

LEXICOGRAPHIE ET HYPERTEXTES

Georges VIGNAUX*, David PIOTROWSKI*, Quyen KIEU**

* CNRS - INaLF
vignaux@poleia.lip6.fr

** Lab. d'Informatique de Paris 6

Résumé : Notre recherche s'inscrit dans la perspective de l'informatisation du dictionnaire Trésor de la Langue Française. Elle a pour objectif de concevoir les fonctionnalités d'une interface de lecture d'un fonds lexicographique. Le prototype HYPERLEX, ici présenté, est programmé sous HYPERCARD, convertible en Java-Script. HYPERLEX a la structure d'une pile de cartes. Chaque carte comporte quatre sous-fenêtres qui constituent des zones de lecture et de traitement des différentes données accessibles. La première sous-fenêtre est destinée à accueillir les données d'un fonds lexicographique (un article de dictionnaire). Les fenêtres 2, 3 et 4 sont, respectivement, les fenêtres d'extraction, d'indexation et d'interprétation. L'opération d'extraction donne au lecteur les moyens de collecter certaines portions textuelles qu'il envisage de travailler. Ce travail va consister à qualifier, par voie d'indexation, les mots ou les groupes de mots des portions extraites. Il s'agit ainsi d'assigner à des textes une qualification sémantico-cognitive univoque susceptible de faire par la suite l'objet de traitements divers. La quatrième fenêtre délivre enfin une vision des différents extraits à travers la grille sémantique des qualifieurs. Elle fournit un état de l'activité d'appropriation et d'interprétation textuelle par indexation sémantique.

Mots clés : dictionnaire, extraction, indexation, interprétation, cognition.

Abstract : Our research is finalized into the perspective of the computerisation of the dictionary Trésor de la Langue Française. Our goal is to conceive the practical modalities of an interface for reading a lexical data base. The prototype, presented here is called HYPERLEX. It is programmed in HYPERCARD and will be converted in Java-Script. HYPERLEX has the structure of a stack. Each card of the stack includes four areas devoted to four different tasks. The first area is devoted to collecting data for the lexical funds (i.e., a dictionary article). The second, third and fourth areas allow the user to operate extractions, indexations and interpretations of the dictionary articles. Extraction consists in collecting text passages that will be later processed. Indexing consists in applying a cognitivo-semantic qualification to words or phrases from the passage. Finally, the interpretation area displays text passages along with the semantic qualifiers thus representing the current state of textual interpretation.

Key words : dictionary, extraction, indexing interpretation, cognition.

INTRODUCTION

Notre recherche s'inscrit dans la perspective de l'informatisation du dictionnaire *Trésor de la Langue Française (TLF)*. Elle a pour objectif de concevoir les fonctionnalités d'une interface de lecture d'un fonds lexicographique. Une telle couche fonctionnelle, qui donne au lecteur la possibilité d'investir l'ensemble des données soit par la voie de procédures de recherche multicritères, soit par la voie d'une « navigation » suivant un système de renvois non directifs (ou réseau de liens), est communément appelée *hypertextuelle*. Elle est ici envisagée comme *système d'aide au parcours* dans des dictionnaires informatisés, au sens où les dispositions fonctionnelles de cette interface pourront susciter une investigation raisonnée du fonds lexicographique et amener ainsi le lecteur à accomplir une authentique « lecture » du dictionnaire, c'est-à-dire un parcours du texte lexicographique au terme duquel il aura pu se construire une représentation personnelle et finalisée des matériaux investis - en mettant en évidence, par exemple :

- (i) des organisations du lexique ;
- (ii) des motivations d'usage de mots ;
- (iii) des regroupements de faits de sens ;
- (iv) des potentialités de signification ;
- (v) des distributions en domaines cognitifs *i.e.* de connaissances et de savoirs.

PROBLÉMATIQUE

Lorsqu'il est question d'informatiser un dictionnaire, la première chose qui vient à l'esprit, c'est de chercher à l'établir sous la forme d'une base de données. Les motivations de cette façon de penser l'informatisation d'un texte lexicographique sont, au fond, très « naturelles ». Elles trouvent leurs racines à la fois dans une conception populaire des dictionnaires, qui voit ces textes comme des recueils de prescriptions normatives sur la langue, et dans des considérations pratiques : versée sous format électronique, la somme d'informations que recèle un dictionnaire est bien plus aisément accessible.

Nous rappellerons les motifs de ces deux points de vue avant de contester la validité d'une réduction des fonds lexicographiques au format de bases de données et d'argumenter en faveur d'une informatisation de type hypertextuelle.

