

Des « normes catégorielles » : structuration cognitive et/ou linguistique des catégories sémantiques

Danièle Dubois*, Jacques Poitou[◇]

“ By category is meant a number of objects that are considered equivalent. Categories are generally designated by names ”. Rosch, 1978 ; p.30

Résumé : Cet article analyse ce qui est désigné, en psychologie cognitive, comme des normes catégorielles, et qu'une approche linguistique peut considérer comme des « listes de termes ». Ainsi ce « produit » d'une activité psycholinguistique particulière (l'énumération de mots isolés) peut, soit être considéré comme le « reflet » de structures catégorielles permettant d'évaluer les hypothèses psychologiques relatives aux prototypes et à la typicalité, soit comme un objet linguistique contraint par les propriétés du lexique et d'un discours très particulier (une liste). Nous tentons de montrer que les premières hypothèses doivent tenir compte des « biais » introduits par l'épaisseur du fonctionnement linguistique et que l'accès aux représentations cognitives, objectivées dans cette tâche, n'est en rien direct. Ce travail entend ainsi illustrer la nécessaire coopération disciplinaire entre linguistique et psychologie dans les sciences cognitives à condition cependant que les observables construits dans chaque discipline ne se trouvent pas réduits à un « cognitif » indéterminé.

Mots clés : listes de mots, catégories, lexique, typicalité

Abstract: **“Categorical norms” : the cognitive and/or linguistic structuring of semantic categories** . This article analyses what, in cognitive psychology, are called “categorical norms”, and a linguistic approach can consider as “lists of words”. Thus, this “production” of a particular psycholinguistic activity (the enumeration of isolated words) can be considered *either* as a “reflexion” of categorical structures, enabling the evaluation of psychological hypotheses concerning prototypes and typicality, *or* as a linguistic object constrained by the properties of the lexicon and a very particular sort of discourse (a list). We aim to show that the psychological hypotheses must take into account the

* Cnrs/ Lcpe

[◇] Lyon-2

“bias” introduced by the layer of linguistic function, so that the access to cognitive representations (as revealed by this task) is by no means direct. This study thus illustrates the disciplinary cooperation between psychology and linguistics that is necessary in cognitive science. A pitfall to be avoided is the reduction of the observable phenomena constructed by each discipline to an indeterminate “cognitive” common denominator.

Key words: lists of words, categories, lexicon, typicality

Ce travail traite de la question générale des relations entre langage et cognition à partir de l'analyse des relations entre formes lexicales et catégories cognitives. Cette question qui relève traditionnellement du champ de la philosophie occupe en effet actuellement une place centrale dans les recherches cognitives, où elle se trouve déplacée sur le terrain de la validation empirique des diverses hypothèses, tant en linguistique qu'en psychologie cognitive. En particulier, le débat opposant des positions relativistes (les formes lexicales déterminent les catégories de pensée) vs universalistes (les catégories de pensées sont universelles indépendamment de la diversité des langues) a été renouvelé dans les recherches sur la catégorisation et les théories des prototypes (Berlin & Kay, 1969 ; Rosch, 1978 ; Dubois, 1986 ; 1991). La sémantique lexicale tend actuellement à faire à nouveau appel à des considérations extralinguistiques (Kleiber, 1997a ; Wierzbicka, 1996) qui, fondées dans les données empiriques de la psychologie ou de l'expérience, garantiraient un fondement cognitif aux structures lexicales (Vandeloise, 1991, par ex.). Les recherches comparatives récentes sur une grande diversité de langues contribuent cependant à apporter des arguments à l'hypothèse d'une autonomie relative des systèmes linguistiques et des catégories cognitives (cf. Lucy, 1992 ; Taylor, 1995 ; Foley, 1997 ; Dubois, 1997, pour des revues récentes), et amènent à dépasser l'opposition entre un universalisme absolu et un relativisme extrême. La question n'est plus de savoir *si* le langage influence la pensée (ou réciproquement), mais plutôt *comment* langage et pensée contribuent à l'élaboration des connaissances (cf. par exemple Waxman, 1999).

Cependant, malgré la prise en compte de plus en plus précise de la diversité des constructions lexicales et des catégories cognitives, une difficulté demeure, celle de mettre en place des procédures de questionnement comparables (toutes choses étant égales par ailleurs) à celles menées dans notre aire culturelle, et qui puissent spécifier plus précisément les relations entre formes linguistiques et catégories cognitives. Ainsi, en marge d'un travail d'élaboration de protocoles expérimentaux universellement valides, sur des terrains où les langues, les comportements langagiers et les cultures sont différents des nôtres, le présent travail reste circonscrit à l'analyse d'un protocole emblématique de la recherche sur la catégorisation : la production de listes de mots isolés en réponse à un terme inducteur.

Ces listes sont désignées, dans la littérature psychologique, sous le vocable de « normes catégorielles ».

En effet, en psychologie, parmi les modes de constitution de la preuve de l'universalité des structures catégorielles (prototypes et typicalité) et des dénominations de base, les résultats de ces protocoles ont été et restent largement utilisés. Ainsi,

“ Item **output** is normally taken to reflect some aspect of storage, retrieval, or category search. Battig and Montague (1969) provided a normative study of the probability with which college student listed **instances** of superordinate semantic **categories**. The order is correlated with prototypicality ratings (Rosch 1975). Furthermore, using artificial **categories** in which frequency of experience with all **items** was controlled, Rosch et al. (1976) demonstrated that the most typical **items** were the first and most frequently produced **items** when subjects were asked to list the **members** of the **category**. ” (Rosch, 1978, p. 35).

On peut noter que, dans ce type de discours, représentatif des recherches psychologiques sur la catégorisation, les formes lexicales ne sont pas désignées comme telles (il est fait mention d' « *item* »), et la notion de catégorie comme celle d'instance n'est pas davantage spécifiée comme objet linguistique et/ou psychologique. En d'autres termes, il y a indétermination et sous-spécification des objets en question, et la relation entre lexique et catégorie cognitive n'est pas travaillée puisque pas même explicitée. Cette absence de fondement ou d'analyse n'empêche pas que ces listes de mots ont été et sont encore largement utilisées en psychologie cognitive comme *outil* d'exploration des structures catégorielles, où elles constituent des « normes » permettant la construction de matériels expérimentaux paramétrés. Elles sont donc investies d'un statut d'instrument de mesure, donnant des valeurs métriques (quantitatives) à la fois à des variables indépendantes dans le paramétrage de matériel expérimental et à des variables dépendantes dans la mesure des réponses.

