

# Conception d'un dictionnaire fondamental de l'informatique et de l'Internet : sélection des entrées

Marie-Claude L'HOMME  
Université de Montréal<sup>1</sup>

## *Abstract*

This article describes a method for selecting terms in a French dictionary on computing and the Internet. First, the general objectives and the foreseen content of the dictionary are presented. Then, we propose a series of lexico-semantic criteria that are designed to help terminologists validate intuitions they have on the value of lexical units from the point of view of their incorporation in a specialized dictionary. The lexico-semantic criteria are applied on a list of specific lexical units generated automatically using corpus comparison methods. We show how an automatic selection of units supported by lexico-semantic criteria helps terminographers select relevant entries and increases systematicity.

## 1. Introduction<sup>2</sup>

On assiste depuis quelque temps à une volonté presque généralisée de diversifier le contenu des dictionnaires spécialisés. Il semble que les dictionnaires élaborés de manière classique ne répondent pas à toutes les exigences de certains de leurs utilisateurs, notamment les traducteurs et les rédacteurs. En mettant l'accent sur les renseignements de nature définitoire, les dictionnaires classiques ont tendance à occulter certaines dimensions du fonctionnement linguistique des termes. Par ailleurs, l'exploitation informatisée des corpus met au jour des données et des phénomènes qui passaient plus facilement inaperçus auparavant.

Parmi les ouvrages illustrant des approches non traditionnelles, se trouve le Binon *et al.* (2000), dictionnaire d'apprentissage consacré au français des affaires. Il propose, pour les termes jugés fondamentaux dans ce domaine, des entrées très élaborées contenant, outre les définitions et des contextes, des listes de cooccurrents, des expressions figées et des termes formellement apparentés. Le Dancette et Réthoré (2000), dictionnaire bilingue consacré aux termes de la distribution, permet à ses utilisateurs d'avoir une prise sur les relations conceptuelles entre termes en mettant au jour, au moyen de longues explications, des ensembles complexes de réseaux<sup>3</sup>. Un ouvrage moins récent, mais tout aussi

novateur, le Descamps (1976), propose pour des termes de géologie des entrées qui s'alignent sur des observations en corpus. On peut également citer certains dictionnaires (par exemple, le Cohen 1986) dont l'objectif premier est de recenser les cooccurrents trouvés dans l'environnement des termes.

Outre les entreprises lexicographiques a proprement parler, on peut lire un nombre croissant de plaidoyers en faveur d'une diversification et d'un enrichissement des descriptions terminographiques (Frawley 1988 ; Jousse et Bouveret 2003 ; Norman 2002, entre autres).

Le travail dont nous rendons compte dans ces pages s'inscrit dans cette mouvance. Il présente les premières étapes de l'élaboration d'un dictionnaire fondamental de l'informatique et de l'Internet en français (appelé DiCoInfo), à savoir les méthodes et critères qui président à la sélection des entrées. Si certains travaux ont permis de définir un vocabulaire fondamental dans une langue (Gougenheim *et al.* 1964) ou un vocabulaire général d'orientation scientifique (Phal 1976), aucune méthode à notre connaissance ne permet d'identifier le vocabulaire de base dans un domaine de spécialité<sup>4</sup>.

Nous avons élaboré une série de critères lexico-sémantiques et nous nous sommes aidée de métriques mises au point en statistique lexicale pour ramener une première liste d'unités lexicales à partir d'un corpus composé de textes d'informatique. Ce travail a permis d'évaluer une méthode de sélection de termes qui pourra être transposée à d'autres domaines et à d'autres langues. Le détail de ce travail constitue l'essentiel du présent article. La section 3 donne quelques généralités sur la méthode de sélection. Les sous-sections 3.1 et 3.2 traitent respectivement de la métrique utilisée et des critères lexico-sémantiques mobilisés pour la sélection des entrées. La section 4, quant à elle, décrit un ensemble de stratégies que nous avons mises au point afin de parfaire la sélection pour certains types d'unités (c'est-à-dire retrouver des unités omises par les métriques et retirer des unités relevées par les métriques que nous ne souhaitons pas conserver).

Mais auparavant, nous donnons, à la section 2, les objectifs généraux du dictionnaire fondamental de l'informatique et de l'Internet ainsi que le contenu envisagé de ses articles. Il importe de souligner que l'élaboration du dictionnaire est en cours et qu'il est susceptible de changer légèrement dans l'avenir<sup>5</sup>.

## **2. Objectifs et contenu provisoire d'un dictionnaire fondamental de l'informatique et de l'Internet**

Ce projet de dictionnaire comporte trois objectifs principaux que nous étudions dans la présente section. D'une manière générale, le dictionnaire est conçu comme une aide à la rédaction ou à la traduction de textes d'informatique. Il comporte de nombreux détails sur le sens et le fonctionnement linguistique des termes, mais évite de donner de longs développements encyclopédiques sur des concepts complexes.

D'abord, il vise à rendre compte des termes fondamentaux dans le domaine de l'informatique et de l'Internet. Par *terme*, nous entendons toute unité lexicale dont le sens peut être associé à un domaine spécialisé préalablement délimité. Ainsi, nous écarterons les unités lexicales non informatiques, même si elles sont

récurrentes dans les textes formant le corpus. Par *fondamental*, nous entendons l'ensemble des unités lexicales susceptibles de se trouver dans de nombreux textes d'informatique. Nous écartons les termes trop étroitement attachés à une spécialisation de l'informatique.

Parmi les unités lexicales retenues, on trouvera des noms (ex. *ordinateur, mémoire, virus*), des verbes (ex. *configurer, formater, activer*), des adjectifs (ex. *compilé, programmable, binaire*) et des adverbes (ex. *numériquement, dynamiquement, virtuellement*). Les unités retenues renvoient, par exemple, à des objets concrets (ex. *microprocesseur, mémoire, carte, bus, réseau*), à des représentations (ex. *caractère, instruction, données*), à des activités (ex. *configurer, naviguer, traitement*), à des propriétés (ex. *compatibilité, asynchrone, lisibilité*), à des animés (ex. *internaute, programmeur, webmestre*) et à des unités de mesure (ex. *mégahertz*).

La liste des entrées retenues et les définitions doivent refléter le plus fidèlement possible le réseau lexical du domaine. Par exemple, la plupart des verbes et des adjectifs se combinent de façon privilégiée avec un terme de nature nominale qui a été retenu font eux-mêmes l'objet d'une entrée. De même, les actants de termes prédicatifs retenus doivent eux aussi être décrits. Enfin, les unités entrant dans une relation paradigmatique ou syntagmatique avec un terme préalablement sélectionné devront également faire l'objet d'une description.

