien
2057	avoid	V	éviter	/d/-/t/	
2053	corner	N	corner		
2051	unit	N	unité		
2047	completely	ADV	complètement		
2043	television	N	télévision		
2043	refuse	V	refuser		
2039	character	N	caractère		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
2039	sister	N	sœur		m.r. : cousin s. : sœur
2036	rich	A	riche		
2023	mention	V	mentionner		
2015	lady	N	dame		m.r. : lady s. : dame
2010	accord	V	accorder		
1999	strike	V	frapper		m.r. : trique s. : frapper
1997	total	A	total		
1997	practice	N	pratique		
1996	ball ¹	N	balle		
1996	ball ²	N	bal		
1995	scene	N	scène		
1994	kitchen	N	cuisine	/ʃ/-/z/	
1992	colour	N	couleur		
1991	attack	N	attaque		
1975	physical	A	physique		
1975	mass ¹	N	masse		
1975	mass ²	N	messe		
1975	success	N	succès		
1975	visit	V	visiter		
1970	attempt	N	attentat		
1970	heat	N	chaleur		m.r. : hot-dog s.t. : chaud
1968	assume	V	assumer		
1967	touch	V	toucher		
1962	apart	ADV	à part		
1954	green	A	jeune		
1944	difficulty	N	difficulté		
1944	science	N	science		
1944	seat	N	centre		
1940	direction	N	direction		
1935	huge	A	énorme	/j/-/ʒ/	s. : juger
1931	basic	A	basique		
1927	foreign	A	étranger	/f/-/p/	m.r. : impur s. : étranger
1926	final	A	final		
1925	encourage	V	encourager		
1918	detail	N	détail		
1917	bear ¹	V	porter	/b/-/p/	

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
1904	strange	A	étrange		
1898	slightly	ADV	légèrement		s. : souplesse
1891	surface	N	surface		
1882	report	V	rapporter		
1881	certain	A	certain		
1880	image	N	image		
1879	structure	N	structure		
1877	marriage	N	mariage		
1877	occasion	N	occasion		
1876	style	N	style		
1876	telephone	N	téléphone		
1874	express	V	exprimer		
1871	past	N	passé		
1871	deal	V	dealer		
1866	set	N	set		
1865	impossible	A	impossible		
1865	remove	V	enlever	/l/-/r/	r. : éliminer
1862	turn	N	tour		
1860	destroy	V	détruire		
1860	regard	V	considérer		s. : regarder
1859	object	N	objet		
1854	drink	N	boisson		m.r. : trinquer s.t. : boire
1854	freedom	N	liberté		m.r. : free-jazz s. : liberté
1852	term	N	terme		
1850	usual	A	usuel		
1849	entirely	ADV	entièrement		
1846	standard	N	standard		
1844	board	N	bord		
1842	college	N	collège		
1841	bird	N	oiseau	/b/-/v/	m.r. : verdier s. : oiseau
1841	distance	N	distance		
1841	step	N	step		
1836	box ¹	N	box		
1836	box ²	N	buis		s. : buxacées
1832	cause	N	cause		
1832	skin	N	stencil		
1830	interesting	A	intéressant		
1829	represent	V	représenter		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
1826	blue	A	bleu		
1823	interested	A	intéressé		
1823	opportunity	N	occasion		r. : opportunité
1823	technology	N	technologie		
1818	sight	N	spectacle		
1815	whole	N	tout		m.r. : holisme s. : tout
1814	recent	A	récent		
1813	team	N	équipe	/t/-/d/	m.r. : demi s. : sport
1811	condition	N	condition		
1804	star	N	star		
1804	hide	V	cache	/d/-/t/	m.r. : honteux s. : cache
1803	weight	N	poids		m.r. : welter s. : poids
1802	exercise	N	exercice		
1802	place	V	placer		
1800	press	N	presse		
1794	claim	V	réclamer		r. : commander
1792	individual	N	individu		
1792	milk	N	lait		r. : milk-shake
1792	public	N	public		
1792	hit	V	heurter		
1784	king	N	roi		m.r. : khan s.t. : chef
1775	individual	A	individuel		
1775	fast	ADV	vite	/f/-/v/	
1774	round	ADV	autour	/r/-/l/, /d/-/t/	r. : alentour
1773	advantage	N	avantage		
1773	range	N	rangée		
1773	present	V	présenter		
1771	gun	N	canon	/g/-/k/	
1768	aware	A	averti		
1764	treat	V	traiter		
1763	secretary	N	secrétaire		
1761	page ¹	N	page		
1761	page ²	N	page		
1752	dinner	N	dîner		
1751	capital	N	capital		
1751	loss	N	perte		m.r. : ablation s. : perte

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
1748	training	N	entraînement		
1747	admit	V	admettre		
1744	hard	ADV	fort		m.r. : hercule s.t. : force physique
1743	wood	N	bois		m.r. : verdure s.t. : arbre
1742	military	A	militaire		
1740	carefully	ADV	avec circonspection	/p/-/f/	
1732	strength	N	force		m.r. : contrainte r. : force
1729	obvious	A	prévisible	/b/-/p/	
1721	ahead	ADV	en avant		m.r. : hardi s. : en avant
1718	campaign	N	campagne		
1717	ordinary	A	ordinaire		
1716	second ¹	N	second		
1716	second ²	N	seconde		
1716	test	N	test		
1715	fly	V	filer		
1713	care	N	garde	/k/-/g/	
1713	maintain	V	maintenir		
1709	charge	N	charge		
1705	sorry	A	triste	/s/-/t/	
1702	travel	V	voyager		m.r. : traveller's chèque s.t. : voyage
1701	forest	N	forêt		
1698	scheme	N	combine		
1695	danger	N	danger		
1693	affair	N	affaire		
1686	opinion	N	opinion		
1685	resource	N	ressource		
1685	improve	V	profiter de	/v/-/f/	
1683	computer	N	ordinateur		m.r. : compatible s.t. : informatique
1683	stick	V	se coincer		
1679	labour	N	labeur		
1678	branch	N	branche		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
1676	European	A	européen		
1676	recently	ADV	récemment		
1676	adult	N	adulte		
1676	prevent	V	prévenir		
1675	record	N	record		
1674	flower	N	fleur		
1674	pleasure	N	plaisir		
1670	speech	N	parole	/p/-/b/	m.r. : bouche s. : parole
1669	expression	N	expression		
1664	conversation	N	conversation		
1663	generally	ADV	généralement		
1663	wash	V	laver		m.r. : vaisselle s. : laver
1662	series	N	série		
1658	coffee	N	café		
1651	being	N	existence		m.r. : bien-être s. : existence
1650	responsibility	N	responsabilité		
1649	normal	A	normal		
1649	bar	N	barre		
1648	colonel	N	colonel		
1646	outside	ADV	dehors	/d/-/t/	m.r. : sortir s. : dehors
1644	factor	N	facteur		
1643	hole	N	trou	/l/-/n/	m.r. : hernie s.t. : cavité
1643	climb	V	grimper	/g/-/k/, /b/- /p/	
1643	hate	V	haïr	/t/-/s/	m.r. : haïssable s. : haïr
1642	middle	A	du milieu		r. : médius
1642	possibly	ADV	peut-être		s. : possibilité
1638	belief	N	opinion	/b/-/p/, /l/-/n/	
1635	return	N	retour		
1634	born	A	né	/b/-/p/	m.r. : apparition r. : naissance
1633	desk	N	pupitre		m.r. : desserte s.t. : meuble
1633	edge	N	bord	/ʒ/-/k/	m.r. : découpé s. : bord
1632	island	N	île		r. : insulaire

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
1629	grass	N	gazon	/s/-/z/	
1629	meal ¹	N	repas	/l/-/n/	r. : menu
1629	meal ²	N	farine		r. : moulin
1627	tea	N	thé		
1626	due	A	dû		
1626	introduce	V	introduire		
1625	warm	A	chaud		m.r. : vermeil s. : chaud
1625	ring ¹	V	entourer	/r/-/l/	m.r. : élinguer s. : entourer
1625	ring ²	V	sonner		m.r. : dring s.t. : sonnette
1622	economy	N	économie		
1622	matter	V	importer		
1620	election	N	élection		
1620	silence	N	silence		
1614	possibility	N	possibilité		
1611	theatre	N	théâtre		
1606	French	A	français		
1603	survive	V	survivre		
1597	approach	N	approche		
1596	message	N	message		
1596	peace	N	paix		r. : pacifier
1596	revolution	N	révolution		
1590	housing	N	logement		m.r. : hacienda s.t. : habitation
1587	arrange	V	arranger		
1584	safe	A	sûr	/f/-/v/	m.r. : sauver s.t. : danger
1584	memory	N	mémoire		
1582	radio	N	radio		
1581	eventually	ADV	finaleme nt	/v/-/f/	
1581	duty	N	droit		
1581	muscle	N	muscle		
1579	farmer	N	fermier		
1578	lift	V	lever	/f/-/v/	
1578	spread	V	écarter	/d/-/t/	
1577	burn	V	brûler		
1575	gold	N	or		m.r. : golden s.t. : jaune
1574	species	N	espèce		
1573	tooth	N	dent	/t/-/d/, /θ/-/t/	r. : denture

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
1570	ear ¹	N	oreille		m.r. : étrier s. : oreille
1570	ear ²	N	épi		m.r. : orge s. : épi
1567	gain	V	gagner		
1567	press	V	presser		
1565	meet	A	convenable	/t/-/s/	m.r. : malséant s.t. : bienséance
1564	race ¹	N	course		r. : cross
1564	race ²	N	race		
1563	yard ¹	N	yard		
1563	yard ²	N	jardin	/j/-/ʒ/	
1562	dress	N	robe		m.r. : dressing s.t. : vêtement
1561	tiny	A	minuscule		m.r. : tenu s.t. : petit
1561	traditional	A	traditionnel		
1561	design	V	dessiner		
1559	reality	N	réalité		
1557	ability	N	aptitude		r. : habileté
1553	complete	A	complet		
1549	control	V	contrôler		
1545	professor	N	professeur		
1545	publish	V	publier		
1544	key	N	clé		
1544	beat	V	battre		
1543	politics	N	politique		
1542	disease	N	maladie		r. : indisposition
1541	order	V	ordonner		
1535	holiday	N	vacances	/d/-/t/	m.r. : halte s.t. : repos
1535	post ¹	N	poteau		
1535	post ²	N	poste		
1535	post ³	N	poste		
1534	protect	V	protéger		
1530	farm	N	ferme		
1527	aspect	N	aspect		
1527	miss	V	manquer	/s/-/k/	
1524	bottom	N	bas	/t/-/s/	m.r. : abaisser s.t. : inférieur
1523	original	A	original		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
1522	bright	A	brillant		
1517	reply	V	répliquer		
1516	product	N	produit		
1516	property	N	propriété		
1511	compare	V	comparer		
1509	conference	N	conférence		
1508	extent	N	étendue		
1507	cancer	N	cancer		
1507	director	N	directeur		
1507	blow	V	faire voler	/b/-/v/	
1505	pain	N	peine		
1504	shout	V	crier	/ʃ/-/k/, /t/-/s/	r. : crisser
1502	mountain	N	montagne		
1499	basis	N	base		
1498	wild	A	violent		
1498	apparently	ADV	apparemment		
1498	actor	N	acteur		
1495	fairly	ADV	assez	/f/-/p/	r. : plutôt

Tableau de l'hyperlexique allemand/français

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
104092	sein	V	signifier		
69018	werden	V	devenir	/r/-/l/	r. : évoluer
59535	haben	V	avoir	/b/-/p/	m.r. : happer s.t. : attraper
25016	können	V	pouvoir		m.r. : chance s.t. : possibilité
20116	so	ADV	ainsi		
19041	noch	ADV	encore		
17252	nur	ADV	seulement		m.r. : encore s.t. : restriction
12872	Jahr	N	an		m.r. : jour s.t. : durée
12159	sagen	V	signifier		
11138	sollen	V	devoir		m.r. : soldat s.t. : obligation
10723	dann	ADV	alors	/d/-/t/	m.r. : tantôt s.t. : moment
10540	groß	A	grand		
10446	wollen	V	vouloir		
10290	geben	V	donner	/g/-/k/	r. : combler
10026	kommen	V	venir		m.r. : chemin s.t. : se déplacer
8140	neu	A	nouveau		r. : néo-
8089	machen	V	faire		m.r. : machine s.t. : effectuer
7841	gut	A	bon		r. : juste
7302	gehen	V	aller	/g/-/k/	r. : conduire
6987	sehen	V	voir		s. : sens
6915	immer	ADV	toujours		m.r. : immédiatement s.t. : instant
6843	deutsch	A	teuton	/d/-/t/	
6771	stehen	V	tenir		
6413	Mensch ¹	N	homme	/ʃ/-/s/	r. : monsieur
6267	weit	A	vaste		
6194	Zeit	N	date	/ts/-/d/	
5874	Welt	N	monde	/v/-/f/	m.r. : foule s. : monde
5811	sehr	ADV	très	/r/-/l/	r. : salement

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
5577	lassen	V	laisser		
5354	Frage	N	question		r. : affaire
5176	sondern	V	séparer	/n/-/l/	r. : isoler
5174	Tag	N	jour	/t/-/d/	r. : décade
5063	Land	N	land		
4894	Prozent	N	pour cent	/ts/-/s/	
4855	mehr	ADV	plus		m.r. : mer s.t. : grande quantité
4757	bleiben	V	rester	/b/-/p/, /b/-/v/	m.r. : épave s. : rester
4687	finden	V	trouver	/f/-/v/, /d/-/t/	r. : inventer
4572	da	ADV	là		s. : dans
4441	Mann	N	homme	/n/-/l/	s. : mâle
4395	Uhr	N	heure		
4362	liegen	V	être allongé		
4295	wissen	V	savoir	/v/-/b/	m.r. : embrasser s.t. : appréhender
4254	Frau	N	femme	/f/-/p/	r. : parité
4213	alt	A	âgé	/l/-/n/	r. : aîné
4171	stellen	V	mettre		r. : installer
4152	Herr	N	monsieur		m.r. : hiérarque s.t. : titre
3987	dürfen	V	avoir la permission de	/f/-/p/	m.r. : dispense r. : permission
3971	halten	V	faire halte		
3820	Seite	N	sens	/t/-/s/	
3598	sprechen	V	parler		m.r. : prêcher s.t. : parole
3533	Leben	N	vie	/l/-/n/	m.r. : anabiose s. : vie
3518	politisch	A	politique		
3460	Volk	N	peuple		r. : foule
3421	tun	V	faire	/t/-/d/	m.r. : donner s.t. : produire
3420	klein	A	petit	/l/-/n/	r. : nain
3192	Arbeit	N	travail	/b/-/v/	
3176	mögen	V	pouvoir		m.r. : marge s.t. : possibilité
3142	zeigen	V	indiquer	/ts/-/d/, /g/-/k/	
3093	Regierung	N	gouvernement		s. : régime