Parallèlement, en linguistique, l'argumentation de Berlin & Kay (1969) est fondée sur l'identification de **termes de base** pour les **catégories** de couleurs.

Un terme de base est défini comme :

« une forme qui présente minimalement les quatre caractéristiques suivantes :

- a) être monolexémique : sa signification n'est pas prédictible de la signification de ses parties
- b) sa signification n'est pas incluse dans celle de tout autre terme de couleur

c) son application ne doit pas être réduite à une classe étroite d'objets

d) il doit être psychologiquement saillant pour les informateurs. Les indices de la saillance psychologique incluent, parmi d'autres, (1) la tendance à apparaître en début de liste suscitée de termes de couleur, (2) la stabilité de référence entre informateurs et entre occasions d'usages, et (3) l'occurrence dans les idiolectes de tous les informateurs ». (p. 6).

Ces quatre principales caractéristiques (auxquelles s'ajoutent des caractéristiques mineures dont le fait de ne pas être un terme d'emprunt d'une autre langue) relèvent de plans d'analyse différents : morphologique pour les deux premières, sémantico-référentiel pour la troisième, cognitif pour la quatrième...

Ainsi, en amont du test de l'hypothèse centrale de Berlin & Kay, celle de la dépendance des formes linguistiques (les termes) vis-à-vis des structures cognitives, la diversité des modes d'inscription lexicale est éliminée, en restreignant le lexique des couleurs aux seuls termes de base, sur les critères précédemment cités, et des critères psychologiques (que l'on cherche à identifier) sont introduits dans la définition même des objets de langue. De plus, au sein même des critères linguistiques, ceux-ci, qu'ils définissent ou excluent des termes comme « termes de base », relèvent à la fois des propriétés du système de la langue (critères morphologiques) (a), de l'usage de la langue (c), et de ses variations géographiques (d) ou historiques (contact de langues et emprunts), ces différents aspects correspondant chacun à des domaines théoriques et des conceptions de la langue et du langage dont la convergence au sein d'une analyse linguistique est postulée, mais dont la compatibilité et la pertinence, vis-à-vis du fonctionnement psychologique ne sont pas explicitées.

Ainsi, ni le statut de ces listes, comme paradigme – c'est-à-dire un type de discours spécifique, résultat d'un type très particulier de production langagière –, ni comme corpus qui puisse être soumis à une analyse linguistique, ni le fait qu'il s'agisse d'un artefact cognitif, résultat de processus psychologiques complexes de rapport au monde mais aussi au langage, ne sont pris en compte, ni en psychologie, ni en linguistique.

A partir de notre travail de collecte et d'utilisation de ces listes (Dubois, 1983 ; 1986 ; Poitou & Dubois, 1999), nous tenterons dans cet article d'expliciter, à l'aide d'analyses croisées, linguistiques et psychologiques, le statut de ces « normes », afin, non pas de récuser leur pertinence opératoire, comme instrument de mesure¹ dans

¹ Nombre d'outils se sont montrés opératoires bien avant que la théorie qui les fonde ne soit élaborée.

l'exploration des catégories cognitives, mais de contribuer à identifier, dans cette situation particulière de production lexicale, la complexité de la mise en relation des mots aux choses sous la dépendance la fois de processus psychologiques individuels de catégorisation, de phénomènes lexicaux (linguistiques), et de processus psycholinguistiques d'énumération de mots. En d'autres termes, il nous a semblé que le questionnement de cet outil pouvait contribuer à identifier plus précisément les contraintes cognitives et les contraintes linguistiques, voire discursives, dans la structuration catégorielle, et du même coup clarifier la coopération pluridisciplinaire de la psychologie et de la linguistique². Nous voudrions développer ici l'idée que les déterminations psychologiques individuelles dans la dénomination et la référence d'une forme lexicale à des catégories cognitives relèvent de deux ordres de contraintes de natures différentes étroitement coopératives et qu'il est souhaitable de distinguer dans les tâches de productions lexicales de termes appartenant à différentes catégories. Nous n'entendons pas ici faire un inventaire exhaustif des processus psycholinguistiques en œuvre dans cette tâche, encore moins une théorie de la production de listes de mots ; Simplement repérer des traces de la diversité de ces ordres de contraintes en jeu dans la production de formes lexicales simples, afin de susciter de nouvelles investigations linguistiques et psychologiques qui puissent confirmer ou infirmer les hypothèses ainsi dégagées.

1. LES « NORMES » CATÉGORIELLES

La méthode de recueil généralement utilisée consiste à demander à des sujets (locuteurs natifs) de produire, à partir d'un mot inducteur qui est un terme générique catégoriel (*meubles, fleurs...*) une liste de termes faisant partie de la catégorie. Ainsi peut-on relever (bien que les consignes ne soient pas toujours explicitement rapportées dans les publications...):

“The purpose of this experiment is to find out what **items** or **objects** people commonly give as belonging to various categories or classes. The procedure will be as follow: First you will be given **the name** or **description** of a category. Then you will be given 30 sec. to write down in the notebook

² En effet, il nous semble qu'on assiste actuellement, au sein des sciences cognitives, à une opacification de la spécificité des domaines des différentes disciplines en même temps qu'à l'effacement de propriétés différentielles des objets constitués dans chacun des domaines, voire à leur dissolution, dans du cognitif sous-spécifié, qu'il est alors facilement possible de réduire à des causes premières ou à des déterminismes simplificateurs, neurophysiologiques, par exemple... (sur le « nomadisme » interdisciplinaire des concepts, voir Stengers, 1987).

as many **items** included in that category as you can. (...)” (Battig & Montague, 1969, p. 2).

“Tape recorded instructions to those reported in Battig and Montague (1969) were presented. Pilot work revealed that subjects of all ages often failed to respond appropriately after these instructions. For example, when given the category heading 'tree' they often wrote 'leaves' 'branches', and so on. Therefore, after the taped instructions the experimenter read the following additional instructions: We want you to list **members** of the category. In the example given to you, lobsters, shrimp, and so on were kinds of seafood: that is they were members of the seafood category...” (Howard, 1980, p. 226).

La notion de « norme » (à travers le nom régulièrement attribué à ces listes³) intervient dans la mesure où les résultats sont analysés après leur compilation sur un groupe relativement important de sujets et nettoyés (ou non) selon un critère de fausseté évidente ou de faible fréquence de citation :

“ All **legible** responses for each category were tallied separately (...)” (Battig & Montague, 1969, p. 2) ... even those which are **obviously inappropriate** to the category name ” (...)” (Battig & Montague, 1969, p. 3).