Deuxièmement, nous accordons une attention particulière aux multiples acceptions spécialisées que peuvent avoir certaines formes lexicales. Chaque article est consacré à un sens particulier. Les sens sont distingués à partir d'un certain nombre de tests lexico-sémantiques qui tiennent compte des interactions qu'un terme entretient avec d'autres termes et unités lexicales apparaissant dans les textes d'informatique.

Enfin, dans chacun des articles, nous dégageons l'ensemble des termes avec lesquels le terme en entrée entretient une relation de nature paradigmatique ou syntagmatique. La liste des relations sémantiques s'inspirent largement des relations retenues dans le Dictionnaire explicatif et combinatoire (Mel'čuk *et al.* 1984-1999), mais l'explication que nous en donnons est vulgarisée et s'aligne sur les explications données dans la version informatisée du DEC, à savoir le DiCo (Polguère 2003). Le tableau 1 présente quelques exemples de relations retenues jusqu'à maintenant. L'explication de la relation sémantique apparaît dans la troisième colonne. La fonction lexicale correspondante, inscrite dans la quatrième colonne, est donnée à titre indicatif.

Terme 1	Terme 2	Description de la relation sémantique : le terme 2 par rapport au terme 1	Fonction lexicale
AFFICHER 1 <sup>6</sup>	AFFICHAGE 1a	Nom du résultat	S <sub>res</sub>
ALGORITHME 1	ALGORITHMIQUE 1	Adjectif	A <sub>0</sub>
ARCHIVER 1	ARCHIVAGE 1a	Nom	S <sub>0</sub>
CHARGER 1	MÉMOIRE 1	Nom de la destination	S <sub>3</sub>
COMPATIBLE 1	INCOMPATIBLE 1	Antonyme	Anti
COMPILER 1	RECOMPILER 1	De nouveau	[De nouveau]
COMPRIMER 1	COMPRESSER 1	Synonymie	Syn
CONVIVAL 1	CONVIVALITÉ 1	Nom	S <sub>0</sub>
INTERNET 1	l'agent NAVIGUE 1 dans ~	L'agent utilise Internet	Real <sub>1</sub>
INTERNET 1	l'agent SE CONNECTE 1 à ~	L'agent se prépare à utiliser Internet	Prepar <sub>1</sub> Real <sub>1</sub>
LOGICIEL 1	le ~ TOURNE 2	Le logiciel fonctionne	Fact <sub>0</sub>
LOGICIEL 1	l'agent EXÉCUTE 1 le ~	L'agent fait fonctionner le logiciel	CausFact <sub>0</sub>
NUMÉRISER 1	NUMÉRISEUR 1	Nom de l'instrument	S <sub>3</sub>
ORDINATEUR 1	CLONE 1	Terme plus spécifique	Syn <sub>→</sub>
ORDINATEUR 1	APPAREIL 1, MACHINE 1	Terme plus général	Syn <sub>←</sub>
PROGRAMMER 1	PROGRAMMEUR 1	Nom de l'agent	S <sub>1</sub>
TABLEUR 1	LOGICIEL 1	Terme plus général	Syn <sub>←</sub>
VIRUS 1	le ~ CONTAMINE 1 le patient	Le virus intervient sur le patient	Fact <sub>2</sub>
VIRUS 1	le ~ INFECTE 1 le patient	Le virus intervient sur le patient	Fact <sub>2</sub>

Tableau 1 : Exemples de termes reliés sémantiquement

La figure 1 illustre l'entrée consacrée au verbe NAVIGUER 1 ; la figure 2, l'entrée nominale INTERNAUTE 1. Les définitions ne sont données qu'à titre indicatif, leur forme définitive n'étant pas encore arrêtée, et la liste des termes reliés sémantiquement n'est pas exhaustive.

#### NAVIGUER 1, V. TR.<sup>7</sup>

**Structure actancielle :** agent (utilisateur 1) navigue dans destination (réseau 2) au moyen de instrument(navigateur 1).

**Définition :** L'utilisateur 1 utilise un réseau 2 au moyen d'un navigateur 1 afin d'y rechercher de l'information 1.

**Contexte(s) :** *À partir de là, on pourra éventuellement retrouver l'information recherchée en naviguant dans les différentes pages offertes. Les requêtes de DNS font partie des requêtes les plus courantes lorsque l'on navigue sur Internet. L'ensemble de ces paramètres constitue la qualité de service (en anglais QOS, Quality of Service) visible par l'utilisateur, qu'il soit une personne voulant naviguer sur le Web ou une entreprise voulant mettre en place des serveurs.*

#### Termes reliés :

Synonyme	SURFER 1
Terme plus spécifique	VISITER 1
Nom	NAVIGATION 1a
Nom de l'agent	INTERNAUTE 1
Nom de la destination	INTERNET 1, WORLD WIDE WEB 1, WEB, TOILE 1, WWW
Nom de l'instrument	NAVIGATEUR 1

Figure 1 : Entrée NAVIGUER 1

**INTERNAUTE 1, N.**

**Structure actancielle :** internaute qui intervient sur patient(page 1, site 1, moteur de recherche 1) dans destination(réseau 2)

**Définition :** Personne qui recherche de l'information dans une page 1 ou un site 1 placés dans un réseau 2, parfois au moyen d'un moteur de recherche 1, également placé dans un réseau 2.

**Contexte(s) :** *En 2002, les Internautes non américains dépasseront en nombre les utilisateurs nord américain du réseau. L'internaute québécois qui accède au site français du moteur Yahoo ! trouvera surtout de l'information relative à la France. L'autoroute de l'information n' est rien si l'internaute ne peut l' utiliser à sa guise. les applications Internet sont les moyens qui permettent à l' usager d' atteindre efficacement les objectifs qu'il s' était fixés avant de prendre la route .*

**Termes reliés :**

Spécifique	VISITEUR 1
Intersection de sens	WEBMESTRE 1
Nom du patient	SITE 1, PAGE 1, MOTEUR DE RECHERCHE 1
Nom de la destination	INTERNET 1, WORLD WIDE WEB 1, WEB, WWW, TOILE 1
Le mot clé intervient sur le patient	l'~ UTILISE 1, CONSULTE 1, VISITE 1 le patient
Nom pour le mot clé intervient sur le patient	UTILISATION 1a, CONSULTATION 1a, VISITE 1a du ~
Le mot clé commence à intervenir sur la destination	l' ~ SE CONNECTE 1, SE BRANCHE 2 à la destination
Nom pour le mot clé commence à intervenir sur la destination	CONNEXION 1a, BRANCHEMENT 2a à la destination
Le mot clé intervient sur la destination	l' ~ NAVIGUE 1 dans / sur la destination
Nom pour le mot clé intervient sur la destination	NAVIGATION 1a de l'~ dans / sur la destination

Figure 2 : Entrée INTERNAUTE<sub>1</sub>**3. Sélection des entrées**

La collecte des termes et des renseignements permettant de procéder à leur description se fait à partir d'un corpus composé d'un peu plus de 50 textes français portant sur l'informatique. Le nombre total de mots s'élève à environ 600 000. Les thèmes abordés dans les textes sont l'initiation à la micro-informatique, l'Internet, le logiciel et le matériel, la programmation, les réseaux et les systèmes d'exploitation. Le détail du corpus d'informatique est présenté dans le Tableau 2<sup>8</sup>.