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
2962	Staat	N	état		
2892	suchen	V	chercher	/z/-/ʃ/	
2884	fahren	V	faire		
2883	stark	A	solide	/r/-/l/	
2830	Hand	N	main		m.r. : handball s. : main
2804	glauben	V	croire		r. : gober
2765	bereits	ADV	déjà	/b/-/p/	s. : auparavant
2765	Teil	N	partie	/t/-/s/	m.r. : cellule s.t. : constitutif
2726	amerikanisch	A	américain		
2709	lang	A	long		
2695	denn	ADV	donc		
2686	Partei	N	parti		
2650	dort	ADV	là	/d/-/t/	m.r. : terre s.t. : lieu
2624	Grund	N	fond	/g/-/k/, /d/-/t/	r. : contenu
2579	gleich	A	égal		
2552	Weg	N	chemin	/v/-/f/, /k/-/ʃ/	r. : fourche
2518	führen	V	porter	/f/-/p/	
2503	Kind	N	enfant		r. : chiard
2461	fragen	V	demander	/f/-/p/, /g/-/j/	r. : prier
2446	Stadt	N	ville		r. : cité
2413	zwar	ADV	certes	/ts/-/s/	
2388	beginnen	V	commencer	/g/-/j/	m.r. : embryon s. : commencer
2381	darauf	ADV	dessus	/d/-/t/	m.r. : thorax s.t. : supérieur
2356	bitten	V	prier	/b/-/v/	s. : inviter
2341	Million	N	million		
2289	natürlich	A	naturel		
2286	Entwicklung	N	développement	/t/-/d/	s. : développer
2278	hören	V	entendre		m.r. : harmonie s.t. : oreille
2269	kurz	A	court		
2259	wichtig	A	important		r. : vital
2250	Woche	N	semaine		r. : week-end
2243	Ende	N	bout		m.r. : nouer s. : bout
2236	jung	A	jeune		
2208	setzen	V	installer		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
2199	sozialistisch	A	socialiste	/ts/-/s/	
2186	Aufgabe	N	tâche	/f/-/v/, /g/-/k/	m.r. : vacation s. : tâche
2185	erhalten	V	recevoir		m.r. : hériter s. : recevoir
2181	erreichen	V	atteindre		m.r. : rencontrer r. : atteindre
2178	nahe	A	proche	/n/-/l/	m.r. : loin s.t. : distance
2167	schwer	A	sévère	/ʃ/-/s/	
2152	wohl	ADV	bien	/v/-/b/	r. : habilement
2149	Beispiel	N	exemple	/b/-/p/	m.r. : pastiche s.t. : imiter
2147	Januar	N	janvier		
2129	bestehen	V	réussir	/b/-/p/	r. : prospérer
2127	denken	V	penser	/d/-/t/, /k/-/g/	m.r. : intriguer s. : penser
2084	Kraft	N	force	/f/-/v/	r. : activité
2065	Mai	N	mai		
2062	sowjetisch	A	soviétique		
2052	fest	A	fort		
2045	scheinen	V	sembler	/ʃ/-/s/, /n/-/l/	
2038	Kampf	N	combat		
2035	Politik	N	politique		
2032	Betrieb	N	entreprise	/b/-/p/	m.r. : patron s. : entreprise
2000	Haus	N	maison	/s/-/t/	r. : hutte
1984	tragen	V	porter	/t/-/d/	s. : diriger
1965	nennen	V	nommer		m.r. : anonyme s.t. : nom
1962	Monat	N	mois		r. : mensuel
1960	spielen	V	jouer	/p/-/b/	m.r. : ballon s. : jouer
1939	Neue	N	nouveau		r. : néo-
1936	einfach	A	facile		
1934	gemeinsam	A	commun	/g/-/k/	
1914	Problem	N	problème		
1907	gelten	V	être valable	/g/-/k/	s. : qualité
1894	Juli	N	juillet		
1894	Stunde	N	heure		r. : instant
1892	Krieg	N	guerre	/k/-/g/	
1870	international	A	international		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
1862	zurück	ADV	en arrière	/ts/-/d/	s. : direction
1824	schreiben	V	écrire	/b/-/p/	r. : scripteur
1810	Auge	N	oeil	/g/-/j/	
1800	nie	ADV	ne ... jamais		m.r. : ne s.t. : négation
1799	brauchen	V	avoir besoin de		r. : réclamer
1790	Gebiet	N	territoire	/g/-/k/, /b/- /p/	m.r. : campagne s.t. : étendue
1786	erklären	V	déclarer		m.r. : clore s. : déclarer
1783	heißen ²	V	hisser		
1780	leicht	A	léger		r. : leste
1778	Lage	N	situation	/g/-/j/	r. : lieu
1769	Bild	N	image	/b/-/p/	m.r. : peinture s.t. : représentation
1766	richtig	A	juste		s. : droit
1764	nein	ADV	non		
1763	Gespräch	N	conversation	/g/-/k/	m.r. : causerie r. : conversation
1750	gewinnen	V	gagner	/g/-/k/	r. : convaincre
1748	Mitglied	N	membre		m.r. : mentalité s. : membre
1746	Arbeiter	N	travailleur	/t/-/s/	r. : bosseur
1744	Art	N	manière		m.r. : art r. : manière
1723	Republik	N	république		
1716	einzel	A	seul		
1711	März	N	mars		
1710	sogar	ADV	même	/z/-/ʃ/	m.r. : changement r. : gradation
1705	Erfolg	N	réussite	/k/-/s/	
1705	gewiß	A	certain	/g/-/k/, /s/-/k/	r. : convaincu
1702	Sinn	N	sens		
1699	französisch	A	français		
1693	ziehen	V	traîner	/ts/-/t/	
1686	treten	V	côcher		m.r. : teinter s.t. : couvrir
1678	voll	A	plein	/f/-/p/	
1668	fallen	V	tomber		m.r. : s'affaler s. : tomber
1664	Präsident	N	président		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
1659	leben	V	vivre	/l/-/n/	m.r. : anabiose s.t. : vie
1653	Name	N	nom		
1652	fordern	V	exiger	/f/-/v/, /r/-/l/	m.r. : vouloir r. : exiger
1647	schnell	A	rapide	/ʃ/-/k/	s. : coule
1643	Kreis	N	cercle		
1635	Geschichte	N	histoire	/g/-/k/, /ʃ/-/z/	m.r. : cause s.t. : événement
1617	Vorsitzende	N	présidente	/f/-/p/, /t/-/d/, /d/-/t/	
1604	Wort	N	mot		r. : verbe
1603	Straße	N	rue	/s/-/t/	m.r. : trottoir s. : rue
1594	Punkt	N	point		
1590	Stelle	N	place		m.r. : installer s.t. : endroit
1586	Preis	N	prix		
1579	Interesse	N	intérêt		
1574	Juni	N	juin		
1570	verstehen	V	entendre	/f/-/p/	m.r. : percuter s.t. : comprendre
1560	frei	A	libre		r. : franc
1558	erwarten	V	attendre		s. : prévoir
1558	Februar	N	février		
1554	Recht	N	droit		
1550	Platz	N	place		
1543	Ziel	N	cible	/ts/-/s/	
1541	verschieden	A	différent	/d/-/t/	m.r. : variante s.t. : différence
1534	weiter	ADV	plus	/v/-/f/	s. : fort
1532	Genosse	N	camarade	/g/-/k/	r. : connaissance
1531	Fall ¹	N	cas		m.r. : filon s.t. : situation
1531	Fall ²	N	drisse	/f/-/p/	s. : palan
1528	sicher	A	sûr		s. : sécurité
1523	treffen	V	trouver	/f/-/v/	
1512	September	N	septembre		
1511	weiß	A	blanche	/v/-/b/, /s/-/ʃ/	
1510	klar	A	clair		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
1509	legen	V	plier	/g/-/j/	
1503	Dezember	N	décembre	/ts/-/s/	
1503	früh	A	précoce	/f/-/p/	m.r. : prompt s.t. : tôt
1503	Werk	N	œuvre		
1498	rund	A	rond		
1479	schaffen ¹	V	créer	/ʃ/-/k/, /f/-/p/	r. : composer
1479	schaffen ²	V	transporter	/ʃ/-/k/, /f/-/v/	r. : convoier
1465	Meinung	N	opinion		m.r. : méninge s.t. : esprit
1449	kennen	V	connaître		
1447	arbeiten	V	bossier	/t/-/s/	
1441	schließlich	ADV	finalemt	/s/-/t/	m.r. : chute s.t. : fin
1435	schön	A	beau	/ʃ/-/k/	r. : canon
1431	April	N	avril		
1428	Minute	N	minute		
1406	Vater	N	paternel	/f/-/p/	
1388	allgemein	A	général	/g/-/k/	m.r. : commun r. : général
1377	Kopf	N	champignon		
1364	viel	ADV	beaucoup		r. : foule
1359	Freiheit	N	liberté		m.r. : racheter s. : liberté
1356	Form	N	forme		
1354	Mädchen	N	filie		m.r. : midinette s. : fille
1338	Raum	N	espace	/r/-/l/	m.r. : limite s.t. : étendue
1337	modern	A	moderne		
1331	schlagen	V	cogner	/ʃ/-/k/	
1325	erscheinen	V	paraître	/ʃ/-/s/	m.r. : scène s. : paraître
1310	notwendig	A	inévitabile	/d/-/t/	
1305	bedeuten	V	signifier	/b/-/v/, /d/- /t/, /t/-/s/	m.r. : avertissement s.t. : signe
1302	warum	ADV	pourquoi	/v/-/f/	m.r. : former s.t. : cause
1295	einzig	A	unique	/ç/-/k/	
1290	laufen	V	courir	/f/-/v/	m.r. : se lever s.t. : se déplacer

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
1282	Oktober	N	octobre		
1281	Vorschlag	N	proposition	/f/-/p/, /ʃ/-/k/	m.r. : procurer s.t. : offrir
1267	Kirche	N	église	/k/-/g/, /r/-/l/	
1266	Stimme	N	voix		m.r. : chanter s. : voix
1264	schließen	V	conclure	/ʃ/-/k/, /s/-/k/	
1257	Meter	N	mètre		
1256	währen	V	durer		r. : vivre
1247	Montag	N	lundi	/t/-/d/	m.r. : mardi s. : lundi
1245	Beziehung	N	relation	/b/-/p/	r. : piston
1242	steigen	V	monter		m.r. : changer s.t. : se déplacer
1237	Freund	N	ami	/f/-/p/	r. : pote
1234	erkennen	V	reconnaître		
1233	Wissenschaft	N	science	/v/-/b/, /s/- /t/, /ʃ/-/k/	r. : botanique
1222	Plan	N	plan		
1217	öffentlich	A	public	/f/-/p/, /n/-/l/	s. : politique
1217	Spiel	N	jeu		r. : sport
1205	Entscheidung	N	décision	/t/-/d/, /ʃ/-/s/	r. : décider
1204	Geld	N	argent		r. : galette
1204	Mittel	N	moyen		r. : méthode
1195	Gruppe	N	groupe		
1189	technisch	A	technique		
1187	britisch	A	britannique		
1187	Zahl	N	nombre	/ts/-/d/	m.r. : duel s. : nombre
1185	besonder	A	spécial	/b/-/p/, /n/-/l/	
1183	Leistung	N	performance	/t/-/d/	m.r. : leader s.t. : compétition
1183	ständig	A	constant	/d/-/t/	
1169	Erfahrung	N	expérience	/f/-/p/	
1168	europäisch	A	européen		
1168	gelingen	V	réussir	/g/-/k/	m.r. : culminant s.t. : atteindre
1168	genau	A	exact	/g/-/k/	s. : conforme
1165	Korrespondent	N	correspondant		
1160	Zukunft	N	avenir	/t/-/d/	m.r. : secondaire

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
					s. : avenir
1156	tief	A	profond		m.r. : tréfonds s. : profond
1155	bieten	V	offrir	/t/-/d/	m.r. : abandonner s.t. : donner
1152	später	A	postérieur		
1148	völlig	A	entier	/f/-/p/	m.r. : plein s. : entier
1145	Bericht	N	bulletin	/r/-/l/	
1142	fast	ADV	presque	/f/-/p/	
1142	Telefon	N	téléphone		
1137	bestimmt	A	décidé	/b/-/p/	m.r. : prêt r. : décidé
1137	Erde	N	terre		m.r. : ados s. : terre
1136	Grenze	N	frontière		
1133	folgen	V	suivre		r. : filer
1131	geschehen	V	se passer	/g/-/k/	m.r. : croche s.t. : durée
1130	eben	A	plan	/b/-/p/	
1123	verkaufen	V	vendre	/f/-/p/, /f/-/v/	m.r. : percevoir s.t. : somme d'argent
1116	Frieden	N	paix	/d/-/t/	m.r. : fraternité s.t. : rapports
1113	Tendenz	N	tendance		
1112	Situation	N	situation		
1104	offen	A	franc		
1098	Schule	N	école	/ʃ/-/k/	
1091	Buch	N	bouquin		
1091	Leute	N	gens		m.r. : clientèle s. : gens
1084	Bedeutung	N	importance	/b/-/p/, /d/-/t/, /t/-/s/	
1083	darin	ADV	dedans	/d/-/t/	s. : intérieur
1082	gar	A	à point	/g/-/k/	m.r. : cru s.t. : cuit
1069	Gefahr	N	danger	/g/-/k/	m.r. : coffre s.t. : sécurité
1068	Minister	N	ministre		
1068	rechnen	V	calculer		m.r. : précompter s. : calculer

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
1063	Schritt	N	pas		r. : trotter
1061	Zeitung	N	journal	/ts/-/t/	m.r. : titre s. : journal
1058	Polizei	N	police		
1052	richten	V	braquer		
1049	Blick	N	regard		r. : loucher
1049	Gesetz	N	loi	/g/-/k/	r. : clause
1047	Wirtschaft	N	économie	/v/-/f/	m.r. : fortune s.t. : biens
1041	Macht	N	puissance		m.r. : machisme s.t. : dominer
1038	ebenso	ADV	pareillement	/b/-/v/	m.r. : à l'avenant r. : pareillement
1036	Gott	N	dieu	/g/-/k/	s. : créateur
1035	Rede	N	parole	/d/-/t/	m.r. : orateur s. : parole
1035	verlangen	V	demander		m.r. : fellaga s.t. : obtenir
1034	Einheit	N	unité		
1031	deutlich	A	clair	/t/-/s/	m.r. : docile s.t. : facile
1030	recht	A	droit		
1022	Mark ¹	N	moelle	/r/-/l/	
1022	Mark ²	N	marche		
1022	Mark ³	N	mark		
1020	Anfang	N	début		m.r. : enfance r. : début
1019	Tod	N	décès	/t/-/d/, /t/-/s/	
1016	Lösung	N	solution		m.r. : élucider s.t. : difficulté
1014	anders	ADV	autrement	/d/-/t/	
1012	Diskussion	N	discussion		
1010	Rolle	N	rôle		
1008	menschlich	A	humain	/ʃ/-/z/	m.r. : humaniser s. : humain
1003	Bereich	N	portée	/b/-/p/	
1003	Besuch	N	visite	/b/-/v/	
997	Wunsch	N	désir	/ʃ/-/z/	r. : visée
992	bereit	A	prêt	/b/-/p/	
990	Höhe	N	hauteur		r. : haut
989	warten	V	soigner	/v/-/f/	m.r. : infirmerie s. : soigner

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
987	Folge	N	suite		r. : fil
983	Professor	N	professeur		
980	Wahl	N	choix	/v/-/b/	m.r. : aboulie s.t. : choisir
979	Satz	N	set		
977	Industrie	N	industrie		
971	demokratisch	A	démocrate		
969	Gedanke	N	pensée	/g/-/k/, /d/-/t/	m.r. : critique s.t. : esprit
969	verlieren	V	perdre	/l/-/n/	m.r. : faner s. : perdre
968	November	N	novembre		
963	Antwort	N	réponse	/t/-/d/, /r/-/n/	m.r. : devinette s. : réponse
963	Versuch	N	épreuve	/f/-/p/	
957	entscheiden	V	décider	/t/-/d/, /ʃ/-/s/	
956	bald	ADV	bientôt		
956	Nacht	N	nuît		
956	Tatsache	N	fait		r. : trait
944	lesen	V	lire		r. : lecteur
943	gerade	A	égal	/r/-/l/	
943	gleichzeitig	A	simultané	/g/-/k/, /ts/- /d/	m.r. : coïncident r. : simultané
943	Kollege	N	collège		
929	gering	A	petit	/r/-/n/	r. : jeune
929	Maßnahme	N	mesure	/s/-/z/	
928	rufen	V	crier	/f/-/p/	m.r. : rappel s.t. : cri
927	eng	A	étroit		m.r. : engoncer s. : étroit
927	erfüllen	V	remplir	/f/-/p/	
927	Foto	N	photo		
926	Bad	N	bain	/t/-/s/	r. : ablution
924	bekannt	A	connu	/b/-/p/, /k/-/s/	m.r. : pressentiment s. : connu
919	Wert	N	valeur	/r/-/l/	
918	gerade	ADV	justement	/g/-/k/, /r/-/l/	m.r. : équilibré s. : justement
916	Reihe	N	rang		
915	wahrscheinlich	A	vraisemblable	/ʃ/-/s/	
913	erinnern	V	rappeler		s. : renaître