Il s'agit donc en fait d'une moyenne statistique considérée comme objectivation d'un fait dominant dans une communauté linguistique considérée comme culturellement homogène⁴, donnant accès à des structures cognitives (génériques, non spécifiées si ce n'est dans leur opposition aux normes associatives (Lieury, Iff, Duris, 1976, par exemple, pour le français) qui prévalaient dans la littérature psychologique associationniste sur la mémoire antérieure au développement du cognitivisme. Dans la mesure où ces listes sont investies du statut de métrique de référence⁵, elles n'ont pas été prises comme objet de recherche des structures et processus mêmes qui les produisent. En particulier elles n'ont jamais été prises au sérieux comme productions linguistiques particulières, résultant de la mise en œuvre de processus cognitifs et linguistiques (d'activités langagières) d'un type particulier. Il a en effet régulièrement été postulé que le cognitif était ainsi directement saisi (« *accédé* »

3 Cf. bilan dans Dubois, 1986: liste des « normes » publiées dans les années 1980 et leur statut dans la littérature psychologique sur les catégories sémantiques.

4 Dans cette perspective d'assigner à ces « normes » le statut d'instrument de mesure, nombre de travaux ont par ailleurs vérifié la fiabilité de telles mesures sur différents échantillons de populations (cf. Howard, 1980 par exemple).

5 C'est explicitement et modestement le cas de la plupart des recherches.

« *normalement reflété* ») au travers du (ce) linguistique sans que l'on reconnaisse à celui-ci son épaisseur (ses différentes faces, signifiants et signifiés et sa structuration spécifique) ou son rôle dans les processus mêmes de production de listes, et donc dans les résultats obtenus. Cela a trois conséquences à partir desquelles nous avons construit notre propre exploration :

1. le lexique produit dans ces listes, considéré comme ensemble d'étiquettes, n'est pas pris en considération comme objet de langue ; rare (à notre connaissance), la recherche de Segalowitz & Poulin-Dubois (1990) prend explicitement en compte une variable linguistique comme telle (la familiarité des mots utilisés y compris celle du signifiant) afin d'en évaluer les corrélations avec d'autres indicateurs des structures catégorielles (jugement de typicalité, par ex.) (*"you are to evaluate how frequently you encounter each word, in written or spoken form, as referring to a member of the category"*) ;
2. le « lexique mental » présumé comme objet psychologique de chaque sujet individuel est posé comme isomorphe au lexique collectif (non défini), ou tout au moins partagé (avec une faible marge d'erreur), par tout locuteur natif dans le cadre d'une sémantique psychologique qui laisse peu de place aux variations dialectales ou autres qui se trouvent précisément problématisées en sociolinguistique par exemple ;
3. en conséquence, les corrélations entre forme lexicale et catégorie sémantique interviennent régulièrement sur les protocoles génériques, moyennés sur l'ensemble des sujets, et l'activité des sujets comme locuteurs, même dans les recherches qui impliquent des comparaisons entre les langues, n'est pas prise en considération (cf. Lambert, Moore, 1966, par ex.).

Nous nous proposons donc ici de reprendre et poursuivre l'analyse de ces « normes » (listes) du double point de vue psychologique et linguistique ; c'est-à-dire, non seulement comme produit de processus de production verbale donnant accès ou reflétant les structures catégorielles, mais aussi comme corpus d'un type particulier suscitant une analyse spécifiquement linguistique, afin d'identifier plus précisément les contributions des structures cognitives et des propriétés de langues à ces données ou « sorties » du système cognitif humain. Du même coup, une telle analyse devrait permettre de dégager les moyens de vérifier plus finement à la fois les hypothèses psychologiques générales sur les structures catégorielles (relatives aux prototypes/stéréotypes et à la typicalité) dans la production individuelle/collective des formes lexicales.

2. UNE ANALYSE DE QUELQUES DONNÉES EMPIRIQUES

Les données résultent d'observations faites lors de la constitution de « normes catégorielles » (Dubois, 1986 ; Poitou, Dubois 1999 ; Hilaire, 2000 ; Dubois & Hilaire, 2002), utilisées par ces auteurs à des fins de recherche sur les catégories sémantiques (Dubois, 1983 ; Hampton, Dubois, 1993 ; Poitou, Dubois, 1999 ; Poitou, 2000 ; Hilaire, 2000 ; Hampton, Dubois, Yeh, soumis). Ces observations sont regroupées ici dans une visée d'analyse et de suggestions de nouvelles hypothèses sur le fonctionnement même de cette activité de production de mots, et des contraintes respectives liées tant aux structures catégorielles qu'aux contraintes linguistiques et langagières. Cette analyse bénéficie à la fois de recherches internes au français et de recherches contrastives avec l'allemand et quelques références aux normes en langue anglaise (Battig, Montague 1969 ; Hampton, Gardiner, 1983). Nous mènerons cette analyse sur trois registres que nous tenterons de spécifier précisément :

- celui des hypothèses psychologiques générales relatives aux structures catégorielles (prototype et typicalité) ;
- celui des similitudes et différences dans l'organisation des lexiques individuels et collectifs, et donc les conditions de mises en correspondance des objets psychologiques et linguistiques ;
- et donc celui des relations entre les termes des différentes listes permettant d'identifier les procédures linguistiques (contraintes des langues, et de leur morphologie lexicale en particulier), et les processus psycholinguistiques (statut et nature des traitements et relations en mémoire individuelle permettant de produire une liste de termes isolés).

2.1. Les corpus de référence

Ce travail s'inscrit en parallèle de la publication de « normes » d'« exemplaires » et de « propriétés » (Cahiers du LCPE N° 5°) utilisables pour des besoins d'expérimentations psychologiques ou d'analyses linguistiques. Le présent travail ne porte que sur les corpus relatifs aux listes d'exemplaires de catégories (une analyse linguistique et psycholinguiste approfondie des structures des propriétés est donnée dans Hilaire, 2000). Il porte plus précisément sur :

- le corpus recueilli en 80-81 sur 75 locuteurs français et ayant déjà donné lieu à une analyse (Dubois 1983 ; 1986). Ce premier corpus porte sur vingt-deux catégories (dix catégories dites « naturelles » : *animaux, arbres, boissons, fleurs, fruits, insectes, légumes, métaux, oiseaux et poissons*) et dix catégories d'artefacts (*armes, bâtiments, instruments de musique, jouets, meubles, outils, récipients, ustensiles, véhicules, et vêtements*) et deux activités, *professions* et *sports*.
- un corpus en langue allemande recueilli en 1993 auprès de 105 jeunes lycéens allemands âgés de 13-14 ans en Rhénanie-Westphalie⁶, portant sur dix des vingt-deux catégories précédentes *Bäume (arbres), Blumen (fleurs), Obst (fruits), Gemüse (légumes), Getränke (boissons), Möbel (meubles), Werkzeuge (outils), Kleidung (vêtements), Berufe (profession), Sportarten (Sports)*.