Les critères qui président à la sélection des termes sont les suivants (certains de ces critères, notamment ceux qui concernent les termes prédicatifs, sont décrits en détail dans L'Homme 1998, 2003, 2004a) :

a. L'unité extraite doit désigner une entité (matériel, logiciel, entité de représentation, unité de mesure ou un animé) du domaine de l'informatique (ex. *archive, carte, données, compilateur, programme ; internaute*) ;

b. S'il s'agit d'unités prédicatives – verbes, nominalisations, adjectifs, etc. –, elles sont extraites si les actants renvoient à des entités du critère a (ex. *charger* : *l'utilisateur charge un logiciel en mémoire; paquet : paquet de données*). Toutefois, la même unité prédicative peut se combiner avec des actants non spécialisés; si elle revêt le même sens avec ces autres actants, elle est éliminée.

c. S'il s'agit de dérivés morphologiques, ils sont sélectionnés s'ils sont sémantiquement apparentés à un terme sélectionné en fonction des critères a ou b (ex. *programme* : *programmer, programmable, reprogrammer, programmation*, etc. ; *archiver* : *archivage, archive*, etc.).

d. S'il s'agit d'une unité lexicale entrant dans une relation paradigmatique (autre qu'une relation morphologique déjà identifiée en c avec un terme sélectionné en fonction des critères a, b ou c, elle est extraite. Par exemple, si l'adjectif *virtuel* est sélectionné parce qu'il se combine avec des actants spécialisés (application du critère b), *réel* devra également être choisi puisqu'il s'oppose à *virtuel* dans certaines de ses acceptions (ex. *mémoire réelle ; mémoire virtuelle*).

Outre ces quatre premiers critères lexico-sémantiques, nous faisons intervenir la fréquence et la répartition. Une unité lexicale peut prétendre au statut de terme si elle est utilisée dans un nombre élevé de textes du corpus.

Subdivisions du corpus d'informatique	Taille des corpus	
	Nombre de textes	Nombre de mots
Initiation à la micro-informatique	8	116 821
Internet	12	102 972
Logiciel	4	78 412
Matériel	5	41 816
Programmation et réseaux	11	38 909
Systèmes d'exploitation	13	221 104
<b>Total</b>	<b>53</b>	<b>600 034</b>

Tableau 2 : Composition du corpus

### 3.1. Constitution d'une liste préliminaire au moyen de méthodes de statistique lexicale

Une liste de termes potentiels a été produite au moyen de méthodes de comparaison de corpus qui retiennent l'attention depuis quelque temps en terminologie (Ahmad *et al.* 1994 ; Chung 2003). Ces méthodes ont été retenues dans le cadre du présent projet car elles permettent : 1) de fonder le choix des entrées sur un corpus spécialisé (et non sur les connaissances préalables qu'un terminographe ou un spécialiste peut avoir d'un domaine) ; 2) de vérifier le critère de fréquence et de répartition évoqué à la section précédente. De plus, elles sont conçues pour dégager les termes simples contrairement à de nombreuses stratégies proposées au cours des dernières années qui se focalisent sur les termes complexes (Bourigault *et al.* 2001).

Les métriques auxquelles nous avons fait appel s'inspirent de travaux visant à dégager le vocabulaire spécifique d'un corpus (Lafon 1980 ; Lebart et Salem 1994) dont l'application à la terminologie a été proposée par Drouin (2003). Il

s'agit ici d'évaluer le caractère spécifique des unités d'un premier corpus, appelé *corpus d'analyse*, à un second corpus, appelé *corpus de référence*. La comparaison se fait sur des listes d'unités étiquetées et lemmatisées<sup>9</sup> au moyen d'un logiciel appelé *TermoStat* mis au point par Drouin (2003). Le programme distribue les unités dans les trois catégories suivantes :

- les spécificités positives (SP+) : celles dont la fréquence est plus élevée que celle observée dans le corpus de référence (nous tenons pour acquis que les termes que nous cherchons se situent dans cette catégorie);
- les formes banales (SP0) : celles dont la fréquence est la même que celle observée dans le corpus de référence ;
- les spécificités négatives (SP-) : celles dont la fréquence est moins élevée que celle observée dans le corpus de référence.

Deux méthodes ont permis de dégager les spécificités lexicales des textes d'informatique :

1. La première méthode consistait à comparer le corpus d'informatique à un corpus journalistique (*Le Monde* 2001 qui comprend environ 30 millions de mots). L'Annexe A présente les 25 unités ayant obtenu la valeur-test la plus élevée à la suite de la comparaison.

Une première évaluation (Lemay *et al.* 2005, à paraître) a montré qu'environ 50 % des spécificités positives générées par cette méthode sont répertoriées dans les dictionnaires spécialisés commerciaux<sup>10</sup>. Les taux de précision obtenus étaient de 48,12 % en ce qui concerne la comparaison avec le Ginguay et de 55,22 % pour la comparaison avec le Collin. Ces chiffres sont repris dans le Tableau 3.

	Nombre d'unités	Nombre de SP+ dans le dictionnaire	Précision	Rappel
SP+	690			
Ginguay	1 290 (854)	332	48,12 %	25,74 % (38,88 %)
Collin	1 311 (945)	381	55,22 %	29,06 % (40,32 %)

Tableau 3 : Précision et rappel de la méthode 1 confrontée au contenu de deux dictionnaires commerciaux

Mentionnons que les valeurs de rappel sont données dans les tableaux 3, 4 et 5 sous deux formes. Le premier pourcentage indique le rappel lorsque toutes les entrées du dictionnaire sont retenues ; le second – placé entre parenthèses – indique le rappel lorsque seules les entrées apparaissant également dans le corpus sont considérées.

2. La seconde méthode (M2) consistait à subdiviser le corpus d'informatique en six sous-corpus représentatifs (matériel, Internet, etc. : la subdivision apparaît dans le tableau 2) et à les comparer à tour de rôle à l'ensemble des textes

d'informatique. L'Annexe B présente les 25 spécificités positives ayant obtenu la valeur-test la plus élevée dans les six sous-corpus.