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
913	Erklärung	N	déclaration	/r/-/l/, /k/-/g/	m.r. : logorrhée s.t. : discours
913	Umsatz	N	chiffre d'affaires		s. : montant
909	Boden	N	sol	/b/-/p/, /d/-/t/	r. : patrie
908	Freitag	N	vendredi	/f/-/v/, /t/-/d/	
908	Sieg	N	victoire		s. : succès
907	bilden	V	produire	/b/-/p/	
906	reden	V	parler		r. : radoter
904	Programm	N	programme		
903	Vertrag	N	contrat	/f/-/p/	r. : prêt
903	Wohnung	N	logement	/v/-/f/	m.r. : fournil s.t. : local
903	zahlreich	A	nombreux	/ts/-/t/	m.r. : téra- s.t. : grand nombre
902	werfen	V	jeter	/v/-/b/, /r/-/l/	r. : balancer
901	Herz	N	cœur		m.r. : hardiesse r. : cœur
900	leisten	V	faire	/t/-/d/	r. : glander
898	Person	N	personne		
892	Gast	N	client	/g/-/k/	
892	persönlich	A	personnel		
890	Konferenz	N	conférence		
889	schlecht	A	mauvais	/ʃ/-/s/	r. : sale
885	Beschluß	N	décision	/b/-/v/, /s/-/t/	r. : volonté
883	Sicherheit	N	sécurité		
878	langen	V	passer		r. : longer
876	Sprache	N	parole		
875	lösen	V	détacher		m.r. : lacer s.t. : attacher
866	praktisch	A	pratique		
866	Produktion	N	production		
862	Maschine	N	machine		
860	Wagen	N	bagnole	/v/-/b/, /n/-/l/	
859	fühlen	V	sentir		m.r. : feeling s. : sentir
858	sterben	V	trépasser	/b/-/p/	
856	gestern	ADV	hier	/t/-/d/	m.r. : agenda s.t. : jour
855	wirken	V	faire	/v/-/f/	
854	Ansicht	N	opinion		s. : assertion

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
850	Jugend	N	jeunesse		m.r. : gigolo s.t. : jeune
847	Brief	N	lettre	/b/-/p/, /r/-/l/	r. : pli
846	erzählen	V	raconter	/ts/-/d/	m.r. : délayer s.t. : exposer
843	Stück	N	pièce		r. : chacun
838	Ausdruck	N	expression		
838	ruhig	A	tranquille	/ç/-/k/	
835	weisen	V	montrer		m.r. : visiter r. : montrer
832	unmittelbar	A	immédiat	/l/-/n/	m.r. : maintenant s.t. : moment présent
831	wahr	A	vrai		
830	bauen	V	construire		m.r. : benne s.t. : matériaux
829	Bank ¹	N	banc		
829	Bank ²	N	banque		
829	Fortsetzung	N	poursuite	/f/-/p/	
824	Universität	N	université		
824	Unternehmen	N	entreprise		m.r. : intérim s. : entreprise
820	inzwischen	ADV	entre-temps	/ts/-/t/, /ʃ/-/k/	m.r. : entrecouper s.t. : intervalle
818	früher	A	ancien	/f/-/p/, /r/-/l/	m.r. : paléo- s. : ancien
813	Dame	N	dame		
812	gern	ADV	volontiers		s. : gré
812	wählen	V	choisir	/v/-/b/	m.r. : aboulie s. : choisir
811	Technik	N	technique		
806	dienen	V	servir	/d/-/t/, /n/-/l/	s. : utile
806	Film	N	film		
806	Thema	N	thème		
805	Mal ¹	N	tache		r. : maille
805	Mal ²	N	fois		r. : multiple
801	Kilometer	N	kilomètre		
797	Mitte	N	milieu		r. : mitan
791	Zimmer	N	pièce	/ts/-/d/	m.r. : demeure s.t. : habitation
790	Abend	N	soir		r. : bonsoir

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
785	besser	A	meilleur	/s/-/t/	s. : bonté
783	Bedingung	N	condition	/b/-/p/, /d/-/t/	r. : prétention
782	Begriff	N	notion	/b/-/p/	r. : a priori
782	genau	ADV	exactement	/g/-/k/	m.r. : conforme r. : exact
778	sozial	A	social	/ts/-/s/	
773	stimmen	V	accorder		s. : instrument
770	Pol	N	pôle		
770	Sozialismus	N	socialisme	/ts/-/s/	
768	besitzen	V	posséder	/b/-/p/, /t/-/d/	
767	falsch	A	fausse	/ʃ/-/s/	
766	Grad	N	grade		
765	System	N	système		
764	Verfügung	N	ordonnance	/g/-/k/	m.r. : verdict r. : décision
763	entscheidend	A	décisif	/t/-/d/, /ʃ/-/s/	s. : décide
762	Tier	N	animal	/t/-/s/	m.r. : serpent s. : animal
759	täglich	A	quotidien	/t/-/d/	m.r. : décade s.t. : jour
756	fehlen	V	manquer		r. : faillir
754	Gewalt	N	force	/g/-/k/	r. : activité
753	Firma	N	firme		
753	verbinden	V	panser	/f/-/p/, /b/-/p/, /d/-/t/	s. : propreté
753	Wahrheit	N	vérité		
750	General	N	général		
746	ehemalig	A	ancien		m.r. : immuable s.t. : longtemps
745	tatsächlich	A	réel	/ç/-/k/	r. : authentique
742	Fortschritt	N	progrès	/f/-/p/, /ʃ/-/s/	m.r. : prospérité s. : progrès
742	kommunistisch	A	communiste		
742	Ordnung	N	ordre		
742	Zustand	N	état	/ts/-/d/	r. : destin
740	kaufen	V	acheter	/f/-/p/	m.r. : accaparer s. : acheter
739	Tür	N	porte		m.r. : trou s.t. : ouverture
739	wenig	ADV	peu	/v/-/b/	r. : brin
735	best	A	meilleur		s. : bonté
735	Stellung	N	position		s. : installation

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
732	betrachten	V	regarder		m.r. : abstraction s.t. : considérer
731	Arm	N	bras		m.r. : armé s. : bras
730	Tat	N	action	/t/-/s/	m.r. : sorte s. : action
729	Richtung	N	direction		m.r. : orienter s. : direction
725	Dollar	N	dollar		
725	erzielen	V	réaliser		
725	militärisch	A	militaire		
724	Leitung	N	direction	/t/-/d/	m.r. : leader s.t. : entreprise
722	ähnlich	A	semblable		r. : analogue
721	Student	N	étudiant		
720	lernen	V	apprendre		m.r. : lire s.t. : s'instruire
720	schwarz	A	noir	/ʃ/-/s/, /v/-/b/	s. : sombre
720	vorher	ADV	au préalable	/f/-/p/, /r/-/l/	
718	Grundlage	N	base	/g/-/k/	m.r. : condition r. : base
718	Milliarde	N	milliard		
718	Nähe	N	proximité	/n/-/l/	m.r. : loin s.t. : distance
717	Musik	N	musique		
717	Zug	N	section		
714	Familie	N	famille		
712	offenbar	A	palpable	/f/-/p/, /n/-/l/, /b/-/p/, /r/-/l/	
711	Beginn	N	commencement		r. : bégaiement
710	langsam	A	long		
710	Natur	N	nature		
709	Bett	N	lit	/b/-/p/, /t/-/d/	r. : paddock
706	direkt	A	direct		
705	Chance	N	chance		
705	national	A	national		
705	Postfach	N	boîte postale		s. : poste
703	Klasse	N	classe		
701	Handel	N	commerce		m.r. : hanse s. : commerce
700	Hotel	N	hôtel		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
700	Rahmen	N	ramette		
699	Bürger	N	bourgeois		
699	Soldat	N	soldat		
698	beschließen	V	décider	/b/-/v/, /t/-/s/	m.r. : volonté s.t. : décision
697	erfolgen	V	s'ensuivre	/n/-/l/	m.r. : infaillible s.t. : conséquence
693	vorhanden	A	existant	/f/-/p/	m.r. : produire s.t. : exister
691	oben	ADV	en haut		m.r. : oblong s.t. : haut
691	Organisation	N	organisation		
689	wünschen	V	désirer	/n/-/l/	r. : vouloir
688	Auto	N	auto		
688	einsetzen	V	instituer		
687	teilen	V	diviser	/t/-/d/	m.r. : delta s. : diviser
685	heben	V	soulever	/b/-/p/	m.r. : hypso- s.t. : hauteur
684	Einfluß	N	influence		
684	Geist	N	esprit	/g/-/k/	r. : conscience
684	sichern	V	assurer		r. : sécuriser
682	Wochenende	N	week-end		
681	Liebe	N	amour		m.r. : libido s.t. : sexuel
677	Element	N	élément		
677	polnisch	A	polonais		
674	Beratung	N	consultation	/b/-/p/	m.r. : aparté s.t. : réunion
673	durchaus	ADV	absolument	/d/-/t/	r. : entièrement
673	Mannschaft	N	hommes		s. : humanité
673	Ort ¹	N	lieu	/t/-/l/	
673	Ort ²	N	front de taille	/t/-/d/	m.r. : redan s.t. : surface verticale
671	Aufbau	N	construction	/b/-/v/	m.r. : vau s. : construction
670	englisch	A	anglais		
669	Amerikaner	N	Américain		
669	äußern	V	dire		r. : assurer
669	Delegation	N	délégation		
669	Rat	N	conseil	/r/-/l/, /t/-/s/	m.r. : leçon

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
					s. : conseil
668	zwingen	V	forcer	/ts/-/t/	m.r. : traîner s. : forcer
664	Opfer	N	offrande		
662	Nation	N	nation		
661	ändern	V	changer	/d/-/t/	s. : autre
661	Kunst	N	art	/n/-/l/	r. : sculpture
659	Linie	N	ligne		
659	überall	ADV	partout	/b/-/p/	
657	fassen	V	prendre	/f/-/p/	s. : posséder
657	Wesen	N	nature	/v/-/f/	m.r. : foncier s. : nature
656	staatlich	A	public		s. : ensemble
654	rot	A	rousse	/t/-/s/	
653	Idee	N	idée		
650	lachen	V	rire	/x/-/g/	m.r. : allégresse s.t. : gaieté
649	Sorge	N	souci		s. : absorbé
648	Luft	N	air	/t/-/d/	s. : fluide
647	friedlich	A	pacifique		m.r. : froideur r. : calme
646	Erkenntnis ¹	N	connaissance	/t/-/s/	
645	Bezirk	N	district	/b/-/p/, /ts/-/t/	m.r. : patriarcat s.t. : circonscription
645	König	N	roi		m.r. : khan s.t. : chef
644	Ausland	N	étranger	/s/-/t/	
644	besuchen	V	visiter	/b/-/p/	m.r. : prospection s. : visiter
640	Gericht	N	mets	/g/-/k/	m.r. : curry s. : mets
640	scharf	A	coupant	/ʃ/-/k/, /f/-/p/	
636	ökonomisch	A	économique		
636	Theater	N	théâtre		
635	Auffassung	N	conception	/f/-/v/, /s/-/t/	m.r. : inventer r. : concevoir
635	erneuern	V	rénover		
635	Prozeß	N	procès	/ts/-/s/	
634	erheblich	A	considérable	/r/-/l/, /b/-/v/	m.r. : élevé r. : considérable
633	hart	A	dur		r. : acharné

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
631	Tätigkeit	N	activité	/t/-/d/	s. : tendance
630	Haltung	N	attitude		m.r. : hauteur s. : attitude
629	fliegen	V	piloter	/f/-/p/, /g/-/k/	m.r. : piqué s.t. : avion
629	freilich	ADV	certes		m.r. : formellement r. : certainement
628	inner	A	intérieur		
627	Schüler	N	écolier	/ʃ/-/k/	
626	Verbindung	N	liaison	/r/-/n/, /d/-/t/	m.r. : affinité s. : liaison
623	historisch	A	historique		
622	ernst	A	sérieux		r. : rassis
622	insbesondere	ADV	spécialement	/b/-/p/, /n/-/l/	
621	Licht	N	clarté		
621	morgen	ADV	demain		m.r. : mardi s.t. : jour
619	Führung	N	port	/f/-/p/	
619	kämpfen	V	combattre		
619	Markt	N	marché		
619	stets	ADV	toujours		m.r. : tout s.t. : totalité
618	begrüßen	V	saluer	/b/-/p/, /r/-/l/, /s/-/t/	m.r. : politesse r. : civilité
618	greifen	V	saisir	/f/-/p/	m.r. : agripper s. : saisir
618	Zuschauer	N	spectateur	/ʃ/-/k/, /r/-/l/	s. : spectacle
617	Dienst	N	service	/t/-/d/	m.r. : décidé s.t. : décision
617	öffnen	V	ouvrir	/f/-/v/	
616	Beitrag	N	contribution	/b/-/p/	s. : impôt
616	geistig	A	intellectuel	/g/-/k/	m.r. : acuité s. : intellectuel
616	interessant	A	intéressant		
616	reichen	V	donner		r. : troquer
615	Angabe	N	indication	/g/-/k/	
614	eröffnen	V	ouvrir	/f/-/v/	r. : entrouvrir
610	mitteilen	V	communiquer		m.r. : mutualiser s.t. : commun
609	Abgeordnete	N	député	/n/-/l/	r. : parlement
609	ausländisch	A	étranger	/s/-/t/	

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
609	Gesicht	N	figure	/g/-/k/	m.r. : caisson s.t. : tête
608	Angst	N	angoisse		
606	erheben	V	élever	/r/-/l/, /b/-/v/	
603	wenden	V	faner	/v/-/f/	
602	Auftrag	N	ordre	/t/-/d/	
602	stoßen	V	choquer	/s/-/k/	
600	Gemeinschaft	N	communauté	/g/-/k/	
599	Mutter	N	mère		r. : mater
599	Tisch	N	table	/t/-/d/, /ʃ/-/s/	r. : desserte
598	Feld	N	peloton	/f/-/p/	
597	erfahren	V	éprouver	/f/-/p/	
597	privat	A	privé		
596	Verlag	N	éditeur	/f/-/p/	s. : publication
595	Institut	N	institut		
594	Eindruck	N	impression	/d/-/t/, /k/-/s/	r. : intuition
594	Forschung	N	recherche	/f/-/p/, /ʃ/-/s/	r. : poursuite
591	Fuß	N	patte	/f/-/p/, /s/-/t/	
590	breit	A	large	/b/-/v/	r. : vaste
590	Kosten ¹	N	gustation	/k/-/g/	
590	Kosten ²	N	coût		
589	rein	A	pur		m.r. : aryen s. : pur
586	liefern	V	livrer	/f/-/v/	
585	Prinzip	N	principe		
583	Demokratie	N	démocratie		
583	Gefühl	N	sentiment	/g/-/k/, /f/-/v/	m.r. : équivoque s.t. : jugement
582	lieben	V	aimer	/b/-/v/	m.r. : alcôve s.t. : amoureux
581	Schluß	N	clôture	/ʃ/-/k/, /s/-/t/	
580	Abschluß	N	clôture	/ʃ/-/k/, /s/-/t/	
578	Urteil	N	arrêt		
577	Ruf	N	cri	/f/-/v/	r. : bravo
576	Tochter	N	filles	/t/-/d/	m.r. : adopter s. : fille
575	Position	N	position		
575	Schiff	N	bateau	/ʃ/-/s/	m.r. : skif s. : bateau
574	Anteil	N	part	/n/-/l/	r. : lot
574	bewußt	A	conscient	/b/-/p/	s. : éprouve

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
574	vorstellen	V	avancer	/f/-/p/	m.r. : péristaltisme s. : avancer
573	Anlage	N	aménagement		m.r. : aligner s.t. : disposer
573	erleben	V	voir	/b/-/v/	m.r. : revoir s.t. : être en présence de
572	darunter	ADV	dessous	/d/-/t/, /r/-/l/	m.r. : talon s.t. : inférieur
571	Sitzung	N	séance	/ts/-/s/	
570	Menschheit	N	humanité	/ʃ/-/z/	m.r. : humaniser s.t. : humain
569	Dorf	N	village	/d/-/t/	r. : trou
568	bedeutend	A	important	/b/-/p/, /d/-/t/	
567	Glück	N	chance		m.r. : gloire r. : succès
567	vorsehen	V	prévoir	/f/-/p/, /z/-/s/	r. : pressentir
566	zahlen	V	payer		r. : salaire
565	schwierig	A	compliqué	/ʃ/-/k/, /r/-/l/, /ç/-/k/	
565	Vorstand	N	président	/f/-/p/, /t/-/d/	
562	antworten	V	répondre	/v/-/f/, /t/-/s/	m.r. : interface s.t. : communication
562	Motor	N	moteur		
562	offiziell	A	officiel	/ts/-/s/	
561	Kenntnis	N	connaissance	/t/-/s/	
558	Gegner	N	adversaire	/g/-/k/	r. : concurrent
557	Krankenhaus	N	hôpital		m.r. : clinicien s.t. : malade
555	Ernst	N	sérieux		m.r. : triste s.t. : qui ne rit pas
553	Freundschaft	N	amitié	/n/-/l/, /t/-/s/	s. : relations
553	Presse	N	presse		
547	Meister	N	maître		
547	vorbereiten	V	préparer	/f/-/p/, /b/-/p/	r. : préparation
546	Schicksal	N	sort		r. : chance
545	eintreten	V	entrer	/t/-/d/	r. : s'introduire
545	Vorbereitung	N	préparation	/f/-/p/, /b/-/p/	
544	Esse	N	cheminée	/s/-/t/	r. : âtre
544	verurteilen	V	condamner	/f/-/p/, /r/-	m.r. :