Dans chacun de ces deux cas, la consigne donnée était la suivante :

« Ecris sous le nom de la catégorie, tous les mots que tu connais dans l'ordre où ils te viennent à l'esprit ».

Des éléments de comparaison seront introduits dans la discussion à partir des « normes » en langue anglaise (Battig & Montague et normes non publiées) élaborées en 1998 pour une recherche en collaboration avec Larry Barsalou, James Hampton (Hampton, Dubois & Yeh, soumis) et des « normes associatives » (non publiées) élaborées en relation avec le travail de Frédéric Lavigne (Lavigne, 1993 ; Dubois & Lavigne, 1994).

⁶ Cette expérience a été menée par Peter Scherfer (université de Wuppertal). Qu'il veuille bien trouver ici l'expression de nos plus vifs remerciements. Nos remerciements aussi à Céline Demit pour son aide au décodage d'un certain nombre de noms de boissons ...

2.2. Quelques repères

L'analyse qui suit vise principalement à repérer des points sensibles pour l'identification du jeu des contraintes cognitives et/ou linguistiques dans la production de ces listes de mots, à partir principalement des décalages entre différents paramètres dans les listes françaises ou entre les langues (français, allemand, anglais) relevant de cultures très proches (et trop souvent confondues !) bien que pouvant susciter, sur certaines catégories, des expériences diversifiées. En d'autres termes, il ne s'agit pas de prouver que les contraintes cognitives influent la production de listes de mots, mais d'examiner comment, quels aspects de la structure des langues interviennent, par la mise en œuvre de processus psycholinguistiques dont il restera à préciser la généralité de fonctionnement dans d'autres situations langagières (discours, textes, dialogues, etc). En outre, l'analyse ne portera que sur quelques concepts centraux de la théorie des prototypes en s'efforçant de repérer en chacun systématiquement, soit des traces du fonctionnement cognitif, soit des faits de langue ou de production langagière. Ainsi, les catégories et les prototypes, traités d'évidence comme des structures cognitives seront questionnés en tant que **termes inducteurs** (1), ou **item lexical disponible et saillant** en langue et candidat prototype (2). De même, la notion de typicalité identifiée à partir de la distribution des exemplaires pourra être ainsi (re)considérée comme le résultat d'une **énumération**, comme type de **discours** très particulier, sans énonciation de marques syntaxiques (3).

3. LES TERMES INDUCTEURS : DES CATÉGORIES ?

3.1. Statut du terme inducteur

On peut d'abord s'interroger sur la signification des termes inducteurs donnés aux sujets et à partir desquels ils produisent les listes de termes pour chaque catégorie. Dans les premiers travaux sur la structure des catégories, ces termes étaient considérés comme relevant d'un même niveau catégoriel, le niveau de base, défini sur le plan cognitif comme correspondant à un optimum cognitif, celui où,

“ (1) in the perceived world, information-rich bundles of perceptual and functional attributes occur that form natural discontinuities, and that (2) basic cuts in categorization are made at these discontinuities ” (Rosch, 1978 ; p. 31).

La conséquence en langue de cette hypothèse de niveau de base est que :

“ we would expect that the most useful and, thus, the most used name for an item to be the basic-level name ” (ibid. p. 35).

Mais cette hypothèse est-elle justifiée sur le plan linguistique ? Ces termes inducteurs renvoient-ils aux mêmes pratiques

langagières ? La question prend toute son importance si l'on considère les listes de termes comme de véritables objets linguistiques, c'est-à-dire comme des textes (certes spécifiques) appelés par un terme, comme des réponses apportées à une question posée par l'expérimentateur et exprimée dans le terme inducteur. Ainsi recadrés dans des pratiques langagières, il apparaît que les termes inducteurs n'ont pas tous le même statut. Dans un contexte discursif plus ordinaire, si l'on essaye, par exemple, d'inventorier les questions dans lesquelles peuvent être insérés les termes " *sports* " et « *boissons* », la différence est frappante : on demande : « *quels sports pratiques-tu habituellement ?* », mais pas « *quelles boissons bois-tu habituellement ?* » De même, on peut dire « *on fait du sport* », mais « *on boit une boisson* » est déviant (*on boit du vin, du coca..*).

Plus généralement, ces différences tiennent à la spécificité de l'organisation linguistique du lexique, aux relations multiples (en particulier syntaxiques) qui la sous-tendent et à l'utilisation du lexique dans les échanges langagiers. Cette structuration ne se réduit pas à un espace bidimensionnel, avec un axe vertical correspondant à des niveaux différents de généralité/inclusion et un axe horizontal correspondant à des relations d'équivalence ou de typicalité comme élément au sein d'une même classe/catégorie (cf. Rosch, 1978 ; Rastier, 1991). Que ce soit sur la « dimension » verticale ou horizontale, on connaît depuis longtemps, en linguistique et en anthropologie, les *lexical gaps*, les trous dans le lexique du point de vue d'une hypothétique organisation hiérarchique (ainsi, il n'existe pas, ni en français ni en allemand, de terme générique simple pour la catégorie dont stylo, crayon, feutre, etc. seraient les exemplaires). Le lexique n'est donc pas structuré comme les nomenclatures linnéennes⁷, comme le suggèreraient les termes utilisés en biologie pour catégoriser les végétaux et les animaux. La question de la pertinence des niveaux de généralité en langue (Kleiber, 1997b) et les liens avec la construction cognitive et historique des nomenclatures et taxinomies devrait constituer l'objet d'une analyse spécifiquement linguistique (cf. Selosse, 2000 à ce sujet).