L'évaluation de cette seconde méthode (Lemay *et al.* 2005, à paraître) révèle que les résultats sont tout à fait comparables à ceux obtenus avec la méthode 1 (en fait, la première méthode obtient des scores un peu plus élevés), même si les unités elles-mêmes diffèrent. Environ 50% des spécificités positives sont également répertoriées dans les dictionnaires spécialisés commerciaux. Les taux de précision et de rappel obtenus par cette méthode sont reproduits dans le Tableau 4.

	Nombre d'unités	Nombre de SP+ dans le dictionnaire	Précision	Rappel
SP+	662			
Ginguae	1 290 (854)	308	45,53 %	23,88 % (36,07 %)
Collin	1 311 (945)	341	51,51 %	26,01 % (36,08 %)

Tableau 4 : Précision et rappel de la méthode 2 confrontée au contenu de deux dictionnaires commerciaux

Comme les deux méthodes produisent des listes différentes et que chacune ramène des termes intéressants en regard des objectifs que nous avons définis, nous avons fusionné les listes avant d'appliquer les critères lexico-sémantiques décrits à la section 3. Après avoir éliminé les doublons, nous avons obtenu 1021 unités. Cette liste combinée a elle-même été comparée au contenu des dictionnaires d'informatique. Les résultats de cette comparaison sont reproduits dans le tableau 5.

	Nombre d'unités	Nombre de SP+ dans le dictionnaire	Précision	Rappel
SP+	1021			
Ginguae	1 290 (854)	452	44,27%	35,04% (52,93%)
Collin	1 311 (945)	512	50,15%	39,05% (54,18%)

Tableau 5 : Précision et rappel obtenus par la liste combinée confrontée au contenu de deux dictionnaires commerciaux

Comme on peut le constater, la combinaison des deux méthodes a pour effet de faire diminuer légèrement la précision, mais accroît le rappel de manière significative.

Cette première évaluation a permis de confirmer l'intérêt de l'identification automatisée des spécificités lexicales pour la confection de dictionnaires spécialisés. Un nombre important d'unités définies comme des spécificités positives apparaissent également dans les dictionnaires spécialisés.

En outre, soulignons que les dictionnaires commerciaux sélectionnés pour cette évaluation et le corpus ayant servi à produire les spécificités lexicales ont un certain nombre de divergences, ce qui, à notre avis, accroît encore davantage l'intérêt des chiffres présentés dans les tableaux 3, 4 et 5. D'abord, de nombreux

textes du corpus étaient plus récents que les dictionnaires. De plus, contrairement au dictionnaire qui fait l'objet du présent article, le *Ginguae* et le *Collin* ont été conçus de manière complètement indépendante du corpus décrit dans le Tableau 2. Malgré ces divergences, les taux de précision oscillent entre 44 % et 51 %.

### 3.2. Validation de la liste de spécificités dans le cadre du projet

Nous avons donc voulu savoir si le calcul des spécificités lexicales et les deux méthodes décrites dans la section précédente ramenaient des termes correspondant à nos propres critères de sélection. La liste de 1021 spécificités (dont la première lettre est *A*, *C* ou *P*, rappelons-le) a donc été analysée et confrontée aux quatre critères lexico-sémantiques énumérés à la section 3.

Un aperçu des unités sélectionnées et rejetées est donné dans le tableau 6. Nous avons également indiqué si le terme était présent dans les dictionnaires ayant servi à l'évaluation précédente. Il est intéressant de constater les divergences entre le *Ginguae* et le *Collin*, d'une part, et entre les deux dictionnaires et notre propre sélection, d'autre part. On remarque d'ailleurs que les séries dérivationnelles ne reçoivent pas un traitement uniforme dans les deux dictionnaires. Par exemple, le *Ginguae* et le *Collin* ont chacun une entrée pour *algorithme*, mais omettent de retenir *algorithmique*.

Unité extraite automatiquement	Partie du discours	Présent dans le Ginguay	Présent dans le Collin	Sélectionné	Méthode
<i>abonné</i>	SBC	oui	oui	oui	1 et 2
<i>abonnement</i>	SBC	non	non	oui	1 et 2
<i>abonner</i>	VB	non	non	oui	1
<i>accès</i>	SBC	oui	oui	oui	1 et 2
<i>accessibilité</i>	SBC	oui	non	oui	1
<i>accessible</i>	ADJ	oui	oui	oui	1
<i>algorithme</i>	SBC	oui	oui	oui	1 et 2
<i>algorithmique</i>	ADJ	non	non	oui	1 et 2
<i>automatique</i>	ADJ	oui	oui	oui	1
<i>automatiquement</i>	ADV	non	non	oui	1
<i>automatisation</i>	SBC	oui	oui	oui	1
<i>caractère</i>	SBC	oui	oui	oui	1 et 2
<i>chemin</i>	SBC	oui	oui	oui	2
<i>clavier</i>	SBC	oui	oui	oui	1 et 2
<i>partage</i>	SBC	oui	oui	oui	1 et 2
<i>partageable</i>	ADJ	oui	non	oui	1
<i>partagé</i>	ADJP	non	non	oui	1 et 2
Unité extraite automatiquement	Partie du discours	Présent dans le Ginguay	Présent dans le Collin	Non sélectionné	Méthode
<i>aller</i>	VB	oui	oui	non	2
<i>absolument</i>	ADV	non	non	non	2
<i>accord</i>	SBC	non	oui	non	1
<i>affirmative</i>	SBC	non	non	non	1
<i>agir</i>	VB	non	non	non	1

<i>ainsi</i>	ADV	non	non	non	1
<i>alcool</i>	SBC	non	oui	non	2
<i>autre</i>	ADJ	non	non	non	1
<i>avantage</i>	SBC	non	non	non	1
<i>capable</i>	ADJ	non	oui	non	1 et 2
<i>centimètre</i>	SBC	non	oui	non	2
<i>chez</i>	PREP	non	non	non	2
<i>choix</i>	SBC	non	oui	non	1
<i>partir</i>	VB.	non	non	non	1
<i>permettre</i>	VB	non	oui	non	1 et 2

Tableau 6 : Sélection de spécificités positives par application de critères lexico-sémantiques

Les résultats de notre propre validation sont donnés dans le tableau 7. Les résultats obtenus à la suite de la comparaison avec les deux dictionnaires commerciaux ont été reproduits pour faciliter la comparaison. Le rappel n'est présenté qu'à titre indicatif puisque la sélection des termes n'est pas encore arrêtée de manière définitive. (L'Annexe C présente les termes dont la première lettre est *A* qui font présentement l'objet d'une entrée dans le dictionnaire d'informatique).