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
				/n/, /t/-/s/	punissable s.t. : punition
543	fort	ADV	parti	/f/-/p/	
542	Sonne	N	soleil	/n/-/l/	
541	Kommission	N	commission		
541	lang	ADV	longuement		
537	selten	A	rare		r. : seul
534	Fenster	N	fenêtre		
531	lange	ADV	longuement		
531	wachsen ¹	V	avancer		
531	wachsen ²	V	farter	/v/-/f/, /s/-/t/	
530	Artikel	N	article		
529	Osten	N	est		
528	Bruder	N	frère	/d/-/t/	r. : fratrie
528	treiben	V	rabattre		
526	Kurs	N	cours		
524	weg	ADV	parti		m.r. : vacant s.t. : absent
523	schwach	A	faible	/x/-/g/	s. : vigueur
523	Unterstützung	N	assistance	/t/-/s/	
522	rasch	A	preste	/ʃ/-/s/	
520	Methode	N	méthode		
520	Qualität	N	qualité		
520	West	N	ouest		
519	Beruf	N	profession	/b/-/p/	
519	Graf	N	comte	/g/-/k/	m.r. : couronne s.t. : noblesse
518	Freude	N	joie	/d/-/t/	r. : fête
518	Mehrheit	N	pluralité	/t/-/d/	m.r. : myriade s.t. : grand nombre
516	Vertrauen	N	confiance	/f/-/p/, /t/-/s/	s. : espérance
515	Doktor	N	docteur		
515	durchführen	V	exécuter	/d/-/t/, /f/-/p/	s. : interpréter
515	italienisch	A	italien		
515	Kommunist	N	communiste		
514	bloß	A	nu	/b/-/p/	s. : poils
514	Menge	N	quantité		m.r. : monceau s. : quantité
513	konkret	A	concret		
513	stetig	A	continu		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
513	Verwaltung	N	administration	/f/-/p/	m.r. : perception s. : administration
511	Gemeinde	N	commune	/g/-/k/	
509	Tagung	N	réunion	/t/-/d/, /g/-/k/	m.r. : discours s. : réunion
508	Künstler	N	artiste	/n/-/l/	r. : sculpteur
507	Kritik	N	critique		
507	Wirkung	N	effet	/v/-/f/	r. : fruit
506	Front	N	front		
506	Titel	N	titre		
504	darum	ADV	autour	/d/-/t/	
504	Regel	N	règle		
504	zählen	V	dénombrer	/ts/-/d/, /l/-/n/	
503	beraten	V	délibérer	/b/-/p/	m.r. : parti s.t. : décision
503	kalt	A	froid	/k/-/ʒ/	r. : gelé
503	Wasser	N	eau	/v/-/f/, /s/-/t/	r. : flotte
502	Aktie	N	action		
502	Schwester	N	sœur	/ʃ/-/k/	m.r. : cousin s. : sœur
497	Aktion	N	action		
497	verändern	V	changer	/d/-/t/	m.r. : versatile s. : changer
496	Wissen	N	savoir	/v/-/b/	m.r. : bases s.t. : connaissances
495	Abteilung ¹	N	division	/t/-/d/	m.r. : diffluence s. : division
495	Abteilung ²	N	bureau		r. : burlingue
495	Parlament	N	parlement		
494	aktiv	A	actif		
494	Berliner	N	Berinois		
493	knapp	A	rare	/p/-/b/	s. : nombre
493	verbessern	V	perfectionner	/f/-/p/	
492	betonen	V	accentuer	/b/-/p/, /t/-/s/	m.r. : passion s.t. : intensité
492	Umstand	N	circonstance	/t/-/s/	s. : situation
491	Politiker	N	politique		
491	Untersuchung	N	examen		m.r. : antisèche s. : examen
490	Sekretär	N	secrétaire		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
489	dunkel	A	sombre	/d/-/t/, /k/-/ʒ/	r. : tragique
488	Mama	N	maman		
487	Verband	N	formation		
482	Bildung	N	culture	/b/-/p/, /d/-/t/	m.r. : épistémè s.t. : ensemble des connaissances
482	Jahrzehnt	N	décennie	/r/-/n/	m.r. : jeunesse s.t. : période
482	keineswegs	ADV	en aucune façon	/v/-/f/	
482	Zone	N	zone		
481	Waffe	N	arme	/v/-/b/, /f/-/v/	m.r. : brave s.t. : ennemi
480	beziehen	V	prendre	/b/-/p/, /ts/-/d/	
480	Wald	N	forêt	/v/-/f/	r. : futaie
479	Sprecher	N	speaker		
474	Bischof	N	bischof		
474	trinken	V	boire		m.r. : trinquer s. : boire
474	Verbesserung	N	perfectionnement	/f/-/p/, /r/-/n/	
473	wohnen	V	habiter		m.r. : avenue s.t. : habitation
472	Ausgabe	N	dépense	/b/-/p/	m.r. : gaspiller s.t. : dépenser
472	entdecken	V	découvrir	/t/-/d/, /d/-/t/	r. : détecter
472	Lauf	N	course	/f/-/v/	m.r. : lièvre s. : course
471	Anlaß	N	raison		m.r. : analyse s.t. : expliquer
470	Unterschied	N	distinction		s. : altérité
469	verfolgen	V	persécuter	/f/-/p/, /g/-/k/	
468	Einsatz	N	mise		m.r. : encaisser s.t. : argent
468	Hinweis	N	indication	/v/-/b/, /s/-/t/	m.r. : habitus s. : indication
468	Stand	N	stand		
468	tot	A	décédé	/t/-/d/, /t/-/s/	
468	Zeichen	N	signe	/ts/-/d/, /ç/-/k/	m.r. : indiquer s. : signe
467	Sport	N	sport		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
467	Wille	N	volonté		
466	dauern ¹	V	durer		
466	drehen	V	tourner	/d/-/t/	
463	Gang	N	canal	/g/-/k/	
463	verwenden	V	employer	/f/-/p/	m.r. : prendre s. : employer
461	Mond	N	lune		m.r. : monde s.t. : Terre
460	Kultur	N	culture		
460	prüfen	V	éprouver	/f/-/v/	
459	Post	N	poste		
458	Ereignis	N	événement	/t/-/n/, /n/-/l/	r. : annales
457	spüren	V	éprouver		
456	Morgen ¹	N	matin	/r/-/n/	s. : minuit
456	Morgen ²	N	arpent de terre	/g/-/k/	m.r. : métrique s.t. : mesure
456	Morgen ³	N	lendemain	/g/-/k/	m.r. : mercredi s.t. : jour
453	glücklich	A	heureux	/k/-/s/	r. : juste
452	Vorgang	N	affaire		
451	Bau	N	construction		m.r. : abri s. : construction
450	Persönlichkeit	N	personnalité		
449	beobachten	V	observer	/b/-/v/	s. : observation
449	schießen	V	shooter	/s/-/t/	
448	Papier	N	papier		
447	Beamte	N	fonctionnaire	/b/-/p/, /t/-/s/	s. : personne
447	Krankheit	N	maladie		r. : rechute
446	Ausschuß	N	comité	/ʃ/-/s/, /s/-/t/	s. : société
446	Flugzeug	N	avion	/f/-/p/, /ts/-/t/	r. : pilotage
445	Größe	N	grosseur		
445	Tonne	N	tonne		
444	erfahren	A	expérimenté	/f/-/p/	s. : expérience

Tableau de l'hyperlexique néerlandais/français

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
474965	hebben ¹	V	avoir	/b/-/p/	m.r. : happer s.t. : attraper
321883	worden ¹	V	devenir	/r/-/l/	r. : évoluer
321883	worden ³	V	faire	/r/-/l/	s. : valeur
245471	er	ADV	là	/r/-/l/	
221958	zullen	V	devoir		m.r. : soldat s.t. : obligation
209812	kunnen	V	pouvoir		m.r. : chance s.t. : possibilité
183791	ook	ADV	aussi	/k/-/s/	
149233	nog	ADV	encore	/x/-/k/	
137589	gaan ¹	V	aller	/x/-/k/	r. : conduire
137589	gaan ²	V	aller	/x/-/k/, /n/-/l/	m.r. : coller r. : aller
135035	komen	V	venir		m.r. : chemin s.t. : se déplacer
134802	zeggen	V	dire	/z/-/ʃ/, /x/-/ʃ/	r. : chuchoter
124471	dan	ADV	alors	/d/-/t/	m.r. : tantôt s.t. : moment
123613	zo	ADV	ainsi		
123371	moeten ¹	V	devoir		m.r. : impôt s.t. : obligation
123371	moeten ²	V	vouloir		m.r. : maître s.t. : volonté
98375	zien ¹	V	voir		s. : sens
98375	zien ²	V	voir		m.r. : sens s. : voir
96976	wel	A	bien		s. : valeur
87505	op	A	fini	/p/-/b/	m.r. : bout r. : terme
86305	maken	V	faire	/k/-/s/	m.r. : mission s. : faire
79616	staan	V	être debout		m.r. : soutenir s.t. : debout
79308	goed	A	bon		r. : juste
76697	willen	V	vouloir		
72886	ander	A	autre	/d/-/t/	
71277	doen	V	faire		m.r. : donner s.t. : produire

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
66787	laten	V	laisser	/t/-/s/	
66735	weten	V	savoir		m.r. : averti s.t. : être au courant
65639	daar	ADV	là	/d/-/t/	m.r. : terre s.t. : lieu
61767	meer	A	plus		m.r. : mer s.t. : grande quantité
58538	denken	V	penser	/d/-/t/, /k/-/g/	m.r. : intriguer s. : penser
58078	mens ¹	N	monde		
57175	toch	ADV	donc	/t/-/d/, /x/-/k/	
54472	geven	V	donner	/x/-/k/, /v/-/f/	s. : confier
52837	blijven ¹	V	rester	/b/-/p/	m.r. : épave s. : rester
52837	blijven ²	V	rester	/b/-/p/	m.r. : épave s. : rester
52618	nu	ADV	maintenant		m.r. : an s.t. : temps
51661	groot	A	grand		
51631	vinden	V	inventer	/d/-/t/	
51564	kijken	V	regarder		r. : chercher
50439	man	N	mâle	/n/-/l/	
48452	jaar	N	année		m.r. : jour s.t. : durée
45945	tijd	N	temps	/t/-/d/	r. : date
45929	krijgen	V	recevoir	/x/-/k/	m.r. : trinquer r. : recevoir
45810	zitten	V	être		r. : se sentir
43573	hand	N	main		m.r. : handball s. : main
42525	lang	A	long		
40957	liggen	V	être allongé		
40727	kind	N	enfant		r. : chiard
39630	dag	N	jour	/x/-/k/	r. : décade
39022	af	A	éreinté	/f/-/v/	m.r. : vieux s.t. : fatigué
38126	vrouw	N	femme		r. : avortement
35984	door	ADV	tout	/d/-/t/	s. : entièrement
35539	lopen ¹	V	marcher	/p/-/b/	m.r. : libre s.t. : se mouvoir
34058	vragen	V	demander		r. : briguer

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
33173	horen ¹	V	entendre		m.r. : harmonie s.t. : oreille
33173	horen ²	V	entendre		m.r. : harmonie s.t. : oreille
33096	altijd	ADV	toujours		
32729	nieuw	A	nouveau		
31729	wezen	V	exister	/w/-/g/	
31168	laat	A	tard	/l/-/n/	m.r. : nuitard s. : tard
31024	heel	ADV	entièrement		m.r. : holo- s.t. : entier
30692	even	A	pair	/v/-/b/	m.r. : binaire s.t. : divisible par deux
30131	mogen ²	V	aimer	/x/-/k/	m.r. : amical s.t. : amitié
29510	aan	A	mis		m.r. : nu s.t. : vêtement
29234	voelen ¹	V	sentir	/v/-/f/	m.r. : feeling s. : sentir
29234	voelen ³	V	être au toucher		m.r. : velours s. : au toucher
29000	steeds	ADV	toujours		s. : instant
27996	plaats	N	place		
27468	zijn	N	le sien		
27079	velen	V	supporter		m.r. : avaler r. : supporter
27078	stellen	V	mettre		r. : installer
26762	leven	V	vivre	/l/-/n/, /v/-/b/	m.r. : anabiose s.t. : vie
26719	huis	N	maison	/s/-/t/	r. : hutte
26001	vallen	V	tomber	/v/-/f/	m.r. : s'affaler s. : tomber
25447	anders	A	autre	/d/-/t/	
25272	moeder	N	mère	/d/-/t/	r. : mater
25142	spreken	V	parler		m.r. : prêcher s.t. : parole
25099	woord	N	mot		r. : verbe
24408	vader	N	père	/v/-/b/	m.r. : baderne s.t. : homme
24275	terug	ADV	en arrière	/t/-/d/, /x/-/k/	s. : direction
24197	werk	N	ouvrage	/k/-/ʒ/	
24119	lijken	V	ressembler	/k/-/s/	m.r. : classe s.t. : commun

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
23699	heel	A	entier		m.r. : holo- s. : entier
23125	ver	A	lointain		m.r. : vergence s.t. : distance
23062	hoofd	N	chef		
22563	toen	ADV	autrefois		r. : d'antan
22157	zeker	A	en sécurité		
21962	bestaan	V	être possible	/b/-/p/, /n/-/l/	
21770	vertellen ¹	V	dire	/v/-/f/	r. : formuler
21720	volgen	V	suivre	/v/-/f/	r. : filer
21621	enkel	A	simple	/k/-/s/	
21514	kennen	V	connaître		
21449	klein	A	petit	/l/-/n/	r. : nain
21418	natuurlijk	A	naturel		
21379	gezicht	N	vue	/x/-/k/	m.r. : consulter s.t. : regarder
21106	zeer	ADV	très	/r/-/l/	r. : salement
20945	waarom	ADV	pourquoi	/r/-/l/	m.r. : volonté s.t. : intention
20811	paar	N	paire		
20472	grote	N	grosse commission	/x/-/k/	s. : excrétion
20423	oud	A	âgé de		m.r. : état s.t. : âge
20154	vraag	N	question	/v/-/f/	r. : affaire
20145	trekken ¹	V	ressembler	/k/-/s/, /n/-/l/	
20145	trekken ²	V	toucher	/k/-/ʃ/	
20145	trekken ³	V	tracer	/k/-/s/	
19643	leven	N	vie	/l/-/n/, /v/-/b/	m.r. : anabiose s. : vie
19260	soms	ADV	quelquefois	/s/-/k/	
19256	wereld	N	monde	/w/-/g/, /r/-/l/	r. : globe
19220	keer	N	fois	/k/-/ʒ/	r. : jour
19185	weg	ADV	parti	/x/-/k/	m.r. : vacant s.t. : absent
18351	noemen	V	nommer		
18171	belangrijk	A	important	/b/-/v/	m.r. : valeur s. : important
18148	nodig	A	nécessaire	/d/-/t/	s. : nature
18068	leggen	V	coucher		s. : allonger
18025	uur	N	heure		
17958	schrijven	V	écrire	/x/-/g/, /v/-/f/	m.r. : -graphe