Le dictionnaire, source normative

Pour le grand public la langue est avant tout perçue comme un système normatif, un ensemble de prescriptions sur les façons d'écrire et d'assembler les mots, dont la maîtrise atteste d'une certaine position dans la hiérarchie sociale. Dans cette perspective, un ouvrage dont la matière est la langue (dictionnaire ou grammaire) est considéré comme une sorte de « table de lois », à laquelle les locuteurs vont se rapporter pour évaluer la justesse de leurs tournures ou encore pour les guider dans leurs formulations et leurs rédactions. L'orthographe, parangon de la

normativité linguistique, constitue dans cette optique la préoccupation majeure, et les dictionnaires, jugés dépositaires de cette norme, seront le plus souvent consultés à cet effet.

Le second usage dominant des dictionnaires, à savoir la recherche du sens des mots, est lui aussi le corrélat d'une conception populaire du lexique, à savoir : le lexique comme nomenclature. Cette conception s'exprime dans l'opinion, répandue, que la richesse du vocabulaire témoigne de l'ampleur du savoir. S'il en est ainsi, c'est que les objets du savoir progressivement considérés, objets du monde ou de pensée, se sont vu en tant que tels attribuer des dénominations (Rosch, 1978 ; Perec, 1985). La connaissance des vocables nommant ces objets et de leur sens qui en donne la description, équivaldrait alors à la stricte connaissance des dits objets. Confusion entre l'ordre des mots et l'ordre des choses, entre l'ordre des signifiés et celui des concepts, qui amène les locuteurs à rechercher dans le dictionnaire, sous la rubrique *définition*, de quoi combler leur ignorance lorsqu'ils rencontrent un terme supposé désigner quelque chose d'ambigu voire d'inconnu. Les principaux types d'emploi des dictionnaires sont donc fortement déterminés par des représentations populaires de la langue et, corrélativement, par le statut socioculturel conféré à ce genre d'ouvrages.

La raison pratique

Dans leurs utilisations les plus rudimentaires comme dans leurs emplois plus savants qui prennent appui sur les structures fines de l'analyse lexicographique (notamment en traitement automatique des langues naturelles), les dictionnaires se présentent donc comme des bases de données : ensembles structurés d'informations à consulter. Tout semble concourir pour affirmer que la transposition du texte lexicographique (du support papier au support électronique) est « naturelle », voire nécessaire ; et on pourra aller jusqu'à dire que l'appareil informatique « procédural », c'est-à-dire l'appareil informatique en tant que système de traitement par application de procédures sur des données structurées, est le complément pratique attendu d'un dictionnaire ou, en d'autres termes, que c'est dans les structures matérielles et opératoires de l'informatique procédurale que le texte lexicographique trouve sa forme aboutie (Bruillard, Piotrowski, Vignaux, 1995).

Les limites de la conception « procédurale »

Cette conception procède d'un certain nombre de présupposés sur la légitimité desquels il convient de s'interroger. Car considérer l'informatique comme la forme « naturelle » du texte lexicographique, c'est affirmer au moins deux choses : d'une part que le contenu du texte source est transposable sur le nouveau support électronique sans pertes, réductions ou déformations, et, d'autre part, que la lecture d'un dictionnaire s'accomplit suivant les voies exactes des traitements de l'informatique. Or ces deux présupposés sont inexacts.

D'une part, de nombreuses études ont montré que l'intelligibilité du texte lexicographique prend appui sur des caractéristiques typographiques et topographiques (positionnement relatifs des composantes textuelles). Ainsi, Ide, Le Maître et Véronis (1991) montrent comment certaines segmentations opérées dans le texte

d'un article en vue de l'établir sous forme de base de données relationnelle ruine son unité sémantique et, corollairement, sa portée informative. En adaptant un exemple donné par ces auteurs, considérons l'article *maintenir* du Petit Robert. Cet article enregistre quatre acceptions, les deux premières étant : (1) *conserver dans le même état ; faire ou laisser durer*, (2) *affirmer avec constance, fermeté*. Dans le cadre d'une base relationnelle, à cette portion d'article correspond le tableau de définitions suivant :

vedette	subdivision	définition
maintenir	1	conserver dans le même état
maintenir	1	faire ou laisser durer
maintenir	2	affirmer avec constance, fermeté.
maintenir

Or, dans un tel tableau, les deux définitions de la première acception sont distribuées dans des cases distinctes et de même rang fonctionnel que la case contenant la définition de la seconde acception. Comme l'observent Ide, Le Maître et Véronis (*ibid.* p. 9) « the two subsenses [de l'acception 1] are more closely related to each other than to sens 2, 3 (...), but the tabular format obscures this fact ». Dans un base relationnelle, on perd donc une information précieuse qui procède de la juxtaposition des deux portions textuelles.