Liste combinée de SP+				
	Nombre d'unités	Nombre d'unités dans le dictionnaire	Précision	Rappel
SP+	1021			
Ginguyay	1,290 (854)	452	44,27 %	35,04 % (52,93 %)
Collin	1,311 (945)	512	50,15 %	39,05 % (54,78%)
Critères lexico-sémantiques	536	506	49,56 %	94,40 %

Tableau 7 : Spécificités retenues à la suite de l'application des critères lexico-sémantiques

Les résultats montrent que sur le total des spécificités, 49,56 % ont été retenues après l'application des critères lexico-sémantiques. La précision se situe entre les précisions obtenues à la suite de la comparaison avec le *Ginguyay* et le *Collin*. Ce résultat nous a quelque peu étonnée. En effet, comme notre sélection s'appuyait directement sur la liste (et, donc, sur le corpus ayant servi à produire la liste de spécificités), nous avons été surprise par le petit écart observé entre notre propre sélection et celle faite par les lexicographes du *Ginguyay* et du *Collin*. En revanche, le taux de rappel est nettement supérieur à celui observé pour les deux dictionnaires commerciaux. Même si la sélection des termes n'est pas encore tout à fait terminée dans le dictionnaire fondamental de l'informatique et de l'Internet, ce taux montre que la technique de spécificités présente un intérêt de premier plan pour la terminologie.

D'une manière générale, ces chiffres révèlent que même si les critères de sélection diffèrent d'un terminographe à l'autre, les taux de précision d'un calcul automatique des spécificités n'excéderont probablement pas 60 % compte tenu de la méthode et des corpus utilisés pour produire cette liste de spécificités.

## 4. Stratégies de correction des imperfections du calcul des spécificités lexicales

Pour corriger une partie des imperfections des listes de spécificités produites automatiquement, il convient de mettre au point des stratégies pour, d'une part, éliminer des unités indésirables et, d'autre part, retrouver des unités qui n'ont pas été définies comme des spécificités. La section 4.1 est consacrée au premier problème alors que la section 4.2 décrit les méthodes déployées pour venir à bout du second. En outre, certaines unités identifiées automatiquement donnent lieu à deux entrées distinctes puisqu'elles ont plus d'un sens. Le traitement réservé à ces unités fait l'objet de la section 4.3

### 4.1. Élimination des unités non terminologiques

Parmi les spécificités positives, se trouvent des unités que nous avons mises de côté systématiquement. Parmi celles-ci citons :

- les noms données à des commandes sous formes d'abréviations ou de noms anglais : ex. *enter*, *abort*, *atl-F4*, *^C*.
- des intitulés de fichiers ou de répertoires : ex. *\Windows\temp* ;
- différentes chaînes de caractères résultant d'un problème de découpage : ex. *acintosh* (plutôt que *Macintosh*).

D'autres unités ont été examinées plus attentivement, à savoir les termes complexes et les noms propres, puisque leur traitement varie en fonction d'un certain nombre de facteurs.

#### a) Termes complexes

Certaines unités définies comme des spécificités font en réalité partie de séquences complexes : par exemple, *économiseur* n'est utilisé qu'en combinaison avec *écran* (dans *économiseur d'écran*) ou *vive* avec *mémoire* (dans *mémoire vive*). Par ailleurs, des spécificités positives n'apparaissent pas intéressantes lorsqu'on les envisage isolément. Toutefois, certaines sont des collocatifs de termes sélectionnés. C'est le cas du verbe *utiliser* qui a été classé comme unité hautement spécifique par la méthode 1, mais *a priori*, il n'apparaissait pas pertinent, puisqu'il se combine avec d'autres unités lexicales.

Le traitement de ces unités varie en fonction de différents critères. D'une manière générale, les termes complexes font l'objet d'une entrée spécifique s'ils ont un sens non compositionnel (ex. *système d'exploitation*, *traitement de texte*). Le traitement réservé aux séquences non compositionnelles ou semi-compositionnelles dépend de la composante (tête ou modificateur) qui entre dans leur formation :

- s'il s'agit d'une séquence compositionnelle, mais que la tête de la séquence ne se rencontre que dans une combinaison spécifique, alors la

combinaison entière fait l'objet d'une entrée (ex. *économiseur d'écran* ; *manette de jeu*) ;

- s'il s'agit d'une séquence compositionnelle ou semi-compositionnelle, mais que le modificateur ne se rencontre que dans cette combinaison, elle est décrite dans l'article consacré à la tête (ex. *mémoire vive* est décrit sous *mémoire* ; *ordinateur personnel* est décrit sous *ordinateur*).

### **b) Noms propres**

La liste de spécificités positives contenait un nombre non négligeable de noms propres. Les annexes A et B, qui ne montrent que les 25 premières formes produites par les deux méthodes, illustrent ce phénomène (*Windows* et *Linux*, dans l'Annexe A ; *Internet*, *Epson*, *Linux*, *France*, dans l'Annexe B).

Les seuls noms propres faisant l'objet d'une entrée sont ceux qui s'utilisent comme des noms communs. Nous avons fait intervenir des critères comme l'emploi des déterminants ou le fait que le nom possède sa propre combinatoire. Par exemple, *Internet* s'emploie souvent avec un déterminant (*l'Internet*) et possède un nombre non négligeable de collocatifs (ex. *naviguer dans l'Internet*, *se connecter à l'Internet*). En vertu de ces critères, les unités comme *Linux*, *Windows* et *Epson* ont été écartées. En revanche, *World Wide Web*, *Internet* et *Pentium* sont retenus.

## **4.2. Recherche de termes n'apparaissant pas dans les spécificités**

Un nombre important de termes ne sont pas ordonnés par le traitement automatique comme des spécificités positives. Les taux de rappel obtenus à la suite de la première évaluation laissent supposer que de nombreux termes sont définis comme des formes banales ou des spécificités négatives. Pour résoudre ce problème, deux des critères lexico-sémantiques décrits à la section 3 sont utilisés.

### **a) Complétion des séries dérivationnelles**

Il arrive que le calcul automatique ne ramène qu'une partie d'une série de termes apparentés morphologiquement et sémantiquement. Par exemple, *automatique*, *automatisation* et *automatiquement* ont été définis comme étant des spécificités positives. En revanche, *automatiser*, qui apparaît pourtant dans le corpus, n'a pas été considéré comme une unité spécifique. De même, *adressable*, *adressage* et *adresse* sont définis comme des spécificités positives, mais *adresser* échappe au repérage automatique. Ces deux verbes seront facilement retrouvés dans le corpus d'informatique au moyen de techniques de recherche simples (avec un caractère jocker, par exemple).