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
					s. : écrire
17954	zaak	N	sujet	/k/-/ʒ/	
17880	bijna	ADV	presque	/b/-/p	m.r. : à peine s. : presque
17876	land	N	terre		m.r. : lande s. : terre
17857	net	ADV	net		
17811	proberen	V	essayer	/b/-/v/	r. : éprouver
17806	erg	A	mauvais		m.r. : ringard s. : mauvais
17790	naam	N	nom		
17748	halen ²	V	respirer avec difficulté		r. : haleter
17652	zetten	V	installer		
17080	werken	V	travailler		r. : œuvrer
17038	vol	A	plein	/v/-/f/	m.r. : full s. : plein
16928	juist	A	juste		
16548	sterk	A	solide	/r/-/l/	
16457	deel ¹	N	partie	/d/-/t/	m.r. : article s.t. : élément
16457	deel ²	N	plancher		m.r. : dalle s.t. : sol
16404	boek	N	ouvrage	/b/-/v/, /k/-/ʒ/	
16312	snel	A	rapide		s. : coule
16115	spelen	V	jouer	/p/-/b/	m.r. : ballon s. : jouer
16009	aantal	N	nombre		s. : ensemble
16006	gebeuren	V	arriver	/x/-/k/, /b/-/p/	r. : s'accomplir
15927	deur	N	porte	/d/-/t/	m.r. : trou s.t. : ouverture
15877	manier	N	manière		
15474	kamer	N	chambre		
15443	water	N	eau		r. : aquatique
15251	jongen	N	garçon		m.r. : gonze s.t. : homme
15121	meisje	N	filles		m.r. : mioche s.t. : enfant
15111	roepen	V	crier		m.r. : rappel s.t. : cri
15065	mooi	A	beau		m.r. : mieux s. : beau
15011	feit	N	fait		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
14944	grond	N	terre		m.r. : croûte s. : terre
14839	wachten ¹	V	attendre	/w/-/g/	r. : guetter
14839	wachten ²	V	prendre garde	/w/-/g/, /t/-/d/	
14811	zin	N	sens		
14756	kort	A	court		
14686	leren	V	apprendre		m.r. : lire s.t. : s'instruire
14570	praten	V	parler	/t/-/d/	r. : plaider
14501	eigenlijk	A	vrai	/x/-/k/	m.r. : actuel s.t. : qui existe
14450	open	A	ouvert		m.r. : épanoui s. : ouvert
14392	probleem	N	problème		
14109	vorm	N	forme	/v/-/f/	
14084	vast	A	fixe	/v/-/f/, /t/-/d/	m.r. : fidèle s.t. : ne change pas
14050	soort	N	sorte		
14028	zeggen	N	dire		m.r. : signifier s.t. : déclarer
13945	zoeken	V	chercher	/z/-/ʃ/, /k/-/ʃ/	
13752	slaan ¹	V	cogner sur		
13752	slaan ²	V	sonner		
13701	stad	N	ville		r. : cité
13676	groep	N	groupe		
13261	jong	A	jeune		
13177	vroeg	A	précoce	/x/-/k/	
13050	samen	ADV	ensemble	/n/-/l/	
12992	stem	N	voix	/s/-/ʃ/	m.r. : chanter s. : voix
12989	wit	A	blanc	/t/-/d/	m.r. : vide r. : blanc
12834	lezen	V	lire		r. : lecteur
12773	lachen	V	rire	/x/-/g/	m.r. : allégresse s.t. : gaieté
12733	bed	N	lit	/b/-/p/	r. : paddock
12651	verder	A	qui reste de	/d/-/t/	m.r. : vestige s. : qui reste de
12617	moment	N	moment		
12498	daarna	ADV	après		r. : derrière
12353	kant	N	côté		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
12305	staat	N	état		
12208	wat	ADV	un peu	/w/-/g/	m.r. : goutte s. : peu
12122	moeilijk	A	difficile		r. : malaisé
12059	weinig	ADV	peu	/w/-/g/	r. : grain
12053	vriend	N	ami		m.r. : frangin s. : ami
12052	avond	N	soir	/v/-/f/	s. : fin
12005	week	N	semaine		r. : week-end
11963	dragen ¹	V	porter		s. : diriger
11963	dragen ²	V	porter	/d/-/t/, /x/-/ʃ/	s. : toucher
11880	geld	N	argent		r. : galette
11735	gedachte	N	idée	/x/-/k/, /d/-/t/	m.r. : critique s.t. : intellectuel
11693	licht	N	lumière		r. : clarté
11667	diep	A	profond	/d/-/t/, /p/-/b/	m.r. : tomber s. : profond
11573	vormen	V	former	/v/-/f/	
11549	wijze ¹	N	façon	/w/-/g/	r. : guise
11429	dienen ¹	V	servir	/d/-/t/, /n/-/l/	s. : utile
11429	dienen ²	V	servir		r. : donner
11360	bekend	A	connu	/b/-/p/, /k/-/s/	m.r. : pressentiment s. : connu
11307	bijvoorbeeld	ADV	par exemple	/b/-/p/	m.r. : approuver r. : confirmer
11257	nacht	N	nuit		
11215	graag	ADV	volontiers	/x/-/k/	s. : fréquent
11163	belang	N	intérêt	/b/-/v/	m.r. : valeur s. : intérêt
10955	leiden	V	mener		m.r. : leader s.t. : être en tête
10944	ontstaan	V	naître	/t/-/d/	m.r. : se distinguer r. : apparaître
10887	ontwikkeling	N	développement	/k/-/ʒ/, /l/-/n/	r. : ontogenèse
10796	vrij	A	libre		r. : franc
10796	zit	N	session	/t/-/s/	
10772	steken ¹	V	piquer	/t/-/d/	m.r. : déchirer r. : percer

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
10772	steken ²	V	être coincé		
10625	gevoel	N	sentiment	/x/-/k/	m.r. : équivoque s.t. : jugement
10596	verdwijnen	V	disparaître	/r/-/l/, /d/-/t/, /n/-/l/	r. : se volatiliser
10588	half	A	demi		m.r. : half-pipe s. : demi
10582	wijzen ³	V	se montrer		m.r. : visible s.t. : être vu
10483	tafel	N	table	/t/-/s/	s. : surface
10356	algemeen	A	général	/x/-/k/	r. : commun
10356	politiek	A	politique		
10292	school	N	école		
10251	gebied	N	domaine	/x/-/k/, /b/-/p/	r. : compétence
10205	bijzonder	A	spécial	/b/-/p/, /z/-/s/, /n/-/l/	
10180	buiten	ADV	dehors	/b/-/v/, /t/-/s/	m.r. : éversion s.t. : extérieur
10111	bereiken	V	atteindre	/b/-/p/	m.r. : approcher s. : atteindre
10107	verhaal	N	histoire	/v/-/f/	r. : affaire
9908	dood	N	mort	/t/-/s/	r. : décès
9850	draaien ¹	V	tourner	/d/-/t/	
9850	draaien ²	V	tourner	/d/-/t/	
9837	sociaal	A	social		
9829	recht	N	droit		
9812	zwaar	A	sévère		
9770	schijnen	V	sembler	/n/-/l/	
9760	hard	A	dur		r. : acharné
9753	lid	N	paupière	/l/-/n/	m.r. : nictation s. : paupière
9747	gevolg	N	conséquence	/x/-/k/, /l/-/n/	m.r. : inconvé- nient s. : conséquence
9735	wijn	N	vin		
9727	maand	N	mois		r. : mensuel
9719	voorbeeld	N	exemple	/v/-/b/, /b/-/v/	m.r. : bréviaire s.t. : enseignement
9711	eeuw	N	siècle		m.r. : avant s.t. : période

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
9705	precies	A	précis		
9677	situatie	N	situation		
9653	mond	N	bouche		s. : mange
9649	opnieuw	ADV	de nouveau		
9636	vroeger	A	antérieur		m.r. : régression s. : antérieur
9600	stuk	N	article		
9579	reden	N	raison	/d/-/t/	m.r. : prétexte s. : raison
9520	voet	N	pied	/v/-/b/	m.r. : botte s. : pied
9453	thuis	ADV	à la maison	/s/-/t/	m.r. : château s. : maison
9419	raken ¹	V	toucher		s. : rencontrer
9419	raken ²	V	devenir	/k/-/ʒ/	m.r. : ériger r. : changer
9381	merken	V	marquer		
9079	ouder	N	parent	/d/-/t/	m.r. : utérin s.t. : père
9050	kracht	N	force		m.r. : -crate s. : force
9026	wonen	V	habiter		m.r. : avenue s.t. : habitation
8987	gebruik	N	usage	/x/-/k/, /b/-/v/	s. : activité
8948	meteen	ADV	tout de suite	/t/-/d/	m.r. : immédiat s. : tout de suite
8926	idee	N	idée		
8892	rol ¹	N	rôle		
8867	ooit	ADV	jamais		s. : temps
8843	antwoord	N	réponse		m.r. : interroger s. : réponse
8798	auto	N	auto		
8748	antwoorden	V	répondre		m.r. : interroger s.t. : réponse
8739	slapen	V	dormir		r. : s'assoupir
8733	dicht	A	dense	/t/-/s/	
8704	kerk	N	église	/k/-/g/, /r/-/l/	
8699	enig	A	unique	/x/-/k/	
8688	meneer	N	monsieur		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
8641	plotseling	A	soudain	/s/-/ʃ/, /l/-/n/	m.r. : prochainement s.t. : peu de temps
8640	onderzoek	N	analyse	/r/-/l/	
8591	rijden ¹	V	rouler	/d/-/t/	m.r. : brouette s.t. : roues
8591	rijden ²	V	transporter		m.r. : radeau s.t. : porter
8562	kans	N	chance		
8557	waarschijnlijk	A	vraisemblable		
8517	plan	N	plan		
8511	ogenblik	N	moment	/x/-/k/, /k/-/ʒ/	m.r. : chronologie s.t. : temps
8494	brief	N	lettre	/b/-/p/, /r/-/l/	r. : pli
8446	aandacht	N	attention	/d/-/t/, /t/-/s/	
8443	langzaam	A	lent		r. : long
8422	richting	N	orientation		
8411	gelden	V	compter	/x/-/k/, /d/-/t/	
8375	geest	N	esprit	/x/-/k/	r. : conscience
8362	beeld	N	image	/b/-/p/	m.r. : peinture s.t. : représentation
8351	sluiten ¹	V	clôturer		
8351	sluiten ²	V	se solder	/t/-/d/	
8289	richten ¹	V	orienter		
8289	richten ²	V	placer	/r/-/l/, /t/-/s/	
8289	richten ³	V	s'adresser	/t/-/s/	
8285	oorlog	N	guerre	/x/-/ʃ/	m.r. : clash s.t. : conflit
8283	eten ¹	V	manger	/t/-/d/	r. : dîner
8283	eten ²	V	dîner	/t/-/d/	
8283	eten ³	V	manger à	/t/-/d/	r. : dîner
8275	zacht	A	doux		s. : sucré
8245	meestal	ADV	généralement		m.r. : masse s.t. : majorité
8245	straat	N	rue		m.r. : trottoir s. : rue
8243	persoon	N	personne		
8238	nauwelijks	ADV	à peine		m.r. : nouvellement s.t. : depuis peu

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
					de temps
8186	heer	N	homme		m.r. : hercule s. : homme
8149	macht	N	pouvoir	/t/-/d/	r. : mandat
8113	schouder	N	épaule		m.r. : coude s.t. : bras
8091	klinken ¹	V	sonner	/k/-/s/	
8091	klinken ²	V	enchaîner		
8090	luisteren	V	écouter	/l/-/n/, /t/-/d/, /r/-/l/	m.r. : inaudible s.t. : entendre
8077	menselijk	A	humain	/s/-/t/	r. : mortel
8070	invloed	N	influence		
8059	hart	N	cœur		m.r. : hardiesse r. : cœur
7982	blik ¹	N	regard		r. : loucher
7977	slecht	A	mauvais		m.r. : blèche s. : mauvais
7954	bezig	A	occupé		m.r. : besogne s.t. : travail
7943	arm	N	bras		m.r. : armé s. : bras
7942	wet	N	loi		m.r. : vertu s.t. : règle
7936	los	A	lâche	/s/-/ʃ/	
7921	bang	A	peureux		m.r. : ombrageux r. : peureux
7920	gang ¹	N	corridor	/x/-/k/	m.r. : conduit s.t. : passage
7920	gang ²	N	gang		
7918	verband	N	bandage	/r/-/l/	m.r. : volant s.t. : bande de tissu
7772	dier	N	animal	/d/-/t/, /r/-/l/	m.r. : éthologie s. : animal
7764	beschouwen	V	considérer		r. : observer
7735	lucht	N	air	/t/-/d/	s. : fluide
7689	bieden	V	offrir	/b/-/p/, /d/-/t/	s. : portée
7640	rug	N	dos	/x/-/ʃ/	m.r. : rachis s.t. : colonne vertébrale
7609	waar	A	vrai		
7608	zorgen	V	se charger de	/z/-/ʃ/	
7584	inderdaad	ADV	en effet	/d/-/t/	m.r. : attester

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
					r. : confirmer
7555	kiezen ¹	V	choisir		
7555	kiezen ²	V	choisir		
7550	been	N	jambe	/b/-/p/, /n/-/l/	r. : pilier
7490	angst	N	angoisse		
7434	laag	A	bas	/x/-/ʃ/	r. : lâche
7403	donker	A	ténébreux	/d/-/t/	
7398	verschil	N	différence		r. : variation
7385	raam	N	fenêtre	/r/-/l/	s. : lumière
7359	minuut	N	minute		
7291	punt ¹	N	point		
7291	punt ²	N	problème		m.r. : point s.t. : question
7291	punt ³	N	pointe		
7224	partij	N	partie		
7199	liefde	N	amour	/v/-/b/	m.r. : libido s.t. : sexuel
7196	ervaring	N	expérience		s. : éprouver
7188	lijden ¹	V	souffrir		m.r. : lourdeur s.t. : supporter
7188	lijden ²	V	souffrir	/d/-/t/	m.r. : plainte s.t. : douleur
7169	Duits	A	allemand	/d/-/t/	r. : teuton
7159	pakken ¹	V	passionner	/k/-/s/	
7159	pakken ²	V	prendre	/k/-/s/	s. : épaissir
7155	direct	A	direct		
7105	hond	N	chien		
7087	voeren ¹	V	mener	/v/-/b/, /r/-/l/	m.r. : balloter s.t. : faire aller
7087	voeren ²	V	fouerrer	/v/-/f/	
7076	relatie	N	relation		
7040	begrip	N	compréhension	/b/-/p/, /x/-/k/	m.r. : opaque s.t. : comprendre
7029	mevrouw	N	madame		s. : mariée
7020	volgend	A	suisant	/l/-/n/	m.r. : venir r. : suisant
7011	systeem	N	système		
6995	stil	A	silencieux		
6986	knikken ¹	V	se casser	/k/-/s/	
6986	knikken ²	V	fléchir	/n/-/l/, /k/-/ʃ/	
6983	doel	N	but	/d/-/t/	m.r. : télé-

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
					s. : but
6968	dienst	N	service	/t/-/d/	m.r. : décidé s.t. : décision
6942	kopen	V	acheter		m.r. : accaparer s. : acheter
6929	proces	N	procès		
6878	onderwijs	N	enseignement	/d/-/t/	r. : instruction
6846	zwart	A	noir	/r/-/l/	r. : sale
6810	rustig	A	tranquille	/s/-/k/	
6784	drinken	V	boire		m.r. : trinquer s. : boire
6760	contact	N	contact		
6753	orde	N	ordre		
6715	functie	N	fonction		
6689	dik	A	épais	/k/-/s/	s. : dense
6689	ongeveer	ADV	environ		
6687	waarde	N	valeur	/r/-/l/	
6680	warm	A	chaud		m.r. : vermeil s. : chaud
6675	paard	N	cheval	/p/-/b/	r. : bardot
6666	vinger	N	doigt	/r/-/n/	m.r. : évagination s. : doigt
6646	vreemd	A	étranger	/v/-/f/	s. : familial
6642	schieten ¹	V	tirer		s. : shooter
6642	schieten ²	V	tirer		s. : shooter
6642	schieten ³	V	shooter		
6625	taal	N	langue	/t/-/d/	r. : dialecte
6600	ontdekken	V	découvrir		r. : déceler
6599	einde	N	fin	/d/-/t/	s. : instant
6585	gesprek	N	conversation	/x/-/k/	m.r. : causerie r. : conversation
6583	mate	N	mesure		r. : métrie
6574	volk	N	foule	/v/-/f/	
6570	kleur	N	couleur		
6556	indruk	N	impression	/d/-/t/, /k/-/s/	r. : intuition
6555	drukken ¹	V	appuyer	/k/-/s/	r. : adosser
6555	drukken ²	V	presser	/k/-/s/	
6524	sterven	V	mourir		r. : crever
6505	pijn	N	peine		
6497	aarde	N	terre		m.r. : ados