En outre, il n'est pas davantage évident que les structures du lexique soient universelles. L'hypothèse d'une diversité et d'une spécificité de chaque langue paraît autrement plus plausible. Ainsi, pour reprendre l'exemple des boissons, le mot de « *boisson* » entre, en français, dans deux expressions, « *distributeur de boissons* » et

7 On utilise les classifications linéennes comme parangon de l'organisation cognitive et lexicale en faisant d'ailleurs fi d'une analyse cognitive et linguistique de cet objet historique, pour lequel on a pu montrer (Selosse, 2000) qu'il s'agissait de fait de l'aboutissement d'un long (sur plusieurs siècles) processus de **réduction** et de systématisation tant de la connaissance que des **discours de description** des plantes collectées par les herboristes.

« *débit de boissons* », le second n'étant guère utilisé dans l'expression de la vie quotidienne. Mais un distributeur de boissons ne fournit pas toutes les boissons (pas de Château-Margaux en distributeur !), le terme a donc dans cet usage concret une extension référentielle et un niveau de généralité moindre que dans « *débit de boissons* ». En allemand, le terme « *Getränk* », ou son pluriel « *Getränke* », apparaît dans au moins trois composés courants dont l'un seulement a un correspondant direct en français, *Getränkeautomat* (distributeur), et deux autres non : *Getränkekarte* (carte des boissons ; en français, on a « *la carte des vins* ») et *Getränkemarkt* (magasin spécialisé dans la vente de boissons en tous genres, alcoolisées ou non). En France, ce genre d'établissement commercial n'est (pour le moins) pas courant, et dans les supermarchés, il n'y a pas de rayon « *boissons* », mais des rayons spécialisés pour différentes sous-catégories (vins, apéritifs, eaux, etc.). C'est dire que les termes « *boissons* » et « *Getränke* », qui pourraient, dans le cadre d'une simple nomenclature, correspondre au même niveau de généralité et auraient, dans ce cadre, la même extension référentielle, ne sont pas équivalents sur le plan intralexical. Il est également notable que dans les « normes catégorielles » de référence que constitue le travail de Battig & Montague (1969), parmi les catégories standard (*color, flower, insect, ...*), il y ait « *alcoholic beverage* » et « *non alcoholic beverage* », sans justification particulière.

Plus généralement, mais pour les mêmes raisons, la distinction, posée dans les premiers travaux sur la catégorisation, entre catégories naturelles et catégories artefactuelles ne peut être transférée directement sur le plan linguistique. Ainsi, on a pu noter (cf. Dubois, 1983 ; 1986) des résultats quantitatifs différents concernant des listes catégorielles produites à partir de noms de catégories artefactuelles et naturelles en ce qui concerne le nombre de termes cités et la fréquence de citation, plus importants dans le cas des catégories artefactuelles (cf. par ex. la différence entre la catégorie des professions et celles des fleurs). Mais les différences entre ces deux catégories ne sont pas systématiques. Elles sont davantage à expliquer comme résultat de processus complexes que comme reflet d'une différence dans la nature intrinsèque des catégories (Dubois & Resche-Rigon, 1995). Il faut les relier à des expériences différentes du monde et à différentes pratiques langagières des sujets à propos de ces domaines de référence. En particulier, on peut faire l'hypothèse que dans l'expérience des catégories naturelles peut entrer, plus que dans les catégories artefactuelles, un savoir encyclopédique enseigné à l'école. Ainsi, ces catégories seraient-elles aussi davantage construites en discours qu'en relation avec une expérience perceptible. Par exemple, les discours sur la baleine et le problème de sa catégorisation ne reflètent généralement qu'une expérience indirecte, iconographique, de ces animaux ou un savoir encyclopédique enseigné de façon systématique dans le cadre de cours de sciences naturelles. A

l'inverse, les discours sur les professions ou les boissons – dont l'analyse n'entre guère dans les programmes scolaires – peuvent être plus directement reliés au vécu quotidien des sujets testés, et donc à d'autres formes de présentation (y compris discursives) des formes lexicales.

Enfin, toujours sur le plan linguistique, les termes inducteurs imposés aux sujets n'ont pas tous le même statut morphologique et sémantique, et celui-ci est variable selon les langues. Ainsi, le mot anglais *furniture* est un collectif massif, en français, *meuble* est un nom comptable et en allemand, *Möbel* est le plus souvent employé au pluriel (avec valeur collective). Au sein d'une même langue, les différences sont aussi nettes : *fruit* s'emploie aussi bien au singulier qu'au pluriel (« *je voudrais un fruit, j'ai acheté des fruits* »), alors que *légume* s'emploie couramment surtout au pluriel. En allemand, *Gemüse* (légumes) est un collectif (son emploi comme nom comptable nécessite son insertion dans un composé : *Gemüsesorte(n)*, de même *Obst* (fruit), tandis que *Frucht* (fruit) est comptable. Certes, cette différence de statut linguistique n'implique pas que leur fonctionnement comme déclencheur de listes soit différent. Au moins la question des effets psycholinguistiques de ces propriétés linguistiques mériterait-elle d'être étudiée.

En bref, alors que les catégories sont données *a priori* et que les termes inducteurs sont considérés comme équivalents entre eux au sein d'une même langue et entre les langues, reflétant un niveau d'organisation catégorielle équivalent, ces quelques exemples conduisent à problématiser le statut de ces termes inducteurs au sein du système linguistique (en particulier du système lexical) et susciter des hypothèses sur les dimensions et caractéristiques spécifiques des langues (telles que massif vs comptable, singulier vs pluriel ... (cf. Wierzbicka, 1990, sur les termes de couleur) qui contraignent leur rôle dans les activités psycholinguistiques de production de listes de termes.

3.2. Des consignes : accès aux représentations ou construction d'un discours ?

Parmi les contraintes d'induction de l'activité de production, il convient également de faire mention du rôle des consignes et formulations dans lesquelles se trouvent inscrits les termes catégoriels. Ces consignes sont très souvent omises dans les publications des résultats d'épreuves de production de listes. On prendra simplement quelques exemples qui semblent intéressants pour évaluer le rôle des consignes dans la production de listes de mots et leurs conséquences sur la structure même des productions qui en résultent, c'est-à-dire une interprétation de ces listes comme produits de pratiques langagières diversement induites par les consignes.