### **b) Recherche d'éléments d'une structure actancielle**

La structure actancielle peut également guider le repérage d'unités qui n'ont pas été définies comme étant des spécificités positives. Par exemple, la liste des actants d'un terme de sens prédicatif peut être relevé automatiquement, mais pas le terme prédicatif lui-même. Il suffit alors de le retrouver dans le corpus.

### 4.3. Distinctions des acceptions

Parmi les unités apparaissant dans la liste de spécificités positives produites automatiquement et sélectionnées à la suite de l'application des critères lexico-sémantiques, certaines sont polysémiques. Deux cas de figure sont observés :

- l'unité a plus d'un sens, mais un seul de ces sens est lié au domaine de l'informatique (ex. *impression* : l'impression d'un document ; cette nouvelle version a fait bonne impression auprès des consommateurs) ;
- l'unité a plus d'un sens, mais deux ou plus sont liés au domaine de l'informatique (ex. *formater* : FORMATER<sub>1</sub> ; *formater une disquette* – FORMATER<sub>2</sub> : *formater un document*).

Ce dernier cas de figure donnera lieu à plus d'une entrée dans le dictionnaire d'informatique, alors que, dans le premier cas, seule l'acception spécialisée est décrite. Une série de critères est appliquée aux contextes dans lesquels apparaissent les unités afin de départager leurs sens. Nous n'aborderons pas ces critères en détail (nous invitons le lecteur à se reporter à L'Homme 1998 et 2003, 2004a pour une présentation systématique des critères servant à distinguer les acceptions).

## Conclusion

Dans cet article, nous avons montré de quelle manière le terminographe peut procéder à la sélection des entrées dans un dictionnaire spécialisé. Nous avons illustré des méthodes appliquées au domaine de l'informatique, mais elles peuvent facilement s'étendre à d'autres domaines. La comparaison de corpus réalisée automatiquement produit une liste d'unités dont environ la moitié se révélera intéressante pour la terminographie. La liste produite est ensuite soumise à l'examen du terminographe qui applique une série de critères lexico-sémantiques pour : 1) retenir les termes intéressants, 2) éliminer ceux qui ne sont pas pertinents et, 3) retrouver les termes que le traitement automatique n'arrive pas à déceler lui-même.

L'évaluation de la liste de spécificités lexicales produite automatiquement a montré que cette technique, bien qu'imparfaite, obtenait des taux de précision se situant autour de 50 % peu importe la méthode d'évaluation utilisée. Ce résultat nous a étonnée en ce qui concerne notre propre évaluation, puisque notre sélection de termes et la liste produite automatiquement s'appuyaient sur le même corpus. Toutefois, les taux de rappel varient en fonction de la méthode d'évaluation utilisée. Lorsque la liste est comparée à des dictionnaires commerciaux, le taux de rappel se situe entre 40 % et 55 %. Notre propre évaluation a montré que le taux de rappel dépassait 90 %. Bien que la sélection des termes dans notre dictionnaire ne soit pas arrêtée de manière définitive, ces résultats montrent à notre avis que le calcul des spécificités est un outil extrêmement précieux en terminologie.

Les méthodes que nous avons décrites sont intéressantes dans la mesure où elles permettent de fonder des choix terminographiques en fonction de ce qui peut

être observé en corpus spécialisés. Elles permettent également au terminographe de ne plus s'appuyer exclusivement sur ses connaissances du domaine (connaissances qui sont, au début d'un projet, souvent partielles) ou, du moins, de les valider avec plus d'assurance au moyen d'une série de contextes.

## Bibliographie

- Ahmad K. A. Davies, H. Fulford et M. Rogers (1994), « What's in a Term ? The Semi-automatic Extraction of Terms from Text ». In M. Snell-Hornby, F. Pochhammer et K. Kaindl (éd.), *Translation Studies. An Interdiscipline*. Amsterdam/Philadelphie : John Benjamins, 267-278.
- ATILF (2003). Laboratoire ATILF (Analyse et traitement informatique de la langue française). Page web consultée le 15 mai 2002, adresse URL : <http://www.atilf.fr/>.
- Binon J., S. Verlinde, J. Van Dyck et A. Bertels (2000), *Dictionnaire d'apprentissage du français des affaires. Dictionnaire de compréhension et de production de la langue des affaires*. Paris : Didier.
- Bourigault D., C. Jacquemin et M.C. L'Homme (2001), *Recent Advances in Computational Terminology*. Amsterdam/Philadelphie: John Benjamins.
- Chung T. M. (2003). « A Corpus Comparison Approach for Terminology Extraction ». *Terminology*, 9(2), 221-246.
- Cohen B. (1986). *Lexique de cooccurrents*. Brossard (Québec) : Linguatex.
- Collin S.M.H., F. Laurendeau et B. Mouget (1996), *Le Bilingue de l'informatique : dictionnaire français-anglais, anglais-français*, coll. « Peter Collins ». Middlesex : Peter Collin.
- Dancette J. et C. Réthoré (2000). *Dictionnaire analytique de la distribution. Analytical Dictionary of Retailing*. Montréal : Les Presses de l'Université de Montréal.
- Descamps J.L. (1976). *Dictionnaire contextuel de français pour la géologie : essai de classement d'un concordance de français scientifique et étude critique*. Paris : Didier.
- Drouin P. (2003), « Term Extraction Using Non-technical Corpora as a Point of Leverage ». *Terminology*, 9(1), 99-115.
- Frawley W. (1988), « New forms of Specialized Dictionaries ». *International Journal of Lexicography*, 1(3), 189-213.
- Ginguay M. (1998). *Dictionnaire français-anglais d'informatique : bureautique, télématique, micro-informatique*, 6<sup>e</sup> éd., 2<sup>e</sup> tirage avec mise à jour. Paris : InterEditions.
- Gougenheim G., R. Michéa, P. Rivenc et A. Sauvageot (1964): *L'Élaboration du français fondamental (1er degré) : étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, coll. « Linguistique appliquée ». Paris : Didier.
- Jousse A.L. et M. Bouveret (2003), « Lexical functions to represent derivational relations in specialized dictionaries ». *Terminology*, 9(1), 71-98.
- L'Homme M.C. (1998), « Définition du statut du verbe en langue de spécialité et sa description lexicographique ». *Cahiers de lexicologie*, 73(2), 61-84.
- L'Homme M.C. (2003), « Capturing the lexical structure in special subject fields with verbs and verbal derivatives : A model for specialized lexicography ». *International Journal of Lexicography*, 16(4), 403-422.
- L'Homme M.C. (2004a), *La terminologie : principes et techniques*. Montréal : Presses de l'Université de Montréal.
- L'Homme M.C. (2004b), « Sélection des termes dans un corpus d'informatique : comparaison de corpus et critères lexico-sémantiques ». In *Euralex International Congress. Proceedings*. Lorient (France), 583-593.
- Lafon P. (1980), « Sur la variabilité de la fréquence des formes dans un corpus ». *Mots*, 1, 128-165.
- Lebart L. et A. Salem (1994), *Statistique textuelle*. Paris : Dunod.
- Lecomte J. (1998), *Le catégoriseur Brill14-JL5 / WinBrill-0.3, INALF/CNRS*. Page web consultée le 13 février 2003, adresse URL : [http://www.atilf.fr/winbrill/BRILL14-JL5\\_WinBrill.doc](http://www.atilf.fr/winbrill/BRILL14-JL5_WinBrill.doc).
- Lemay C., M.C. L'Homme et P. Drouin (2005, à paraître), « Two Methods for Extracting "Specific" Uniterms from Specialized Corpora : Experimentation and Evaluation ». *International Journal of Corpus Linguistics*.