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
					s. : terre
6487	glas	N	verre		r. : glace
6459	artikel	N	article		
6447	onmiddellijk	A	immédiat	/d/-/t/, /l/-/n/	m.r. : maintenant s.t. : moment présent
6431	dood	A	mort	/t/-/s/	r. : décédé
6389	persoonlijk	A	personnel		
6389	stoel	N	chaise		m.r. : selle s.t. : siège
6355	ruimte	N	place	/r/-/l/, /t/-/s/	
6336	over	A	fini	/v/-/f/, /r/-/n/	
6327	taak	N	tâche		
6291	moeite	N	peine	/t/-/d/	m.r. : merder s.t. : difficulté
6285	rood	A	rouge	/t/-/s/	m.r. : roussir s.t. : roux
6250	verandering	N	changement	/d/-/t/, /r/-/l/	r. : versatilité
6229	muur	N	mur		
6194	passen ¹	V	passer		
6194	passen ²	V	mesurer		r. : apprécier
6191	geluid	N	son	/x/-/k/, /l/-/n/	r. : canard
6147	duren	V	durer		
6140	dokter	N	docteur		
6095	betrekking	N	emploi	/b/-/p/	m.r. : pratique s.t. : activité
6066	bestaan	N	existence	/b/-/p/	r. : présence
6058	gedrag	N	conduite	/x/-/k/	
6056	zee	N	mer	/z/-/s/	r. : océan
6003	bewust	A	conscient	/b/-/p/	s. : éprouve
5978	kennis ¹	N	connaissance		
5978	kennis ²	N	connaissance		
5951	raad	N	conseil	/r/-/l/, /t/-/s/	m.r. : leçon s. : conseil
5880	regel	N	règle		
5841	groeien	V	croître	/x/-/k/	
5832	min	A	moins		
5818	boom	N	arbre	/b/-/p/	m.r. pommier s. : arbre
5818	dorp	N	village	/d/-/t/	m.r. : terroir s.t. : rural

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
5809	gelijk	A	égal		
5806	geschiedenis	N	histoire	/x/-/k/	m.r. : coïncidence s.t. : événements
5792	economisch	A	économique		
5789	koud	A	froid	/k/-/g/, /t/-/s/	r. : glacé
5735	kop	N	caboche	/p/-/b/	
5721	stand ¹	N	attitude		
5721	stand ²	N	stand		
5696	ernstig	A	sérieux		r. : critique
5695	familie	N	famille		
5674	periode	N	période		
5646	erop	ADV	dessus	/r/-/l/, /p/-/f/	m.r. : plafond s.t. : supérieur
5622	aard	N	nature	/r/-/n/	
5593	resultaat	N	résultat		
5589	keren ¹	V	tourner		s. : courbe
5589	keren ²	V	retourner		m.r. : charnière s.t. : tourner
5589	keren ³	V	se tourner	/k/-/s/	m.r. : suivre s.t. : aller dans une direction
5578	normaal	A	normal		
5565	tonen ¹	V	montrer	/n/-/l/	r. : étaler
5565	tonen ²	V	donner une impression	/t/-/s/, /n/-/l/	r. : sembler
5565	tonen ³	V	se montrer	/t/-/d/	m.r. : se donner s. : se montrer
5558	breed	A	large	/b/-/v/	r. : vaste
5554	rond	A	rond		
5521	opzicht	N	égard		m.r. : porter s.t. : concerner
5483	organisatie	N	organisation		
5456	speciaal	A	spécial		
5455	bezitten	V	posséder	/b/-/p/, /t/-/d/	
5439	patiënt	N	patient		
5425	broer	N	frère		
5421	opeens	ADV	tout à coup	/s/-/k/	m.r. : piquer s.t. : brusquement
5418	opvatting	N	idée	/t/-/s/	r. : aperçu
5412	leveren	V	livrer		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
5408	daarvoor	ADV	devant		
5398	totaal	A	total		
5380	overheid	N	autorité	/v/-/f/	r. : force
5358	afstand	N	distance	/f/-/p/	r. : espace
5358	eind	N	bout		m.r. : entame r. : bout
5336	durven	V	oser	/r/-/l/	m.r. : déluré s.t. : hardiesse
5298	vrijheid	N	liberté		m.r. : free-jazz s. : liberté
5290	leeg	A	inoccupé	/l/-/n/, /x/-/k/	
5281	blauw	A	bleu		
5270	klaar	A	clair		
5242	stap	N	pas		r. : étape
5241	schudden ¹	V	secouer	/d/-/t/	r. : cahoter
5241	schudden ²	V	être secoué	/d/-/t/	m.r. : cahoter s. : être secoué
5215	karakter	N	caractère		
5191	Amerikaans	A	américain		
5174	ziekte	N	maladie		m.r. : séquelle s. : maladie
5164	dringen ²	V	pousser		s. : entraîner
5164	hoogte	N	hauteur		
5159	theorie	N	théorie		
5142	bloed	N	sang		m.r. : boudin s. : sang
5139	Frans	A	français		
5138	beleid	N	politique	/b/-/p/	
5135	strijd	N	querelle	/s/-/k/	
5132	gauw	A	rapide		m.r. : aigu s. : rapide
5114	zorg	N	soin		r. : sérieux
5106	besluiten ¹	V	terminer	/s/-/t/, /t/-/s/	m.r. : aboutissement r. : terme
5106	besluiten ²	V	se décider	/b/-/v/	m.r. : velléitaire s. : se décider
5094	beter	A	meilleur		s. : bonté
5082	licht	A	léger		r. : leste
5081	dochter	N	filie		m.r. : adopter s. : filie
5072	stappen	V	marcher		m.r. : clopiner

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
					s. : marcher
5062	droom	N	rêve		m.r. : dormir s.t. : sommeil
5058	schrijver	N	écrivain		
5050	tuin	N	jardin		r. : tonnelle
5035	bedrijf	N	profession	/b/-/p/	
5033	meter ¹	N	mètre		
5033	meter ²	N	marraine		m.r. : matrone s.t. : femme
5030	springen	V	sauter		s. : projeté
5005	onder	ADV	en bas		m.r. : indice s. : en bas
4994	vandaag	ADV	aujourd'hui		m.r. : vendredi s.t. : jour
4952	krant	N	quotidien		
4921	verhouding	N	rapport	/v/-/f/, /r/-/n/, /d/-/t/	r. : affinité
4915	centraal	A	central		
4899	trap	N	escalier		m.r. : rampe s. : escalier
4889	rest	N	reste		
4887	regering	N	gouvernement		m.r. : régner s.t. : pouvoir
4886	muziek	N	musique		
4882	werpen	V	jeter	/r/-/l/	m.r. : volée s. : jeter
4875	scherp	A	coupant		
4873	treden ¹	V	aller		r. : se rendre
4873	treden ²	V	fouler		m.r. : étreindre r. : presser
4849	openen ¹	V	s'ouvrir		r. : s'épanouir
4849	openen ²	V	ouvrir		m.r. : épanoui s.t. : ouvert
4838	gek	A	fou	/k/-/s/	r. : agacer
4834	bank	N	banc		
4830	blad	N	plateau		
4816	schip	N	bateau		m.r. : skif s. : bateau
4815	rekening	N	note	/k/-/g/, /n/-/l/	m.r. : régler s. : note
4813	voren	ADV	avant	/v/-/b/	r. : d'abord
4802	zak	N	sac		
4798	borst ¹	N	poitrine		r. : buste

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
4789	soldaat	N	soldat		
4786	papier	N	papier		
4762	breken ¹	V	casser		r. : ébrécher
4762	breken ²	V	percer	/b/-/p/	
4759	fles	N	flacon		
4752	waarheid	N	vérité		
4750	tenminste	ADV	du moins	/t/-/d/	
4746	prachtig	A	splendide	/r/-/l/, /t/-/d/	
4723	grens	N	frontière		
4715	maal ¹	N	fois		r. : multiple
4715	maal ²	N	portion	/l/-/n/	m.r. : menu s.t. : mets
4708	beroep	N	profession	/b/-/p/, /p/-/f/	
4706	minister	N	ministre		
4702	wind	N	vent		
4699	koffie	N	café		
4698	toestand	N	situation	/t/-/s/	
4691	stoppen ¹	V	stopper		
4691	stoppen ²	V	s'arrêter		r. : stopper
4691	stoppen ³	V	repriser		r. : stopper
4675	oom	N	oncle		m.r. : homme s.t. : mari
4671	oor	N	oreille		m.r. : étrier s. : oreille
4671	poging	N	tentative	/p/-/b/	m.r. : bégalement s. : tentative
4668	oorzaak	N	raison		
4626	vlak	A	lisse	/k/-/s/	
4620	buurt	N	environs	/b/-/v/	
4618	informatie	N	information		
4614	mening	N	opinion		m.r. : méninge s.t. : esprit
4605	ruim	A	étendu	/r/-/l/	m.r. : limité s. : étendu
4598	methode	N	méthode		
4594	gooien	V	jeter	/x/-/k/	m.r. : canon s.t. : lancer
4594	lip	N	lèvre		r. : lippe
4594	positie	N	position		
4582	gezin	N	famille	/x/-/k/	m.r. : classe s.t. : ensemble

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
					de personnes
4581	morgen	N	matin		m.r. : amorce r. : commencement
4580	huilen	V	hurler		
4579	algemeen	N	ensemble	/x/-/k/	r. : commun
4574	voorzichtig	A	prudent	/t/-/s/	m.r. : avertissement s.t. : prudence
4557	enorm	A	énorme		
4522	foto	N	photo		
4507	film	N	film		
4500	activiteit	N	activité		
4494	modern	A	moderne		
4477	trachten	V	tâcher		
4464	geluk	N	bonheur		m.r. : gloire r. : succès
4458	beslissing	N	décision	/b/-/v/, /s/-/t/	r. : volonté
4442	basis	N	base		
4438	niveau	N	niveau		
4438	steen ¹	N	pierre	/n/-/l/	s. : solide
4438	steen ²	N	dalle	/t/-/d/, /n/-/l/	
4438	steen ³	N	forteresse	/s/-/ʃ/	m.r. : château r. : forteresse
4429	noodzakelijk	A	indispensable	/k/-/s/	
4423	zuster	N	sœur	/t/-/d/	m.r. : ascendance s.t. : parents
4419	vertonen ¹	V	présenter	/t/-/d/	m.r. : en évidence r. : montrer
4419	vertonen ²	V	se montrer	/t/-/d/	m.r. : en évidence s.t. : être vu
4416	lijn	N	ligne		
4406	tante	N	tante		
4401	schuiven ¹	V	pousser	/v/-/f/	m.r. : souffler s. : pousser
4391	blij	A	heureux		r. : bon
4387	vuur	N	feu	/v/-/f/	r. : foudre
4386	prijs	N	prix		
4379	eten	N	nourriture	/t/-/d/	m.r. : dîner

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
					s.t. : repas
4377	aardig	A	gentil	/r/-/l/	m.r. : élégant s.t. : grâce
4364	terecht	A	juste	/t/-/d/, /r/-/l/	m.r. : délicatesse s. : juste
4352	leeftijd	N	âge		m.r. : longtemps s.t. : temps
4350	kloppen ¹	V	frapper		s. : coups
4350	kloppen ²	V	battre		s. : coups
4344	reactie	N	réaction		
4340	telkens	ADV	à tous les coups		
4326	verkopen ²	V	vendre	/v/-/b/, /r/-/n/, /p/-/f/	m.r. : bénéfice s.t. : vente
4322	eveneens	ADV	également		m.r. : à l'avenant r. : de même
4319	tegelijk	ADV	en même temps	/t/-/s/, /x/-/k/, /l/-/n/	m.r. : synchronisé s. : en même temps
4318	praktijk	N	pratique		
4297	neus	N	nez		
4292	bellen	V	sonner	/b/-/p/	m.r. : appeler r. : sonner
4284	sturen ¹	V	adresser	/t/-/d/	
4284	sturen ²	V	conduire	/t/-/d/	
4283	stilte	N	silence		
4279	daarin	ADV	dedans	/d/-/t/	s. : intérieur
4261	inzicht	N	intention		
4259	terrein	N	terrain		
4249	koning	N	roi		m.r. : khan s.t. : chef
4247	model	N	modèle		
4245	boer	N	plouc	/b/-/p/, /r/-/l/	
4234	zingen	V	chanter	/z/-/ʃ/	
4232	vliegen ¹	V	voler		r. : voltiger
4223	gat	N	trou	/x/-/ʃ/	r. : chatière
4200	cultuur	N	culture		
4200	kunst	N	art	/n/-/l/	r. : sculpture
4197	reageren	V	réagir		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
4194	genieten ²	V	jouir	/x/-/k/	m.r. : content s.t. : plaisir
4188	kapitein	N	capitaine		
4186	vorig	A	dernier	/v/-/f/, /r/-/n/	r. : final
4163	bezoek	N	visite	/b/-/v/	
4155	grijpen	V	attraper		
4150	reis	N	voyage		r. : croisière
4136	zwarte	N	Noir	/t/-/d/	m.r. : vaudou s. : Noir
4131	spel	N	jeu		r. : sport
4125	druk	N	pression	/k/-/s/	
4107	oplossing	N	solution		m.r. : élucider s.t. : difficulté
4102	haar ¹	N	cheveu		m.r. : hirsute s. : cheveu
4102	haar ²	N	poils		m.r. : haire s. : poils
4093	achten	V	se soucier de		r. : s'inquiéter
4093	missen	V	manquer	/s/-/k/	
4088	element	N	élément		
4059	vlug	A	rapide		r. : véloce
4055	vogel	N	oiseau	/x/-/k/	r. : aviculteur
4051	schuld	N	dette		r. : solde
4048	aanvaarden	V	accepter	/r/-/l/	m.r. : avaler s. : accepter
4045	stof ¹	N	poussière	/f/-/p/	s. : suspension
4045	stof ²	N	étoffe		
4040	keuken	N	cuisine		r. : kitchenette
4038	politie ¹	N	police		
4038	politie ²	N	agent de police		m.r. : policier s. : agent de police
4037	Parijs	A	parisien		
4026	leuk	A	drôle	/k/-/s/	m.r. : glousser s.t. : rire
4017	onderzoeken	V	examiner	/z/-/s/	r. : disséquer
4017	wensen	V	vouloir	/n/-/l/	
4006	trouwen ¹	V	se marier	/t/-/d/	m.r. : divorce s.t. : mariage
4006	trouwen ²	V	épouser	/t/-/d/	m.r. : divorce s.t. : mariage
3999	bloem	N	fleur	/b/-/p/	s. : plante
3987	voort	ADV	en avant		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
3971	bouwen ¹	V	bâtir	/b/-/v/, /w/-/g/	m.r. : wagonnet s.t. : matériaux
3964	goed	N	crédit	/x/-/k/	
3954	begin	N	début		m.r. : bégaïement r. : commencement
3951	factor	N	facteur		
3951	succes	N	succès		
3941	kost	N	coût		
3940	natuur	N	nature		
3930	arts	N	docteur	/t/-/d/	
3913	hotel	N	hôtel		
3912	verstaan ²	V	s'entendre	/v/-/f/	r. : fraterniser
3897	gesteld	A	attaché		s. : goût
3890	fijn	A	fin		
3883	studie	N	étude		
3882	geheel	N	tout	/x/-/k/	m.r. : cohorte s.t. : ensemble
3879	rekenen ¹	V	calculer		m.r. : reconnaître s.t. : déterminer
3879	rekenen ²	V	compter		r. : recenser
3875	vriendelijk	A	aimable	/v/-/b/	m.r. : bienveillant r. : aimable
3869	aspect	N	aspect		
3866	voegen ¹	V	joindre	/v/-/b/	m.r. : bague s. : joindre
3857	teken	N	signe	/t/-/d/	s. : indice
3851	figuur	N	figure		
3841	nationaal	A	national		
3840	treffen	V	frapper	/f/-/p/	
3835	toon	N	ton		
3833	plek	N	place		
3821	peter	N	parrain	/p/-/b/	s. : baptême
3791	morgen	ADV	demain		m.r. : mardi s.t. : jour
3783	ziek	A	malade	/z/-/ʃ/	m.r. : check-up s.t. : santé