D'autre part, si les dictionnaires sont difficilement réductibles à des bases de données, ce n'est pas du seul fait de la composition de leurs contenus textuels, c'est aussi du fait des ambiguïtés des catégories de l'analyse lexicographique. On sait que pour constituer une base de données, il est nécessaire de disposer d'une liste arrêtée et univoque de propriétés descriptives (les types d'*attributs*). Or il se trouve que les formes descriptives de la lexicographie ne sont ni univoques, ni en nombre fini. Deux exemples simples le montreront. Considérons, à la suite de F. Henry (1993), les articles de *obligation* (niveau C) et *ordre* (niveau IIIA1a) dans le *TLF*, soit :

OBLIGATION

- *Vieilli. Avoir obligation (de qqc. à qqn).*

Être redevable (de bons offices, de quelque bienfait à quelqu'un)

ORDRE

• *Avoir, recevoir ordre de + inf, Avoir reçu la mission, la consigne de.*

On voit qu'une même information de structure syntaxique (à savoir *élément verbal + complémentation*) est introduite dans *obligation* à l'aide d'un « indicateur de construction » (en gras) et dans *ordre* sous la forme d'un « syntagme-type » (élément d'illustration, en italique). Les types d'attribut « construction grammaticale » et « syntagme-types » se recouvrent donc ; ils ne déterminent pas une catégorisation univoque des informations lexicographiques.

En désaccord avec les exigences de la mise en forme informatique procédurale (univocité et explicitation), les contenus lexicographiques ne sont donc pas, du moins complètement, réductibles à une somme d'informations distribuées suivant un système des rapports et de catégories stabilisés et réguliers : on a affaire à des signifiés plutôt qu'à des concepts. Autrement dit encore, la forme du texte lexicographique est une forme : (i) matériellement déterminée et (ii) susceptible de montrer différents visages (Nanard, 1994 ; Piotrowski, 1996).

Ce qui amène à contredire le second présupposé : puisque le contenu lexicographique ne suit pas les formes d'un ordre abstrait, univoque et arrêté, sa signification se construit dans sa « saisie » : le sens du texte lexicographique met en jeu les dimensions constitutives d'une activité de lecture, à savoir : (i) les compétences du sujet cognitif et (ii) la forme matérielle du texte, en vue de produire (iii) une interprétation de ce texte. La pratique du dictionnaire ne suit donc pas les voies des traitements informatiques qui, à l'inverse supposent l'existence de contenus (i) univoquement déterminés et (ii) constitués indépendamment des méthodes qui permettent d'y accéder (les actes de lecture).

Pour conclure

Les dictionnaires ne sauraient donc être réduits à de simples bases de données : ce ne sont pas des recueils d'informations univoques et structurées (en fait l'idéal de la lexicologie), mais d'authentiques textes que le lecteur doit s'approprier en produisant une interprétation. L'informatisation d'un dictionnaire doit être pensée comme une « hypertextualisation » c'est-à-dire comme une inscription du texte lexicographique dans un cadre de fonctionnalités de *lecture* (Balpe, 1990 ; Nielsen, 1990 ; Laufer & Scavetta, 1992). Il ne s'agira plus alors de procéder simplement à des extractions de données mais d'effectuer des *navigations*, en d'autres termes, d'effectuer des *parcours de lecture* au fil desquels des données sont sélectionnées, collectées, puis qualifiées et ordonnées dans la perspective d'en produire une interprétation.

Dans le paragraphe suivant, nous présenterons la structure fonctionnelle du système HYPERLEX. Pour la motiver, nous proposerons en préalable, en suivant le fil d'un exemple simple, d'examiner quelles fonctionnalités et quelles opérations automatisables sont susceptibles de favoriser une investigation « raisonnable » du texte lexicographique (*i.e.* un parcours qui se conclut par la construction d'une représentation personnelle du sens des matériaux lexicaux abordés et/ou qui invite à prolonger la lecture).

LE PROTOTYPE HYPERLEX

Un exemple de lecture assistée

Supposons que le lecteur choisisse de débiter ses investigations lexicographiques par le vocable *salade*. Dans un premier temps, donc, il va charger l'article *salade* dans le système HYPERLEX. Dans la fenêtre consacrée aux contenus d'articles, va alors apparaître, par exemple, le texte suivant (version abrégée de *salade* du *Grand Robert*)

salade

-A.

- 1. *De la salade ; une salade* : mets fait de feuilles d'herbes potagères crues, assaisonnées d'huile, de vinaigre et de sel (vinaigrette), souvent servi avant le fromage ou le dessert.

- 2. *Par ext.* (Rabelais, Lettre à d'Estissac, 15 févr. 1536). *Une salade* : plante cultivée, légume dont on fait la salade.