- Mel'čuk I. et al. (1984-1999), *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I-IV*. Montréal : Les Presses de l'Université de Montréal.
- Mel'čuk I., A. Clas, A. et A. Polguère (1995), *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve (Belgique) : Duculot / Aupelf - UREF.
- Namer F. (2000), « FLEMM: A Rule-based Morphological Analyzer of Inflected French Words ». *T.A.L.*, 41(2), 523-548.
- Norman G. (2002), « Description and prescription in dictionaries of scientific terms ». *International Journal of Lexicography*, 15(4), 259-276.
- Polguère A. (2003), « Collocations et fonctions lexicales : pour un modèle d'apprentissage ». In F. Grossmann et A. Tutin (éd.), *Les collocations. Analyse et traitement*, coll. Travaux et recherches en linguistique appliquée. Paris : Éditions de Werelt, 117-142.
- Rayson P. et R. Garside (2000), « Comparing Corpora Using Frequency Profiling ». In *Proceedings of the Workshop on Comparing Corpora, 38<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (ACL 2000)*, Hong Kong, 1-6.
- Phal A. (1976), *Vocabulaire général d'orientation scientifique (VGOS) : part du lexique commun dans l'expression*. Paris : CREDIF, Didier.
- Sager J.C. (1990), *A Practical Course in Terminology Processing*, Amsterdam / Philadelphia : John Benjamins.
- Temmerman R. (2003), « Innovative methods in specialised lexicography ». *Terminology*, 9(1), 117-135.

## Notes

<sup>1</sup> Observatoire de linguistique Sens-Texte (OLST), Département de linguistique et de traduction, Université de Montréal, C.P. 6128, succ. Centre-ville, Montréal (Québec) H3C 3J7 ; [marie-claude.l'homme@umontreal.ca](mailto:marie-claude.l'homme@umontreal.ca).

<sup>2</sup> Nous aimerions remercier les étudiants qui ont prêté main-forte dans le traitement des données : Sahara Iveth Carreño Cruz et Philippe Hanscom. Nous remercions également Anne-Laure Jousse et Chantal Lemay dont les travaux ont permis de parfaire les méthodes de sélection des termes d'informatique. Enfin, un merci à Patrick Drouin qui a fait des modifications à son programme afin de le rendre compatible avec les objectifs de ce projet.

<sup>3</sup> Temmerman (2003) fait une étude comparative du Binon *et al.* (2000) et du Dancette et Réthoré (2000).

<sup>4</sup> Les travaux d'Ahmad et Rogers (1994) et ceux de Chung (2003) dont il sera question plus loin visent à tester des méthodes d'identification automatique de termes. Toutefois, ils ne cherchaient pas à mettre au point une méthode de sélection d'entrées dans un dictionnaire spécialisé.

<sup>5</sup> Une version en ligne peut être consultée à l'adresse suivante : <http://olst.ling.umontreal.ca/dicoinfo>.

<sup>6</sup> Les chiffres placés à la suite des termes renvoient à leur numéro d'acception.

<sup>7</sup> La version informatisée du dictionnaire est réalisée sous forme de base de données relationnelles. L'entrée, la structure actancielle et la définition sont placées dans une première table. Une seconde table contient les contextes. Enfin, une troisième table modélise les différentes relations sémantiques. L'utilisateur peut également avoir accès aux articles des termes reliés. Par exemple, en sélectionnant INTERNAUTE<sub>1</sub>, à partir de la table des termes reliés à NAVIGUER<sub>1</sub>, l'utilisateur accède directement à l'article décrivant ce terme.

<sup>8</sup> Depuis que la sélection des termes décrite dans cet article a été faite, le corpus a été enrichi de nouveaux textes et compte actuellement environ 1 million de mots.

<sup>9</sup> L'étiqueteur utilisé est celui de Brill et, plus spécifiquement, l'adaptation qui en a été faite dans WinBrill (ATILF 2003). Le lemmatiseur est FLEMM (Lecomte 1998; Namer 2000).

<sup>10</sup> Les dictionnaires en question sont le Collin (1996) et le Ginguay (1998). L'évaluation a porté sur les spécificités positives dont les premières lettres sont A, C et P et sur les entrées correspondantes dans les dictionnaires choisis. Ces lettres ont été choisies parce qu'elles contenaient le plus grand nombre de formes. Il est à signaler que nous n'avons retenu, dans la liste de spécificités positives, que les mots lexicaux.

**Annexe A**  
**Liste des 25 premières spécificités obtenues**  
**par la comparaison avec le corpus Le Monde 2001**

Vingt-cinq premières SP+			
Unité	Partie du discours	Fréquence brute	Valeur test
Fichier	SBC	3956	360.825
commande	SBC	1902	201.749
Option	SBC	1486	182.338
Serveur	SBC	1166	180.477
Utilisateur	SBC	1117	167.307
configuration	SBC	845	162.82
Utiliser	VB	1996	161.788
Répertoire	SBC	1003	153.668
système	SBC	2699	152.979
disquette	SBC	609	148.431
ordinateur	SBC	1283	140.484
touche	SBC	855	138.648
logiciel	SBC	1166	137.933
imprimante	SBC	537	137.692
disque	SBC	1093	129.889
mémoire	SBC	1240	125.77
windows	SBP	613	125.066
clavier	SBC	579	122.24
caractère	SBC	1096	118.081
recommander	ADJP	516	117.004
linux	SBP	442	116.261
bit	SBC	382	110.72
paramètre	SBC	455	109.856
interface	SBC	412	109.64
permettre	VB	2625	108,873

**Annexe B**  
**Liste des 25 premières spécificités obtenues**  
**par subdivision du corpus d'informatique en sous-corpus**