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
3780	blijkbaar	A	manifeste		r. : flagrant
3779	dame	N	dame		
3774	stukje	N	article		
3752	tand	N	dent	/t/-/d/	
3745	schenken	V	concéder		
3742	kaart	N	carte		
3740	term	N	terme		
3738	bureau	N	bureau		
3722	opdracht	N	ordre		
3721	beseffen	V	se rendre compte	/b/-/p/, /f/-/v/	r. : s'apercevoir
3707	rennen	V	courir	/r/-/l/	s. : élans
3690	Engels	A	anglais		
3685	effect	N	effet		
3685	rand	N	rand		
3684	gebouw	N	bâtiment	/x/-/k/	m.r. : cabane s. : bâtiment
3660	althans	ADV	en tout cas		
3660	huid	N	peau	/t/-/s/	m.r. : housse s.t. : enveloppe
3652	plant	N	plante		
3650	uitvoeren	V	exporter	/f/-/p/	
3638	rust	N	repos		r. : arrêt
3633	ervaren	V	éprouver		
3627	uitdrukking	N	expression	/k/-/s/	m.r. : adresse s. : expression
3622	schrikken	V	sursauter	/k/-/s/	
3590	midden	ADV	au milieu		m.r. : médian s. : au milieu
3587	tekst	N	texte		
3573	beweren	V	prétendre	/b/-/p/	m.r. : prouver s.t. : certain
3573	knie	N	genou	/k/-/ʒ/	
3567	telefoon	N	téléphone		
3566	kosten	V	coûter		
3558	bos ¹	N	botte	/s/-/t/	
3558	bos ²	N	bois		
3557	optreden	V	apparaître	/d/-/t/	
3540	hoop ¹	N	tas		m.r. : charpie s.t. : amas
3539	gering	A	petit	/r/-/n/	r. : jeune
3534	individueel	A	individuel		

Fréquence	Lemme	Catégorie grammaticale	Équivalent français	Manipulations	
				Ajustement formel	Ajustement sémantique
3532	structuur	N	structure		
3529	roman	N	roman		
3528	slagen	V	réussir	/x/-/k/	s. : succès
3515	burger	N	bourgeois		
3510	kring	N	cerne		
3508	grijs	A	gris		
3498	eens	A	d'accord	/s/-/t/	r. : s'entendre
3494	slaap	N	tempe	/s/-/t/	
3484	bestuur	N	direction	/b/-/p/, /t/-/d/	m.r. : présider s.t. : diriger
3482	kwestie	N	question		
3480	geloof	N	foi		m.r. : église s. : foi
3478	eventueel	A	éventuel		
3475	wezen	N	être		m.r. : visage s.t. : personne
3467	dichter	N	poète		m.r. : dactyle s.t. : poésie
3464	keuze	N	choix		s. : choisir
3444	vlees	N	viande	/l/-/n/	r. : venaison
3439	voorwaarde	N	condition	/v/-/f/, /r/-/l/, /d/-/t/	m.r. : formalité s. : condition
3435	trein	N	train		
3435	vangen ²	V	être cinglé	/v/-/f/	m.r. : fondu s.t. : fou
3432	praktisch	A	pratique		
3425	helder	A	clair		m.r. : halo s.t. : éclat
3421	uitspraak	N	prononciation	/r/-/l/	s. : articuler
3415	nergens	ADV	nulle part	/r/-/l/	
3404	ochtend	N	matin		m.r. : création r. : début
3390	rode	N	roux		m.r. : ardent r. : roux
3388	zijde	N	côté		m.r. : side-car s. : côté
3383	winnen	V	gagner	/w/-/g/	
3377	voorstel	N	proposition		m.r. : ouvertures r. : proposition

Il nous a semblé intéressant de présenter les listes des mots opaques de chaque langue par rapport au français, et ce pour deux raisons : d'abord pour que la totalité du corpus utilisé apparaisse dans notre travail, et ensuite pour laisser la possibilité d'examiner ces mots qui n'ont pas été considérés comme transparents. « Mots opaques » s'entend par opposition à « mots transparents » ; il s'agit donc ici des mots qui ne sont pas les représentants d'hyperlexèmes.

Tableau des mots anglais opaques

Fréquence	Lemme	Catégorie grammaticale
687085	be	V
242066	have	V
91464	not	ADV
80717	do	V
76541	say	V
51830	go	V
51269	much	ADV
45312	out	ADV
42717	get	V
36958	see	V
35874	think	V
35871	up	ADV
30968	now	ADV
28954	only	ADV
28805	very	ADV
25802	may	V
23445	even	ADV
21871	the	ADV
20077	want	V
19742	well	ADV
18602	thing	N
18448	would	V
18344	down	ADV
17760	should	V
17657	back	ADV
16068	never	ADV
16014	day	N
15641	still	ADV

14957	become	V
14935	life	N
14882	on	ADV
14651	ask	V
14061	here	ADV
13855	yes	ADV
13783	again	ADV
13567	world	N
13126	mean	V
12146	off	ADV
9955	need	V
9866	quite	ADV
9384	eye	N
9246	room	N
9199	rather	ADV
8487	perhaps	ADV
8426	almost	ADV
8413	often	ADV
8395	show	V
8324	hold	V
8053	how	ADV
8047	ever	ADV
8022	little ¹	A
8022	little ²	A
7838	happen	V
7540	few	N
7516	however	ADV
7312	little	N
7106	why	ADV
7049	high	A
6956	help	V
6689	read	V
6555	together	ADV
6501	already	ADV
6389	lot ¹	N
6256	about	ADV
5924	bad	A
5684	enough	ADV
5660	soon	ADV
5580	might	V
5527	least	N
5226	sometimes	ADV
5220	understand	V

5183	eat	V
5086	early	A
5046	once	ADV
5036	spend	V
4875	else	ADV
4533	indeed	ADV
4483	watch	V
4463	little	ADV
4455	either	ADV
4284	death	N
4182	meet	V
4178	need	N
4118	shall	V
4080	hope	V
4030	ago	ADV
3958	today	ADV
3868	usually	ADV
3797	strong	A
3689	son	N
3601	therefore	ADV
3491	alone	A
3459	main	A
3373	wish	V
3359	yet	ADV
3254	low	A
3216	before	ADV
3207	early	ADV
3175	listen	V
3162	where	ADV
3083	wonder	V
3028	instead	ADV
2861	while	N
2644	mouth	N
2618	dead	A
2608	fail	V
2604	heavy	A
2587	income	N
2524	end	V
2482	deal ²	N
2468	available	A
2467	anyway	ADV
2465	nearly	ADV
2425	maybe	ADV

2423	help	N
2336	tend ¹	V
2275	hang	V
2258	close	A
2244	ready	A
2232	summer	N
2223	easily	ADV
2220	things	N
2217	earth	N
2207	hardly	ADV
2176	newspaper	N
2175	stare	V
2134	yeah	ADV
2123	merely	ADV
2063	thank	V
2043	through	ADV
2012	afraid	A
1949	afternoon	N
1917	bear ²	V
1896	along	ADV
1856	weapon	N
1854	close	ADV
1845	ought	V
1747	somewhere	ADV
1727	worry	V
1662	high	ADV
1648	behaviour	N
1646	highly	ADV
1611	statement	N
1602	quiet	A
1599	own	V
1594	hell	N
1588	by	ADV
1582	Sunday	N
1540	egg	N
1518	otherwise	ADV
1518	Jew	N
1503	sky	N

Tableau des mots allemands opaques

Fréquence	Lemme	Catégorie grammaticale
47745	nicht	ADV
14287	müssen	V
13998	ja	ADV
10724	wie	ADV
8678	schon	ADV
7811	ganz	A
7510	wieder	ADV
7458	jetzt	ADV
7252	hier	ADV
6413	Mensch ²	N
4902	nun	ADV
4867	nehmen	V
4644	hoch	A
4255	einmal	ADV
3952	solch	A
3766	bringen	V
3762	doch	ADV
3487	etwa	ADV
3477	Sie	N
3416	erst	ADV
3269	alle	ADV
3073	dabei	ADV
2991	meinen	V
2960	dazu	ADV
2872	eigen	A
2772	vielleicht	ADV
2706	heuen	V
2636	heute	ADV
2503	möglich	A
2444	damit	ADV
2423	bisher	ADV
2208	setzen	V
2197	besonders	ADV
2117	mal	ADV
2090	davon	ADV
2078	Möglichkeit	N
2059	dafür	ADV
2043	letzt	A

1867	wirklich	A
1858	darüber	ADV
1821	versuchen	V
1789	wo	ADV
1783	heißen ¹	V
1700	gehören	V
1685	überhaupt	ADV
1612	deshalb	ADV
1599	kaum	ADV
1580	spät	A
1558	Sache	N
1526	Gesellschaft	N
1523	bekommen	V
1518	zusammen	ADV
1486	hin	ADV
1477	Ergebnis	N
1476	sitzen	V
1461	allerdings	ADV
1454	einig	A
1445	damals	ADV
1445	sofort	ADV
1430	Ding	N
1393	westdeutsch	A
1361	darán	ADV
1321	handeln	V
1301	gegenwärtig	A
1293	wesentlich	A
1289	Hilfe	N
1285	Jahrhundert	N
1267	wirtschaftlich	A
1255	Mittwoch	N
1253	oft	ADV
1242	Vertreter	N
1230	Bevölkerung	N
1225	helfen	V
1215	bestimmen	V
1195	Sonntag	N
1193	berichten	V
1190	Angebot	N
1186	Dienstag	N
1177	her	ADV
1174	August ¹	N
1174	August ²	N

1172	Forderung	N
1166	bezeichnen	V
1159	dagegen	ADV
1158	eigentlich	A
1150	Donnerstag	N
1145	erhöhen	V
1138	folgend	A
1129	sonst	ADV
1118	daher	ADV
1110	Zusammenarbeit	N
1098	eben	ADV
1098	Verhältnis	N
1096	Zusammenhang	N
1094	dadurch	ADV
1094	entstehen	V
1085	Arzt	N
1081	gar	ADV
1075	gesamt	A
1073	ergeben	V
1061	Verhandlung	N
1021	Augenblick	N
1021	wissenschaftlich	A
1014	einigen	V
1001	Ministerpräsident	N
968	plötzlich	A
966	hinaus	ADV
939	genug	ADV
937	übrig	A
932	Außenminister	N
921	befinden	V
911	bekennen	V
908	ebenfalls	ADV
896	entsprechen	V
896	verlassen	V
888	gleichen	V
880	Gewerkschaft	N
879	zugleich	ADV
861	Voraussetzung	N
849	sogenannt	A
845	feststellen	V
841	bisherig	A
837	vergessen	V
836	aufnehmen	V

827	Vorstellung	N
826	betragen	V
825	Schwierigkeit	N
824	Sohn	N
811	günstig	A
807	annehmen	V
807	insgesamt	ADV
807	Vernunft	N
802	vertreten	V
790	Bewegung	N
782	entsprechend	A
782	künftig	A
778	lieb	A
770	eigentlich	ADV
766	behaupten	V
753	enthalten	V
749	erfolgreich	A
747	bestätigen	V
744	melden	V
741	zunächst	ADV
736	beweisen	V
723	heraus	ADV
718	beteiligen	V
716	holen	V
711	heutig	A
708	Wettbewerb	N
701	Amt	N
698	Anspruch	N
697	beschäftigen	V
694	Lehrer	N
691	Wirklichkeit	N
689	Mitarbeiter	N
688	nämlich	A
683	nämlich	ADV
680	gesellschaftlich	A
668	hoffen	V
661	Öffentlichkeit	N
656	Verantwortung	N
655	übernehmen	V
647	Geschäft	N
646	Erkenntnis ²	N
634	stattfinden	V
623	Bayer	N

619	hinter	A
618	Hauptstadt	N
610	mindestens	ADV
608	überzeugen	V
606	allein	A
606	teilnehmen	V
605	Arbeiterklasse	N
600	Hoffnung	N
592	danach	ADV
587	jedenfalls	ADV
586	manchmal	ADV
580	Gelegenheit	N
580	hervor	ADV
579	grundsätzlich	A
578	veröffentlichen	V
577	daraus	ADV
571	je	ADV
570	verhindern	V
568	behandeln	V
562	darstellen	V
562	Sonabend	N
560	erforderlich	A
552	bezahlen	V
551	ablehnen	V
550	lediglich	ADV
547	endgültig	A
545	bemühen	V
542	begründen	V
542	Vergangenheit	N
541	weiterhin	ADV
533	freien	V
532	Fernsehen	N
531	Landwirtschaft	N
523	echt	A
523	Jahrestag	N
523	Weltkrieg	N
521	gefährlich	A
519	danken	V
518	sorgen	V
516	Hälfte	N
516	lächeln	V
515	hängen	V
515	verschwinden	V

514	Gegensatz	N
508	Auseinandersetzung	N
508	genügen	V
504	Wissenschaftler	N
502	verzichten	V
500	immerhin	ADV
499	Antrag	N
499	Parteitag	N
499	zeihen	V
496	Gewinn	N
490	unten	ADV
490	wobei	ADV
489	verstärken	V
488	herzlich	A
488	niemals	ADV
485	zuerst	ADV
484	Himmel	N
484	verwirklichen	V
484	Vorjahr	N
481	überraschen	V
478	Anerkennung	N
477	benutzen	V
474	Hof	N
474	zuvor	ADV
473	Hochschule	N
471	Beteiligung	N
471	Nachrichtendienst	N
470	zusätzlich	A
468	wenigstens	ADV
467	diesmal	ADV
466	dauern ²	V
465	verpflichten	V
464	dringen	V
460	vermögen	V
458	ziemlich	A
457	Bewußtsein	N
457	diskutieren	V
457	Nachricht	N
456	endlich	A
456	selbstverständlich	A
456	verantwortlich	A
455	jeweils	ADV
455	wiederum	ADV

454	begreifen	V
450	empfangen	V
449	zuletzt	ADV
447	stellvertretend	A
446	eignen	V
444	leid	A

Tableau des mots néerlandais opaques

Fréquence	Lemme	Catégorie grammaticale
852027	zijn	V
474965	hebben ²	V
395743	niet	ADV
321883	worden ²	V
67496	uit	ADV
66149	al	ADV
64664	weer	ADV
58078	mens ²	N
54262	hoe	ADV
48135	maar	ADV
47494	hier	ADV
46508	houden ¹	V
46508	houden ²	V
46508	houden ³	V
43709	nemen	V
40226	eens	ADV
36693	alleen	A
36536	brengen	V
36461	beginnen ¹	V
36461	beginnen ²	V
35539	lopen ²	V
35539	lopen ³	V
35255	toe	ADV
34753	oog	N
34126	nooit	ADV
31573	eigen	A
31159	om	A
30340	zelfs	ADV
30131	mogen ¹	V
29234	voelen ²	V
27367	misschien	ADV
27148	mee	ADV
23459	alleen ¹	ADV
23459	alleen ²	ADV
23450	mogelijk	A
22864	geval	N
22159	echter	ADV
21770	vertellen ²	V

21233	waarin	ADV
20889	doen	N
19597	blijken	V
18883	duidelijk	A
18438	vooral	ADV
18184	begrijpen	V
18170	weg	N
17748	halen ¹	V
17731	heen	ADV
17227	gebruiken ¹	V
17227	gebruiken ²	V
17121	slechts	ADV
17053	hoog	A
17004	helemaal	ADV
16661	meest	A
16572	binnen	ADV
16517	waar	ADV
16432	echt	A
15709	ding	N
13782	daarom	ADV
13646	geloven	V
13509	gewoon	A
13245	bepalen ¹	V
13245	bepalen ²	V
12567	beetje	A
12378	lichaam	N
12228	waarop	ADV
11870	later	A
11779	reeds	ADV
11769	laatst	A
11749	helpen	V
11746	hangen ¹	V
11746	hangen ²	V
11718	betekenen	V
11604	geheel	A
11549	wijze ²	N
11421	waarvan	ADV
10582	wijzen ¹	V
10582	wijzen ²	V
10441	ervan	ADV
10197	hoeven ¹	V
10197	hoeven ²	V
10172	bedoelen	V