- 3. (1690 ; qualifié par un compl. de nom, un adj.). Mets froid fait de légumes, cuits ou crus, de viande, d'oeufs, de crustacés, etc. seuls ou en mélange, assaisonnés d'une vinaigrette, généralement servi en hors-d'oeuvre.
- 4. *Salade de fruits* : fruits menus ou coupés, servis froids accommodés avec un sirop, une liqueur.
- B. *Fig.*
- 1. (1856). *Fam.* Mélange confus, réunion hétéroclite. - Confusion, désordre, enchevêtrement.
- 2. [a] (1901). *Vendre sa salade* (péj.), se dit d'un camelot qui bonimente, et, par ext., d'un artiste médiocre qui se produit (- Fourmi, cit. 8).
- [b] (1890 ; souvent au plur.). Histoires, mensonges.
- 3. *Pop.* Chercher une, des salades à qqn, une querelle, des histoires

Nous avons vu précédemment que le sens du texte lexicographique procède d'une mise en rapport de ses éléments constitutants. Un type de rapport particulièrement intéressant pour orienter la lecture est celui de la « récurrence ». La présence dans le texte d'éléments répétés signale en effet l'existence d'un certain relief sémantique. Dans HYPERLEX, une fonction de *recoupement* permet de faire apparaître les termes « insistants » : il s'agit donc de donner au lecteur des pistes susceptibles d'orienter son parcours de lecture vers des « points cardinaux » de l'article. Dans notre exemple, la fonction « recoupement » met en relief les vocables *mets* et *mélange*. Et le lecteur, sera alors naturellement « invité » à en savoir plus sur le contenu des vocables *mets* ou *mélange*. Une fonction de navigation suivant des liens définitionnels lui permet d'accéder rapidement à l'article *mets*, soit (version abrégée) :

METS [me] n. m.

- Chacun des aliments cuisinés qui entrent dans l'ordonnance d'un repas*

Corrélativement une fonction automatisable de recherche de relations sémantiques « robustes » entre éléments des différents articles consultés permet de pointer une opposition entre *mélange* et *ordre* (cette fonction n'est pas encore implémentée dans HYPERLEX). En effet, dans le *Robert électronique* on a (les mots récurrents sont soulignés) :

Définition (ordonnance) = mise en ordre, disposition selon un ordre

Antonyme (ordre) = désordre

Synonyme (mélange) = confusion, désordre

Synonyme (désordre) = confusion

Par ailleurs, un rapide coup d'oeil à l'article *hors-d'oeuvre* permet de lire :

- 2. (1690). *Cour.* Petit plat que l'on sert au début du repas, après le potage et avant les entrées* (ainsi appelé, d'après Littré, «parce que, originairement, ils ne faisaient pas partie de l'ordre régulier dans lequel les plats principaux étaient rangés»).

Ce qui confirme la présence de la notion d'*ordre* (en mots : *ordre*, et en idées : au *début* du repas, *après* le potage et *avant* les entrées) dans le sens de *salade*.

L'opposition [ordre/désordre] semble donc un structure susceptible de guider raisonnablement l'investigation de l'article *salade*. Ce sera au lecteur de tester cette hypothèse. Pour ce faire il dispose de fonctionnalités d'*indexation* qui permettent d'assigner aux vocables une identité sémantique répondant au schéma interprétatif choisi (repérage [ordre/désordre]). Ainsi, il pourrait assigner les étiquettes (entre crochets) suivantes (on choisira de distinguer l'ordre relatif à l'organisation du repas et celui relatif à la composition du plat) :

SALADE

-A.

- 1. *De la salade ; une salade* : mets [aliment] [ordre/repas] fait de feuilles d'herbes potagères [aliment=végétal] crues, assaisonnées d'huile, de vinaigre et de sel (vinaigrette), souvent servi avant le fromage ou le dessert.[ordre/repas]

- 2. *Par ext.* (Rabelais, Lettre à d'Estissac, 15 févr. 1536). *Une salade* : plante cultivée, légume [aliment=végétal] dont on fait la salade.

- 3. (1690 ; qualifié par un compl. de nom, un adj.). Mets [aliment] [ordre/repas] froid fait de légumes, cuits ou crus, de viande, d'oeufs, de crustacés, etc [aliment=X] seuls ou en mélange [désordre/plat], assaisonnés d'une vinaigrette, généralement servi en hors-d'oeuvre [aliment] [désordre/repas]

- 4. *Salade de fruits* : fruits menus ou coupés, servis froids accommodés avec un sirop, une liqueur.

-B. *Fig.*

- 1. (1856). *Fam.* Mélange confus, réunion hétéroclite. - Confusion, désordre, enchevêtrement. [désordre]

Une fonction d'*interprétation* permet ensuite de donner une lecture synthétique des articles suivant le filtre des catégories sémantiques choisies. On obtient :

SALADE

-A.