Vingt-cinq premières SP+				
Unité	Partie du discours	Fréquence brute	Valeur test	Sous-corpus
internet	SBC	786	41.1712	SC2
mémoire	SBC	737	30.4842	SC1
option	SBC	1258	29.3401	SC6
site	SBC	306	28.9233	SC2
recherche	SBC	346	28.7659	SC2
imprimante	SBC	200	25.9019	SC4
papier	SBC	104	25.0858	SC4
fichier	SBC	2388	23.0704	SC6
implémentation	SBC	55	22.85	SC5

epson	SBP	44	22.101	SC4
recommander	ADJP	496	21.4903	SC6
algorithme	SBC	61	20.2627	SC5
synchronisation	SBC	63	20.0507	SC5
remplissage	SBC	38	19.994	SC5
noyau	SBC	438	19.9696	SC6
contrat	SBC	44	19.9632	SC5
laser	SBC	62	19.879	SC4
électronique	ADJ	202	19.8612	SC2
linux	SBP	422	19.6613	SC6
france	SBP	154	19.6566	SC2
étendue	SBC	75	19.1454	SC3
configuration	SBC	660	19.0063	SC6
imprimant	ADJ	52	18.9279	SC4
feuille	SBC	67	18.8884	SC4
réseau	SBC	292	18.7129	SC5

### Annexe C

#### Termes sélectionnés dans le dictionnaire (commençant par la lettre A)

abandon 1a; n. m.	activation 2a; n. f.	agréable 1; adj.
abandonner 1; v. tr.	activer 1; v. tr.	aide 1; n. f.
abonné 1; n. m.	activer 2; v. tr.	ajouter 2; v. tr.
abonnement 1a; n. m.	adaptateur 1; n. m.	aléatoire 1; adj.
abonnement 1b; n. m.	administrateur 1; n. m.	algorithme 1; n. m.
abonner 1; v. intr.	administration 1a; n. f.	algorithmique 1; adj.
accéder 1; v. tr.	administrer 1; v. tr.	algorithmique 2; n. f.
accéder 2; v. tr.	adressable 1; adj.	alias 1; n. m.
accéder 4; v. tr.	adressage 1a; n. m.	alimentation 1a; n. f.
accepter 1; v. tr.	adresse 1; n. f.	alimentation 2a; n. f.
accepter 2; v. tr.	adresse 2; n. f.	alimenter 1; v. tr.
accès 1a; n. m.	adresse 3; n. f.	alimenter 2; v. tr.
accès 2a; n. m.	adresse 4; n. f.	alimenter 3; v. tr.
accès 4a; n. m.	adresse relative 1; n. f.	allocation 2a; n. f.
accessible 1; adj.	adresser 1; v. tr.	allouer 2; v. tr.
accessible 4; adj.	affectation 1a; n. f.	allumage 1a; n. m.
accessoire 1; n. m.	affectation 2a; n. f.	allumer 1; v. tr.
accessoire 2; n. m.	affecter 1; v. tr.	alphabet 1; n. m.
accomplir 1; v. tr.	affecter 2; v. tr.	alphabet 2; n. m.
acheminement 1a; n. m.	affichage 1a; n. m.	alphabétique 1; adj.
acheminement 2a; n. m.	affichage 1b; n. m.	alphanumérique 1; adj.
acheminement 3a; n. m.	affichage 2a; n. m.	améliorable 1; adj.
acheminer 1; v. tr.	affichage 2b; n. m.	amélioration 1a; n. f.
acheminer 2; v. tr.	afficher 1; v. tr.	amélioration 1b; n. f.
acheminer 3; v. tr.	afficher 2; v. tr.	amélioré 1; adj.
actif 2; adj.	agenda 1; n. m.	améliorer 1; v. tr.
action 1a; n. f.	agrandir 1; v. tr.	amorçable 1; adj.
activable 1; adj.	agrandissement 1a; n. m.	amorçage 1a; n. m.
activation 1a; n. f.	agrandissement 1b; n. m.	amorcer 1; n. f.

- amorcer 1; v. tr.  
 amovible 1; adj.  
 analogique 1; adj.  
 analyse 1a; n. f.  
 analyse 2a; n. f.  
 analyser 1; v. tr.  
 analyser 2; v. tr.  
 analyseur 1; n. m.  
 analyste 1; n. m.  
 analytique 1; adj.  
 animation 1a; n. f.  
 animation 1b; n. f.  
 annuaire 1; n. m.  
 annuaire 2; n. m.  
 annulation 1a; n. f.  
 annuler 1; v. tr.  
 antislash 1; n. m.  
 antivirus 1; n. m.  
 aperçu 1; n. m.  
 apostrophe 1; n. f.  
 apparaître 1; v. intr.  
 appareil 1; n. m.  
 apparition 1a; n. f.  
 appel 1a; n. m.  
 appeler 1; v. tr.  
 applicatif 1; adj.  
 applicatif 2; n. m.  
 application 1; n. f.  
 application 2; n. f.  
 appui 1a; n. m.  
 appuyer 1; v. tr.  
 arborescence 1; n. f.  
 arborescent 1; adj.  
 arbre 1; n. m.  
 architecture 1; n. f.  
 archivage 1a; n. m.  
 archive 1; n. f.  
 archivé 1; adj.  
 archiver 1; v. tr.  
 arrêt 1a; n. m.  
 arrière-plan 1; loc. adj.  
 ascendant 1; adj.  
 ascii 1; n. m.  
 assemblage 1; n. m.  
 assemblage 3; n. m.  
 assembleur 1; n. m.  
 assembleur 2; n. m.  
 assertion 1; n. f.  
 assigner 1; v. tr.  
 assistant 1; n. m.  
 assistant 2; n. m.  
 astérisque 1; n. f.  
 asymétrique 1; adj.  
 asynchrone 1; adj.  
 attachement 1a; n. m.  
 attachement 1b; n. m.  
 attacher 1; v. tr.  
 attaquer 1; v. intr.  
 attente 1; loc. adv.  
 attribuer 1; v. tr.  
 attribuer 2; v. tr.  
 attribut 1; n. m.  
 attribut 2; n. m.  
 attribut 3; n. m.  
 attribution 1a; n. f.  
 auteur 1; n. m.  
 authentification 1a; n. f.  
 authentifier 1; v. tr.  
 auto-extractible 1; adj.  
 automatique 1; adj.  
 automatique 2; adj.  
 automatiquement 1; adv.  
 automatisation 1a; n. f.  
 automatiser 1; v. tr.  
 autonomie 1; n. f.  
 autorisation 1a; n. f.  
 autoriser 1; v. tr.  
 auto-test 1; n. m.  
 auxiliaire 1; adj.  
 azerty 1; n.