« *Normes catégorielles* » vs « *normes associatives* »

Par exemple, si on observe le contraste en français entre les résultats des deux consignes qui utilisent les mêmes termes génériques, celles qui suscitent des « normes associatives » (fréquemment utilisées dans les années béhavioristes de la psychologie, Jodelet, 1965), et celles des « normes catégorielles » développées dans la mouvance cognitive, le paysage qui en découle est très différent (cf. tableau 1, ci-dessous).

ordre de fréquences	consigne CAT	consigne ASS
1	chêne	forêt
2	pommier	bois
3	sapin	racines
4	cerisier	feuille(s)
5	poirier	branche(s)
6	hêtre	vert
7	bouleau	tronc
8	pin	nature
9	peuplier	sève
10	prunier	oxygène
11	saule	vie
12	érable	force
13	oranger	fruit

Tableau 1 : Items lexicaux les plus fréquemment cités selon deux types de consignes (catégorielles CAT vs associatives ASS)

Ainsi, à partir du même terme inducteur *arbre*, par exemple, la consigne Associative (ASS) :

« Citez les mots qui vous viennent le plus rapidement à l'esprit à partir de chacun des termes écrits en haut de chacune des pages de votre livret. »

conduit à une liste de mots de diverses catégories syntaxiques (massivement des substantifs mais aussi des adjectifs), et sémantiquement, à des mots référant à des classes collectives (*forêt*), des « parties de » (*racines, feuilles, branches...*), des matières (*bois*), des propriétés (*vert*), et des associés thématiques abstraits (*nature, vie...*) (Dubois, Lavigne, 1994).

Ce résultat contraste avec les consignes « catégorielles » qui demandent :

« d'énumérer, dans l'ordre où ils leur venaient à l'esprit, les noms des objets appartenant aux diverses catégories données en tête de chaque feuillet »,

et qui conduisent quasi-exclusivement à des substantifs (dans le cas des catégories nominales classiquement étudiées) et sémantiquement (et comme il est certes demandé...) à des exemplaires de la classe (cf. plus loin, Tableau 6).

Ce résultat peut être considéré comme trivial, puisqu'on obtient ce que l'on demande. Cependant, là encore, le reflet de cette flexibilité des processus psychologiques d'accès aux structures de la mémoire qui se révèlent très sensibles à l'orientation de la tâche à partir d'un même mot inducteur n'est pas problématisée. Et en tout cas, pas comme le résultat de processus lexico-discursifs et psycholinguistiques qui réguleraient des processus cognitifs. En d'autres termes, la dimension discursive et langagière de ces deux types de listes pourrait être prise en compte, et cette absence de prise en compte reflèterait simplement le présumé implicite d'une indépendance des items lexicaux vis-à-vis de leur inscription discursive, fondée sur une image des structures du lexique qu'il s'agit précisément de tester.

On notera que l'hypothèse d'un lexique comme ensemble de formes indépendantes des registres textuels où ils font sens, en accord avec la conception à la fois dictionnaire et nomenclaturale du lexique qui prévaut dans le sens commun de notre culture (et donc celle des sujets sollicités), autorise par ailleurs le traitement collectif des données. En contraste, lorsque cette conception du lexique n'est pas aussi intériorisée ni donc partagée par les sujets interrogés et l'expérimentateur, lors de recherches dans des cultures plus distantes des nôtres⁸, la consigne catégorielle peut être jugée insolite et requérir de l'expérimentateur une reformulation plus explicite sur ce qu'il entend obtenir. En témoigne, dans l'une des premières expériences de Rosch (Rosch, 1973), la difficulté de certains sujets à répondre correctement, difficulté qui devrait être l'objet d'une interrogation sur la pertinence culturelle des consignes plutôt que d'être ignoré⁹.

8 et qui, en raison d'une absence de tradition écrite, livresque et donc dictionnaire par exemple, ne partage donc pas la même histoire des théories sur le langage, ne peuvent mettre en œuvre les activités psychologiques métalinguistiques qui sont disponibles dans nos cultures pour permettre la production spontanée de telles listes.

9 tout comme lors du traitement des réponses, l'élimination des données non conformes: "Two judges reviewed the resulting tables and indicated cases in which an attribute was clearly and obviously false" "Then attributes were deleted from the tabulations" from Rosch et Mervis 1975 ; p. 578.

“ Instructions were [...] to list all the objects you can think of. Try not to just free associate - for example, if bicycles just *happened to remind you of your father, don't write down father* ”.

Quelle que soit cependant notre ignorance des divers processus psychologiques que peuvent susciter ces consignes, les listes résultantes peuvent être considérées comme un corpus, un texte d'un type particulier et faire l'objet d'une analyse linguistique susceptible d'informer sur les propriétés de l'organisation lexicale à la fois comme objet de langue et comme accès aux structures linguistiques intériorisées par des locuteurs et mises en œuvre dans l'exercice très particulier de production de listes de mots selon une diversité de consignes, qui peuvent être différemment interprétées et différemment maîtrisées par différents locuteurs. Les quelques paragraphes suivants suggèrent de telles hypothèses d'analyse de ces « normes », à partir des principaux concepts psychologiques relatifs aux structures catégorielles qu'elles sont censées mesurer.

4. LE PREMIER TERME DE LA LISTE : PROTOTYPE OU DISPONIBILITÉ LEXICALE ?

Le premier terme cité (notion ambiguë qui recoupe à la fois le terme cité en premier par chacun des sujets interrogés et le terme le plus fréquemment cité apparaissant en premier dans la liste moyenne calculé sur l'ensemble des sujets) est utilisé pour identifier le prototype de chaque catégorie. Même si on peut mettre en évidence de fortes corrélations entre ces deux mesures de la prototypie (Battig & Montague, 1969 ; Hampton & Gardiner, 1983 ; Dubois, 1983 ; Poitou, 2000), il est difficile de considérer comme homogènes tous les exemplaires cités en premier à travers la diversité des scores selon les différentes catégories, et il pourrait s'avérer intéressant de pointer davantage sur les distorsions et décalages que sur ces corrélations entre ces deux indices : la fréquence de citation par l'échantillon de sujets et le rang moyen.

Ainsi, par exemple, parmi les fruits, l'avocat qui n'atteint qu'une fréquence de citation de 15 sur les 75 sujets interrogés et se situe ainsi sur la base d'un classement sur la fréquence au 23^e rang est donné avec un score de rang moyen de 7,87 qui le situe cette fois au 9^e rang sur cet indicateur (cf. Dubois, 1986) ; cela signifie non seulement qu'il existe d'importantes différences inter-individuelles sur les représentations et dans la gestion de la tâche, mais aussi que d'autres facteurs, et en particulier des dimensions linguistiques des mots désignant ces prototypes, peuvent intervenir.