- 1. [aliment] [ordre/repas] [aliment=végétal] [ordre/repas]

- 2. [aliment=végétal]

- 3. [aliment] [ordre/repas] [aliment=X] [désordre/plat] [désordre/repas]

- 4.

-B. *Fig.*

- 1. [désordre]

On met ainsi en perspective la double nature du vocable *salade* : dans son principe « d'ordre », en tant qu'elle est située dans l'ordonnement du repas (acception 1) la salade désigne précisément le végétal assaisonné [aliment=végétal] ; dans son principe de « désordre » (acception 3), la salade désigne un plat perturbant la distribution régulière des mets [désordre/repas] et ne détenant pas d'identité alimentaire précise [désordre/plat] et [aliment=X]. Cette observation conduit alors à revenir sur l'acception *au figuré*. En effet, dans l'emploi au figuré (qui procède

généralement de la substitution du terme générique), la composante [désordre] n'est pas spécifiée (relativement au plat ou au repas).

Le traitement hypertextuel accompli débouche donc sur une interrogation : le désordre auquel font référence des expressions comme *il a fait une salade indescriptible* est-il un désordre de dérégulation d'un bon ordonnancement (type [désordre/plat]) ou un désordre de non-organisation, non spécification (type [désordre/aliment]) ? Cette interrogation peut être produite automatiquement par le système qui signalera la sous spécification de la clé d'index [désordre]. Et le système hypertextuel invitera ainsi le lecteur à de nouvelles explorations cognitives (Vignaux, 1994, 1996).

Par exemple, le lecteur pourra parcourir les différentes exemplifications de *salade* (au figuré) et tester le rendement des deux types du désordre que véhicule la salade. Il observera peut-être que seul le désordre de « non organisation » semble attesté. (par exemple : « *Ils s'amusaient à dire du mal des gens, à se moquer des figures inconnues (...) Du Roy déclara : - Quelle salade de société !* » (Maupassant, *Bel-Ami*, II, VII) ou : « *Ce qui était épouvantable pour lui, Alexandre, c'était le désordre. La salade des unités sur les routes, la cohue des réfugiés, ces femmes et ces enfants juchés sur des charrettes ...* » (Robert Merle, *Week-end à Zuydcoote*, p. 51). Ce qui induit de nouvelles incitations pour explorer le fonds lexicographique, en vue de recueillir des informations susceptibles d'expliquer l'absence du désordre de type « désorganisation » dans le sens métaphorique de *salade*, ou encore pour réfuter les indexations accomplies.

A travers cet exemple élémentaire, on aura suggéré comment certaines fonctionnalités de lecture/qualification peuvent donner au lecteur des vecteurs raisonnés d'investigation du texte lexicographique et le motiver à en extraire des représentations sémantiques et cognitives (ici, par indexation).

La structure fonctionnelle d'HYPERLEX

Le prototype ici présenté est programmé sous HYPERCARD. Ce contexte de programmation a été choisi pour quatre raisons principales : (i) comme tout langage-objet, HYPERCARD offre une grande souplesse de programmation qui facilite l'élaboration de versions successives, (ii) HYPERCARD comporte des outils de création et de gestion d'objets hypertextuels (boutons, fenêtre, menus, texte...) directement adaptés à nos fins, (iii) HYPERCARD permet d'enregistrer des versions d'HYPERLEX comme « applications portables », c'est-à-dire comme programmes fonctionnant sans faire appel à la « base » d'HYPERCARD, (iv) SUPERCARD, une version compatible PC de HYPERCARD, a été récemment mise au point, ce qui évite à HYPERLEX de se retrouver cloîtré au monde Mac. Par ailleurs, le code de SUPERCARD étant convertible en Java-Script, HYPERLEX est aussi susceptible d'être transporté sur WWW, notamment en connexion au TLF informatisé disponible sur le site de l'INaLF.

HYPERLEX a la structure d'un empilement de cartes. Chaque carte de la pile comporte quatre sous-fenêtres qui constituent des zones de lecture et de traitement des différentes données accessibles. À l'ouverture d'HYPERLEX (Figure 1) apparaît la première carte (vide) ainsi qu'une barre générale de commandes.