10109	veranderen	V
9954	pas	A
9843	daarbij	ADV
9631	bovendien	ADV
9623	mogelijkheid	N
9615	middel	N
9516	weten	N
9389	bepaald	A
9074	betreffen	V
9039	waarbij	ADV
8892	rol ²	N
8626	beweging	N
8591	rijden ³	V
8489	boven	ADV
8480	allemaal	ADV
8411	verschillen	V
8376	werkelijk	A
8362	Nederlands	A
8351	sluiten ³	V
8345	verwachten	V
8100	daarvan	ADV
8083	tenslotte	ADV
8045	daarmee	ADV
7997	zoon	N
7982	blik ²	N
7958	immers	ADV
7793	heten	V
7759	herinneren ¹	V
7759	herinneren ²	V
7613	waarmee	ADV
7591	gezien	A
7558	maatschappelijk	A
7550	genoeg	ADV
7535	gemakkelijk	A
7489	behoren	V
7396	gelukkig	A
7396	verschijnen	V
7255	betekenis	N
7028	glimlachen	V
6893	volledig	A
6888	gezegd	A
6881	verschillend	A
6852	verliezen ¹	V

6852	verliezen ²	V
6839	menen	V
6811	behoefte	N
6809	eenmaal	ADV
6808	gedaan	A
6760	eenvoudig	A
6674	ergens	ADV
6555	drukken ³	V
6548	neer	ADV
6531	waardoor	ADV
6437	werkelijkheid	N
6157	bewegen ¹	V
6157	bewegen ²	V
6018	vergeten ¹	V
6018	vergeten ²	V
5908	daardoor	ADV
5886	namelijk	ADV
5806	beschrijven	V
5728	overal	ADV
5715	sprake	N
5688	bekijken	V
5684	nogal	ADV
5677	hoofdstuk	N
5663	overigens	ADV
5601	allerlei	A
5579	omstandigheid	N
5555	trouwens	ADV
5548	toekomst	N
5420	omgeving	N
5399	hopen	V
5376	bevinden	V
5355	vervolgens	ADV
5354	wetenschap	N
5346	vertrekken ¹	V
5346	vertrekken ²	V
5332	betalen	V
5280	lief	A
5246	ontwikkelen ¹	V
5246	ontwikkelen ²	V
5233	omhoog	ADV
5191	zwijgen	V
5186	verklaren ¹	V
5186	verklaren ²	V

5177	voorstellen ¹	V
5177	voorstellen ²	V
5164	dringen ¹	V
5040	maatschappij	N
4997	genoemd	A
4996	beneden	ADV
4984	samenleving	N
4953	voorbij	ADV
4923	hulp	N
4914	ontmoeten	V
4910	houding	N
4908	gevaar	N
4803	wetenschappelijk	A
4801	opnemen	V
4798	borst ²	N
4787	behandelen	V
4709	hoek	N
4704	daarop	ADV
4684	gegeven	N
4592	herkennen	V
4585	schreeuwen ¹	V
4585	schreeuwen ²	V
4581	verlaten ¹	V
4581	verlaten ²	V
4543	ontvangen	V
4535	hoge	N
4526	dikwijls	ADV
4446	uiteindelijk	A
4437	geleden	A
4435	gelegenheid	N
4414	handelen	V
4411	wakker	A
4401	schuiven ²	V
4398	aanwezig	A
4382	loop	N
4326	verkopen ¹	V
4294	gevoelen	N
4291	onmogelijk	A
4291	verschijnsel	N
4285	veroorzaken	V
4254	behandeling	N
4233	hemel	N
4232	vliegen ²	V

4211	volkomen	A
4205	huwelijk	N
4194	genieten ¹	V
4182	verklaring	N
4175	verleden	N
4124	zover	ADV
4113	staren	V
4070	verlangen ¹	V
4070	verlangen ²	V
4069	bevatten	V
4065	moeilijkheid	N
4035	tegenwoordig	A
4018	tevens	ADV
4010	herinnering	N
3978	dagelijks	A
3976	ziekenhuis	N
3971	bouwen ²	V
3949	onderwerp	N
3948	erin	ADV
3946	belangstelling	N
3941	vrijwel	ADV
3914	fluisteren	V
3912	verstaan ¹	V
3905	verenigen ¹	V
3905	verenigen ²	V
3899	verdienen ¹	V
3899	verdienen ²	V
3873	links	A
3866	voegen ²	V
3849	ontbreken	V
3723	eraan	ADV
3721	zognaamd	A
3719	zelden	ADV
3700	bedenken ¹	V
3700	bedenken ²	V
3683	waaraan	ADV
3682	beperken ¹	V
3682	beperken ²	V
3676	adem	N
3663	voorkomen ¹	V
3663	voorkomen ²	V
3656	uiteraard	ADV
3630	gebeurtenis	N

3608	inhoud	N
3596	ineens	ADV
3590	lekker	A
3575	genomen	A
3570	geboren	A
3558	bedoeling	N
3540	hoop ²	N
3528	zichtbaar	A
3520	bevolking	N
3501	beleven	V
3488	vergelijken	V
3480	weigeren	V
3469	helft	N
3451	dwingen ¹	V
3451	dwingen ²	V
3451	verrichten	V
3447	haast	ADV
3445	aankomen ¹	V
3445	aankomen ²	V
3436	vertrouwen	V
3435	vangen ¹	V
3431	bewijzen	V
3427	betrekken ¹	V
3427	betrekken ²	V
3422	straks	ADV
3422	zomer	N
3420	eis	N
3381	vaststellen	V
3380	gezond	A
3373	lukken	V

Index

- acquisition**, 13, 46, 49, 62
- ajustement formel**, 57, 94, 128, 129,
130, 138, 140, 148, 149, 150, 153, 154,
172, 173, 176, 178, 179, 181, 182, 231,
232, 236, 237, 240, 244, 247, 253, 256,
257, 258, 261, 262, 264
- ajustement sémantique**, 57, 69, 81, 94,
153, 154, 155, 156, 157, 167, 168, 171,
172, 173, 179, 182, 211, 227, 232, 240,
244, 247, 253, 257, 261, 264
- alignement**, 28, 114, 126, 265
- allotopie**, 172, 173
- amorçage sémantique**, 52, 165, 166
- apprentissage**, 8, 10, 11, 37, 42, 47, 49,
62, 86, 98, 181, 258
- approximation**, 46, 55, 56, 57, 58, 59,
60, 62, 64, 70, 186, 260
- association lexicale**, 159
- champ phonologique**, 140, 147, 148,
149, 176, 177, 236
- champ sémantique**, 140, 159, 160, 161,
162, 163, 164, 168, 212, 214, 217, 218,
229, 230
- compréhension**, 8, 9, 10, 13, 14, 37, 38,
39, 40, 41, 42, 43, 44, 45, 46, 48, 49,
50, 51, 53, 56, 59, 60, 62, 63, 66, 79,
80, 91, 98, 107, 155, 156, 168, 173,
181, 182, 184, 189, 200, 204, 253, 257,
263, 265
- contexte**, 20, 21, 54, 57, 59, 61, 70, 71,
72, 82, 97, 106, 107, 113, 122, 124,
125, 149, 155, 164, 168, 169, 173, 181,
193, 257, 261, 263
- contrat de lecture**, 170, 171
- équivalence**, 12, 124, 125, 126, 175, 176,
177, 181, 189, 194, 261
- équivalent**, 12, 57, 63, 126, 154, 155,
156, 157, 168, 172, 175, 176, 183, 189,
190, 191, 193, 194, 195, 196, 197, 198,
209, 210, 212, 215, 217, 218, 220, 221,
224, 227, 228, 229, 230, 231, 232, 233,
234, 235, 236, 237, 240, 255, 262
- faux ami**, 8, 12, 70, 95, 96, 97, 98, 100,
101, 104, 105, 112, 114, 117, 123, 128,
155, 156, 256, 261
- forme**, 11, 12, 13, 17, 20, 21, 22, 23, 24,
25, 26, 31, 33, 34, 51, 55, 57, 61, 63,
64, 69, 70, 71, 75, 78, 79, 80, 81, 83,
85, 86, 87, 88, 90, 91, 92, 93, 94, 95,
96, 98, 99, 100, 105, 117, 125, 127,
128, 129, 138, 145, 146, 149, 153, 154,
155, 156, 157, 168, 177, 183, 189, 191,
210, 240, 247, 253, 258, 261, 262, 265
- graphème**, 26, 71, 102, 103, 104, 129,
130, 136, 142, 143, 148, 191, 195, 196,
198, 229, 230, 236
- homographie**, 22, 96, 98
- homonymie**, 21, 22, 85, 123, 155, 156,
160, 176, 191
- horizon d'attente**, 50, 170, 171
- hyperforme**, 87, 88
- hyperlangue**, 89, 90, 263, 264
- hyperlexème**, 87, 89, 238, 264, 392
- hyperlexique**, 14, 89, 91, 180, 181, 182,
188, 210, 232, 233, 237, 238, 241, 242,
245, 246, 247, 254, 263, 264, 293, 319,
344, 368
- hyperphonème**, 87, 88, 91, 140, 264
- hypersémème**, 87, 88, 89, 168, 264

- hypersystème**, 82, 84, 85, 87, 89, 91, 140, 168
- ICE**, 13, 42, 44, 47, 48, 50, 53, 54, 55, 57, 60, 61, 62, 63, 64, 69, 72, 129, 140, 145, 182, 258, 263, 265
- inférence**, 53, 54, 55, 57, 58, 61, 203
- intercompréhension**, 10, 13, 14, 16, 25, 32, 33, 34, 37, 38, 40, 41, 43, 44, 46, 47, 48, 61, 63, 64, 71, 72, 73, 76, 81, 82, 89, 98, 128, 183, 186, 258, 260, 263, 264
- interforme**, 85, 86, 87
- interlangue**, 86, 87, 90
- interlexème**, 83, 84, 85, 86, 87, 264
- interlexique**, 14, 83, 86, 91, 180, 181, 182, 188, 189, 190, 198, 199, 201, 202, 204, 210, 232, 233, 237, 242, 245, 247, 254, 263, 293, 307, 313
- interprétation**, 45, 46, 51, 52, 53, 54, 55, 62, 125, 128, 138, 172, 262
- intersémème**, 85, 86, 87
- intersystème**, 82, 83, 84, 85, 86, 87, 90
- isotopie**, 168, 170, 171, 172, 181, 265
- langue étrangère**, 8, 10, 11, 13, 34, 36, 38, 41, 44, 46, 48, 49, 50, 57, 59, 60, 61, 62, 63, 64, 83, 86, 89, 93, 95, 96, 98, 123, 128, 129, 130, 149, 154, 168, 173, 189, 264
- langue maternelle**, 10, 11, 13, 32, 35, 38, 40, 42, 43, 46, 49, 57, 59, 60, 63, 72, 73, 87, 93, 95, 117, 128, 129, 138, 149, 154, 155, 156, 168, 176, 211, 260, 261, 262
- lettre**, 11, 16, 26, 27, 29, 31, 59, 81, 84, 93, 98, 102, 105, 106, 108, 111, 112, 113, 114, 115, 116, 117, 130, 146, 149, 157, 165, 167, 175, 191, 197, 230, 231, 256, 261, 265
- lexie**, 19, 20, 163, 194
- lexique**, 8, 9, 10, 11, 12, 13, 17, 23, 32, 41, 78, 83, 84, 90, 91, 105, 107, 117, 149, 159, 160, 183, 207, 211, 214, 227, 250, 258
- morphème**, 17, 18, 19, 20, 79, 154, 163, 185, 194
- mot grammatical**, 185, 186, 187, 223
- mot lexical**, 185, 186, 187, 189, 199, 204, 207, 223, 238, 241, 246, 255, 258, 261
- mot-relais**, 70, 71, 72, 168, 171, 176, 178, 179, 209, 210, 224, 227, 228, 229, 230, 231, 233, 234, 239, 243, 255, 256, 265
- multilinguisme**, 33, 34, 35, 36, 40
- opacité**, 11, 26, 71, 72, 76, 261
- opaque**, 32, 53, 54, 65, 70, 71, 72, 76, 81, 175, 177, 194, 203, 252, 260, 262, 293, 392, 396, 403
- parasynonyme**, 124, 125
- parasynonymie**, 85, 124, 125, 176, 189, 261
- paronymie**, 85, 123, 155, 156, 176
- phonème**, 18, 59, 65, 71, 87, 88, 101, 102, 103, 104, 116, 128, 129, 130, 131, 136, 138, 140, 142, 143, 145, 146, 148, 150, 151, 153, 154, 163, 172, 177, 191, 195, 196, 197, 198, 229, 230, 236, 261, 265
- plurilinguisme**, 33, 36, 39, 72
- polysémie**, 21, 22, 160, 191
- proximité articulatoire**, 129, 136, 138, 145, 146, 153
- proximité formelle directe**, 82, 95, 98, 105, 112, 117, 123, 128, 129, 148, 156,

- 177, 178, 189, 210, 255, 256, 260, 261, 262
- proximité formelle indirecte**, 82, 87, 128, 129, 145, 148, 154, 156, 172, 173, 177, 178, 181, 255, 260
- proximité sémantique directe**, 82, 117, 128, 157, 173, 175, 176, 177, 178, 181, 189, 210, 255, 260
- proximité sémantique indirecte**, 82, 89, 140, 153, 154, 157, 158, 164, 167, 172, 173, 175, 177, 178, 181, 210, 211, 255, 256, 260
- redondance**, 59, 60, 106, 113
- sème**, 88, 89, 91, 92, 162, 163, 164, 165, 167, 169, 170, 172, 173, 174, 175, 176, 177, 178, 181, 211, 212, 215, 220, 223, 228, 229, 230, 231, 255, 261
- sème de transfert**, 165, 168, 171, 172, 173, 174, 176, 178, 179, 210, 224, 228, 229, 234, 235, 236, 265
- sémème**, 88, 162, 163, 164, 165, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 181, 211, 212, 215, 255
- sens**, 9, 10, 12, 13, 17, 20, 21, 22, 23, 24, 26, 34, 45, 50, 51, 53, 55, 57, 58, 60, 61, 63, 69, 93, 107, 118, 122, 123, 124, 126, 154, 155, 156, 157, 158, 160, 163, 165, 168, 169, 171, 173, 177, 181, 189, 190, 193, 207, 253, 258, 261
- signe graphique**, 102, 103, 129
- signifiant**, 12, 22, 65, 93, 94, 102, 103, 120, 154, 155, 156, 163, 165, 189
- signification**, 11, 12, 17, 20, 21, 22, 23, 24, 51, 56, 57, 62, 63, 64, 66, 69, 70, 71, 72, 78, 81, 85, 88, 91, 92, 93, 94, 96, 109, 117, 118, 120, 122, 123, 124, 126, 127, 154, 155, 156, 158, 162, 163, 164, 167, 169, 181, 189, 190, 191, 215, 216, 217, 220, 223, 240
- signifié**, 12, 17, 22, 25, 66, 93, 94, 95, 102, 103, 118, 120, 154, 155, 156, 163, 164, 165
- synonyme**, 120, 123, 124, 217
- synonymie**, 85, 123, 124, 167, 176, 261
- trame consonantique**, 101, 104, 105, 112, 117, 148, 149, 157, 175, 176, 177, 190, 191, 195, 196, 197, 198, 234, 235, 236, 237, 256, 262, 265
- transparence**, 11, 12, 13, 14, 16, 22, 23, 24, 25, 26, 29, 31, 32, 33, 60, 63, 64, 69, 71, 72, 74, 75, 76, 80, 81, 82, 83, 84, 85, 90, 91, 92, 94, 95, 127, 128, 175, 176, 177, 178, 179, 181, 182, 188, 193, 194, 198, 199, 200, 201, 202, 203, 204, 206, 207, 208, 209, 210, 227, 228, 237, 238, 239, 240, 241, 243, 244, 245, 248, 250, 252, 253, 255, 257, 258, 260, 261, 263, 265
- transparent**, 11, 12, 14, 23, 25, 31, 32, 52, 54, 55, 57, 63, 64, 65, 69, 70, 71, 72, 80, 81, 82, 83, 84, 87, 91, 96, 117, 155, 175, 177, 178, 179, 181, 182, 186, 189, 190, 191, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 210, 213, 217, 218, 220, 221, 224, 227, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 260, 261, 262, 264, 265, 392
- valeur**, 22, 120, 121, 123, 125, 162, 163
- vocabulaire**, 8, 9, 10, 25, 43, 49, 62, 64, 66, 79, 80, 161