4.1. Prototype : normativité de la représentation ou disponibilité lexicale ?

Si on analyse plus précisément quelques-uns de ces hiatus entre les indices globaux que sont la fréquence de citation (considérée

comme le pourcentage global de sujets ayant produit le terme, quel que soit son rang de citation) et le rang de citation (le rang *moyen* dans lequel chacun des sujets a produit le terme dans sa production individuelle), on peut en effet introduire quelques hypothèses sur la nature et le statut de ces différents indicateurs.

On prendra, à titre d'exemple, les données des listes allemandes (Poitou, Dubois 1999, Poitou, 2000)¹⁰, telles que représentées dans le tableau 2 ci-dessous.

	Fréquence de citation	Rang moyen de citation
Cola	98,1 % (1)	2,8
Bier [bière]	84,6 % (2)	7,1
Wasser [eau]	82,7% (3)	5,4
Fanta	67,3% (4)	3,7
Sprite	61,5% (5)	5,7
Tee(arten) [thé, tisane]	59,6 % (6)	9,9
Wein [vin]	59,6% (6)	9,6
...		
Red Bull	26,9% (18)	7,1

Tableau 2 : Fréquence de citation et rang moyen pour les items les plus fréquents de la catégorie Getränke

On peut ainsi remarquer que les boissons froides non alcoolisées ont un rang de citation inférieur aux boissons alcoolisées ou chaudes (voir notamment les cas opposés de *Bier*, massivement cité, mais relativement tard dans les listes et de *Red Bull* (boisson gazeuse tonique et euphorisante), peu cité, mais relativement haut dans les listes). Cette différence de saillance pour les sujets, que l'on peut sans trop de risque imputer à un biais d'échantillonnage (nos sujets allemands étaient des lycéens de 13-14 ans), n'en manifeste pas moins le fait que ces listes sont des objets sociolinguistiques sensibles à l'échantillonnage de la population de référence, tant sur le plan de leur expérience (le type de boisson qu'ils boivent...) que sur celui de l'usage qu'ils font de la dénomination.

Cela rejoint l'argumentation que nous avons déjà mentionnée (Poitou, 2000) sur la double imputation de la saillance des items lexicaux candidat prototypes : comme reflet d'une **expérience** partagée par les différentes communautés interrogées, expérience à la

¹⁰ Les données allemandes sont disponibles sur le site Internet <http://nte.univ-lyon2.fr/~poitou>

fois pratique et quotidienne et discursive des lectures de matériaux linguistiques. Dans aucun des cas, cependant, ces valeurs collectives ne sont directement imputables à des représentations perceptives singulières, mais elles sont davantage liées à des pratiques et connaissances partagées y compris les pratiques langagières (cf. Dubois & Resche-Rigon, 1993 ; Dubois, 1993).

L'identification même du ou des prototypes, comme exemplaire(s) représentatif(s) de la catégorie, effectuée sur la base de la fréquence de citation, demeure elle aussi problématique quant à son statut de représentation cognitive partagée ou de phénomène lexical. Si on reprend les termes cités par 75 % des sujets comme base de référence (largement utilisée sans pour autant être justifiée, cf. par ex. Dubois 1983), on peut ainsi observer des contrastes entre l'allemand, le français et l'anglais (Battig & Montague, 1969), qui interviennent différenciellement selon les catégories. (cf. tableau 3 ci-dessous) :

Werkzeuge	Sportarten	Blumen	Möbel	Obst
Hammer	Fussball	Rose	Stuhl Schrank Tisch Bett	Apfel Banane
carpenter's tool	sports	flower	article of furniture	fruit
hammer (saw)	football baseball basketball	Rose	chair table (bed) (sofa)	apple orange
outils	sports	fleurs	meubles	fruits
marteau	football	Rose	chaise table armoire lit	pomme poire orange cerise

Gemüse	Berufe	Bäume	Kleidung	Getränke
-	-	-	T-Shirt Hose	Cola
vegetable	profession	tree	article of clothing	beverage
-	doctor	oak (maple)	shirt socks	-
légumes	professions	arbres	vêtements	boissons
carottes poireau pomme de terre	-	chêne pommier	pull-over pantalon robe chaussette manteau	eau

Tableau 3 : items cités par plus de 75 % des sujets dans trois langues.

On observe ainsi que si, pour les légumes, ni les listes allemandes, ni les listes anglaises ne permettent d'identifier un prototype (défini comme item cité par au moins 75 % des sujets), les listes françaises quant à elles en fourniraient trois sur le même critère, à savoir *carotte*, *poireau*, *pomme de terre*, qui relèvent davantage de la pratique du pot-au-feu (Rastier, 1991) ou d'un répertoire de ritournelles (*des pommes*, *des poires* et *des ...*) que d'une hypothétique similitude perceptive permettant l'émergence d'un prototype structurant l'espace catégoriel.

Si on étend l'exploration des candidats prototypes au-delà du critère de 75 % des réponses, la comparaison des dix premiers exemplaires (des catégories des vêtements en français, allemand et anglais, pris à titre d'exemple, cf. tableau 4, ci-dessous) permet de développer à nouveau une argumentation qui suggère la pluralité des processus à la fois cognitifs, expérientiels et langagiers dans les résultats de ces listes catégorielles.

<i>ordre sur les fréquences</i>	<i>vêtements</i> F /75	<i>Kleidung</i> D/ 53	<i>articles of clothing</i> E/ 442
1	pull-over (67)	T-Shirt (46)	shirt (352)
2	pantalon (66)	Hose (40)	socks (330)
3	robe (60)	Hemd (38)	pants (318)
4	chaussettes (60)	Pull-over (35)	shoes (274)
5	manteau (60)	Schuhe (34)	blouse (261)
6	jupe (55)	Unterhose (31)	skirt (261)
7	chemise (54)	Jeans (31)	coat (260)
8	veste (43)	Socken (30)	dress (240)
9	slip (31)	Jacke (27)	hat (201)
10	chemisier (30)	Bluse (25)	sweater(163)

Tableau 4 : Inventaire des dix items les plus fréquemment cités pour la catégorie des vêtements, (le nombre entre parenthèses indique le nombre de sujets et/ou de citations)

Pour cette catégorie, si la structure catégorielle interne apparaît très semblable à partir des fréquences de citation, les différents items sont cependant ordonnés différemment. Peut-on parler de prototypes alors que la distribution des exemplaires laisse plutôt penser à un inventaire de la garde-robe minimale dans nos cultures occidentales : les vêtements du haut, du bas, de dessus, de dessous ? L'ordre des différents items dépendant ainsi d'autres facteurs que la représentativité comme élément de la classe, des facteurs à la fois cognitifs (davantage expérimentiels que perceptifs) et langagiers, liés tant à la stratégie de déclinaison d'une séquence de mots qu'à des phénomènes de langue telle la disponibilité lexicale et certaines propriétés morphologiques, non partagées entre le français et l'allemand par exemple.