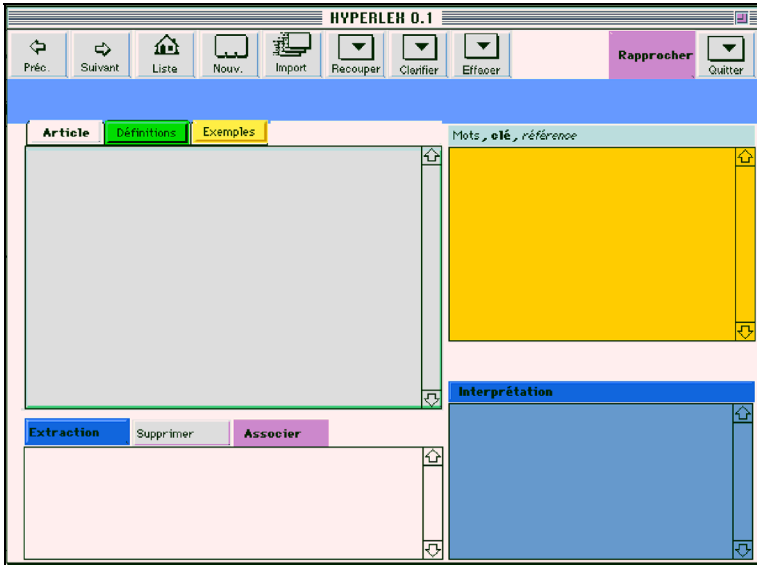


Figure 1.

Les quatre fenêtres sont disposées suivant un ordre qui correspond à la chronologie d'un cycle élémentaire de lecture, soit :

1	3
2	4

La première fenêtre : les données

La première sous-fenêtre est destinée à accueillir les données d'un fonds lexicographique élaboré par ailleurs ; par exemple, l'article *peur* (Figure 2). Les boutons qui lui sont associés (**article**, **définition**, **exemple**) permettent d'accéder aux différentes composantes ou sous-composantes de ce fonds.

Cette fenêtre pourrait sembler superflue puisque les fonctionnalités d'accès et d'affichage des articles et des éléments d'articles d'un dictionnaire électronique sont généralement livrées avec celui-ci. En fait, la fenêtre 1 s'avère nécessaire parce que les fonctionnalité d'HYPERLEX ne sont pas simplement des fonctionnalités « classiques » de lecture (ou d'accès) mais aussi des fonctionnalités de traitement et d'écriture. Aussi, il faut définir un champ de données propres où les opérations d'HYPERLEX peuvent être effectuées. C'est la raison d'être de la fenêtre 1. Par ailleurs, comme cette fenêtre doit permettre d'effectuer des traitements sur les différentes sortes de données disponibles (définitions, exemples...), ainsi que de donner une présentation des traitements effectués, il sera donc nécessaire de la structurer suivant les formes et les catégories des articles. Se pose alors le problème de la compatibilité de la structure de la fenêtre 1 avec la structure des données lexicographiques — ce qui implique certains choix sur la position fonctionnelle d'HYPERLEX.

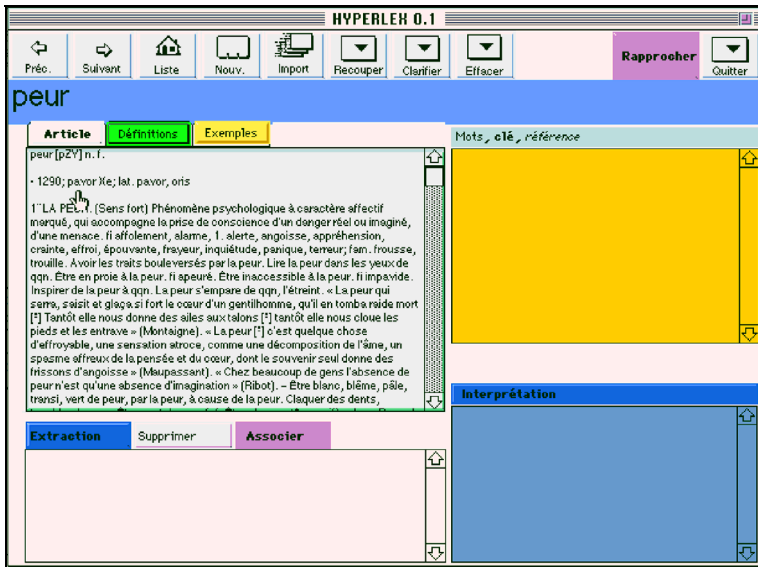


Figure 2.

Dans un premier temps, nous avons choisi de concevoir le prototype d'HYPERLEX comme une surcouche fonctionnelle générique et, pour ce faire, nous sommes restreints à trois types de contenus. Cette option n'est que temporaire. D'une part, par nécessité : un outil de lecture lexicographique ne saurait en effet se limiter à des catégories descriptives aussi rudimentaires que celles choisies, et d'autre part, par programmation : parce qu'il est à terme prévu d'utiliser les ressources du TLF et, partant, il faudra réaliser une connexion consistante TLF/HYPERLEX.

Les fenêtres 2, 3 et 4 sont, respectivement, les fenêtres d'extraction, d'indexation et d'interprétation. Rappelons qu'avec la fenêtre 1, elles constituent les quatre éléments d'une carte et qu'à chaque carte est donc associé un article du fonds.

La barre de commandes



La barre de commandes comporte deux sortes de boutons. D'une part des boutons de gestion des cartes : **Précédent**, **Suivant**, **Liste**, **Nouvelle** et **Import**, et, d'autre part, les boutons de traitement des données lexicographiques : **Recouper**, **Clarifier** et **Effacer**.

Le Recoupelement

L'action sur le bouton **Recouper** (barre de commandes) a donc pour effet de faire apparaître en gras les unités récurrentes du texte. Ainsi, par exemple, le mot *peur* et certaines unités grammaticales (formants de l'antidictionnaire), les vocables *mort* et *danger* sont des éléments récurrents de l'article *peur*. Cette information qui

focalise l'attention du lecteur et qui dévoile certaines circonstances et/ou certains aspects de la peur délivre une suggestion sur d'éventuels prolongements à donner à la lecture.

Dans les versions futures d'HYPERLEX, il sera important d'améliorer le fonctionnement de la commande *recouplement*, notamment en utilisant un lemmatiseur et en élargissant le champ d'action du recouplement aux produits de l'indexation (cf. infra). Pour retourner au texte « non recoupé », il suffit de cliquer sur un des boutons **article**, **définition**, **exemple**.

La deuxième fenêtre : l'extraction

La fenêtre d'extraction permet au lecteur de collecter les fragments de texte qu'il juge intéressants ou pertinents à son projet de lecture. Les textes extraits entrent alors dans une logique de lecture « confrontative » (i.e. mise en rapport oppositif) : en superposant dans un même cadre divers fragments textuels, le lecteur est naturellement porté à les aborder en effectuant des synthèses et des différenciations entre les significations qu'il identifie. L'opération d'extraction porte sur des éléments de la première fenêtre qui sont sélectionnés par action *cliquer/glisser* de la souris. Les portions textuelles ainsi extraites apparaissent en gras dans leur fenêtre d'origine (par exemple, dans le champ *définition* de la carte *peur* — Figure 3)

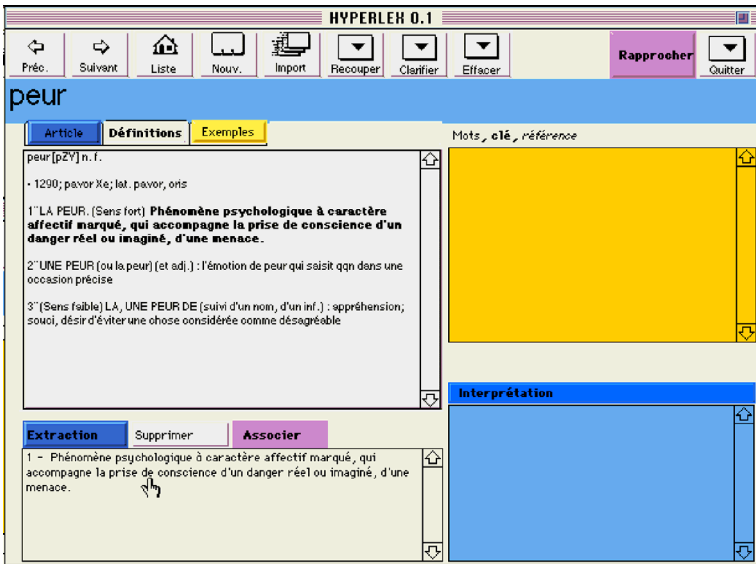


Figure 3.

La troisième fenêtre : l'indexation

L'opération d'extraction donne donc au lecteur les moyens de collecter certaines portions textuelles qu'il envisage de travailler à ses fins propres. Ce travail va essentiellement consister à qualifier, par voie d'indexation, les mots ou les groupes de mots des portions extraites. Il s'agit ainsi d'assigner à des textes en langue naturelle une qualification sémantico-cognitive univoque susceptible de faire par la suite l'objet de traitements divers.

En général, la lecture confrontative que favorise la fenêtre d'extraction va porter sur des extraits puisés dans différents articles. Les extraits sélectionnés seront rassemblés dans la fenêtre **extraction** d'une carte de **synthèse**. L'opération de qualification des éléments extraits par indexation va être effectuée avec le bouton **Associer**. Dans un premier temps on sélectionne dans la fenêtre **extraction** l'élément à qualifier, puis on clique sur le bouton **Associer**. Apparaît en incrustation un champ destiné à introduire le nom de la clé d'index qualifiant les unités sélectionnées. Après avoir validé le choix de la clé d'index, la relation de qualification, par exemple « phénomène psychologique » = « état » (suivi du nom de la carte où cette qualification a été opérée), est enregistrée et affichée dans la troisième fenêtre (Figure 4) qui permet le suivi des qualifieurs introduits. Les éléments ayant fait l'objet d'une qualification apparaissent alors en gras dans la fenêtre d'extraction.

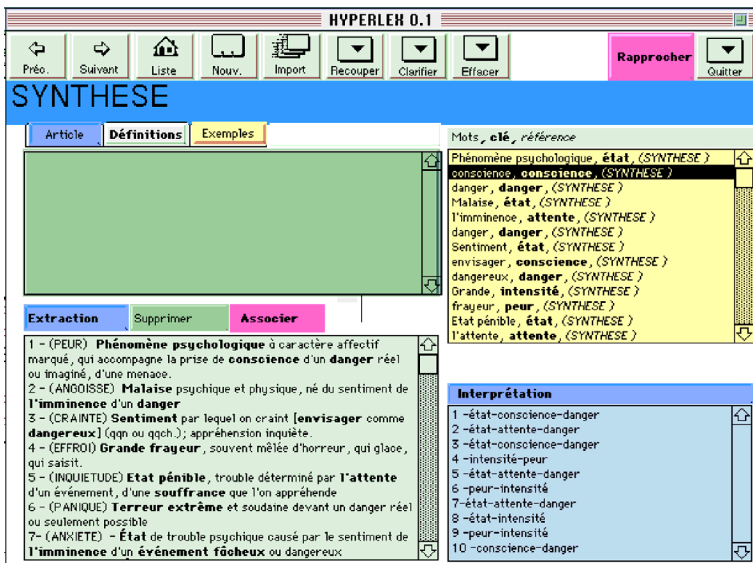


Figure 4.

La quatrième fenêtre : l'interprétation

La quatrième fenêtre délivre une vision des différents extraits à travers la grille sémantique des qualifieurs. Elle fournit un état de l'activité d'appropriation textuelle par indexation sémantique.

Après avoir accompli un travail d'assignation de spécifications interprétatives, le lecteur peut souhaiter le visualiser sous une forme synthétique ; c'est l'objet du bouton **Rapprocher** qui fait apparaître un tableau de distribution de traits.

À l'examen de ce tableau, le lecteur est en mesure d'opérer un certain nombre de constats (par exemple la prééminence ou les distributions complémentaires de certains traits) susceptibles de l'inciter à affiner son approche, par exemple en introduisant des nouvelles qualifications sémantiques ou en

prolongeant sa lecture vers d'autres articles, ou qui peuvent aussi lui suggérer des configurations de sens qu'il juge pertinentes au recueil et à la construction personnelle de *domaines cognitifs* et de faisceaux de sens issus de son activité de lecture interprétative.

À la suite de quoi, il ne lui reste plus qu'à cliquer sur le bouton **Quitter**.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Balpe J.-P. (1990). *Hyperdocuments, hypertextes, hypermédiats*, Paris, Eyrolles.
- Bruillard E., Piotrowski D., Vignaux, G. (1995). *Du dictionnaire considéré comme hypertexte*, Rapport d'Étape, INaLF, 120 p.
- Henry F. (1993). « Analyse descriptive du Trésor de la Langue Française », in *Pour l'informatisation du TLF*, Rapport Préalable, Nancy : CNRS-INaLF.
- Ide N., Le Maître J., Veronis J. (1991). *Outline of a model for lexical databases*, Documents du GRTC-CNRS, Marseille, n° 496.
- Laufer R. & Scavetta D. (1992). *Texte, hypertexte, hypermedia*, Paris : PUF ; Que-sais-je ?
- Nanard M. (1974). « L'apport des travaux de recherche dans les hypertextes aux techniques éducatives », *Actes du séminaire Hypermédiats, Éducation et Formation*, IUFM de Créteil, p. 7-34.
- Nielsen J. (1990). *Hypertext and Hypermedia*, Boston : Academic Press.
- Perec G. (1985). *Penser/Classer*, Paris : Hachette.
- Piotrowski D. (1996). *Dynamiques et structures en langue*, Paris : CNRS-Editions, à paraître.
- Rosch, E. (1978). *Cognition and Categorization*, Hillsdale, New Jersey : Lawrence Erlbaum.
- Vignaux G. (1992). *Les sciences cognitives : une introduction*, Paris : La Découverte, 1992 ; Le Livre de Poche, 1994.
- Vignaux G. (1996). « Hypertextes, dictionnaires : approche sémantique, perspective cognitive », in D. Piotrowski, (éd.), *Lexicographie en Informatique, Autour de l'informatisation du Trésor de la Langue Française*, Actes du Colloque International de Nancy, Paris : Didier Érudition.