4.2. Prototype : Représentation saillante ou inventaire lexical ?

En effet, si on considère, dans la catégorie allemande des vêtements (Kleidung), les deux sous-catégories des vêtements de dessus pour le haut et le bas, on constate que le terme Hose est lexicalement saillant en raison (1) de sa généralité (il peut désigner pratiquement tout vêtement du bas), (2) de la faiblesse de l'inventaire lexical dans ce domaine (il est pratiquement seul et donc un bon candidat parce qu'il n'y en a guère d'autres), (3) de sa simplicité par rapport à beaucoup d'autres termes en -hose. Mais il est d'abord nécessaire de prendre ici encore une diversité d'objets en œuvre dans la notion même de disponibilité lexicale. Il faut en effet distinguer (cf. Poitou, 2000) :

- l'inventaire global (ou supposé tel) de la communauté linguistique (les mots des dictionnaires) ;
- l'inventaire global tel qu'il apparaît dans l'expérience auprès des sujets testés, considérés comme un groupe homogène ; cet inventaire correspond aux termes les plus disponibles pour ce groupe ; sans la limitation temporelle de l'expérience, il pourrait être plus étendu ;
- l'inventaire des termes cités par chaque sujet (liste ordonnée).

Si l'on prend l'exemple des vêtements haut/bas et dessus/dessous, à partir des données relatives aux termes cités en premier (entre les rangs 1 à 5 par plus de 1 sujet, cf. tableau 5 ci dessous),

	<i>Tête</i>	<i>Haut</i>	<i>Bas</i>	<i>Pieds</i>	<i>problèmes</i>
<i>Dessus</i>	Hut (3) Mütze (2)	Pullover (33) T-Shirt (32) Sweatshirt (8) Bluse (8) Jacke (5) Schlips (2)	Hose (41) Jeans (21) Short (3)	Schuhe (6)	Rock (8) Body (2) Unterwäsche (2) Kleid (2)
<i>% citation</i>	2,5	14,66	21,66	6	
<i>Dessous</i>		BH (3)	Unterhose (15) Slip (3) Stapse (2)	Socken (11) Strümpfe (6) Strumpfhose (2)	
<i>% citation</i>		3	6,66	6,66	

Tableau 5 : termes référant aux vêtements « du haut » et « du bas », « de dessus » et « de dessous », ainsi que leurs fréquences de citation.

on observe à la fois nettement plus de termes pour « le haut », ce qui n'a rien d'étonnant puisque l'inventaire disponible est plus grand, mais aussi, et surtout en premier rang, une nette dominance de termes pour les vêtements du bas (34) par rapport aux termes pour les vêtements du haut (11), ce qui n'est pas étonnant du fait de leur saillance dans l'inventaire lexical¹¹.

Les analyses quantitatives, qui poseraient ici *Hose* comme prototype des vêtements, sans relier sa fréquence de production dans les listes à l'analyse de l'inventaire lexical conduit ainsi à des contradictions. Du fait que *Hose* est plus cité que *Pullover*, on ne peut pas déduire que *Hose* est plus typique que *Pullover*. *Hose* est un terme plus générique (et morphologiquement simple alors que *Pullover* est construit), et en outre les instanciations des éléments de

11 Dans l'inventaire global cité par tous les sujets, on trouve pour le bas dessus (entre parenthèses le nombre de citations) 16 types (soient 12,7 % de l'ensemble des termes cités) et 123 tokens (15,9 %) :

– Hose et composés (68) : Hose (45), **Radlerhose** (8), Kordhose (4), Badehose (2), Turnhose (1), Reithose (1), Leinenhose (1), Latzhose (1), **Joggingshose** (1), **Jeanshose** (1)
– autres (55) : Jeans (31), Short (7), Leggings (9), Boxershorts (5), Minirock (2), Bermuda (1)

la sous-catégorie *Hose* n'ont pas de représentations lexicales aussi simples morphologiquement que *Hose*. Ce fait peut expliquer pourquoi *Hose* apparaît massivement en rang 1. En bref, dire que *Hose* est le prototype de la catégorie n'a donc pas de sens dans une analyse des productions lexicales prises comme simples dénominations d'objets du monde, et sans description de son fonctionnement comme forme lexicale d'une langue.

Pour une catégorie donnée, la typicalité (comme simple description des résultats des listes catégorielles ou saillance, Poitou, 2000) dépend donc (au moins) :

- a) de données expérientielles (au sens large et sans parler de son statut de représentation individuelle ou collective) concernant les référents des termes de la catégorie ;
- b) de la nature du terme lui-même (propriétés linguistiques) ;
- c) de l'inventaire lexical disponible. Il faudrait y ajouter (surtout si l'on situe ces expériences dans le cadre de l'acquisition des catégories) la mémorisation de ces différents termes.

Les variables langagières et lexicales contribuent donc largement à l'émergence d'un terme en tête de liste. Ce fait devrait conduire à la prudence quant à l'hypothèse une relation directe entre le rang et la fréquence de citation dans de telles listes et le statut de prototype comme représentation cognitive représentative de la catégorie.

Si on se réfère cette fois à l'exemple des arbres, il s'avère à nouveau difficile d'interpréter les différences d'ordre permettant d'identifier le ou les prototypes d'arbres, sans faire appel à une diversité de processus incluant en particulier des processus lexicaux.

ARBRES (F)	BAUM (D)	TREE (E)
Chêne	Eiche	oak
Pommier	Tanne	maple
Sapin	Fichte	pine
Cerisier	Birke	elm
Poirier	Buche	apple
Hêtre	Kastanie	birch
Bouleau	Apfelbaum	cherry
Pin	Ahorn	dogwood
Peuplier	Birnenbaum	spruce
Prunier	Kirschbaum	redwood
Saule	Pappel	peach
Erable	Blätter	weeping willow
oranger	Kiefer	walnut

Tableau 6 : Citations des items les plus fréquents de la catégorie « arbre » dans trois langues

Si *chêne* semble bien émerger, est-ce vraiment en raison de sa saillance perceptive dans l'expérience sensible des sujets interrogés ? En effet pourquoi le chêne serait-il plus présent dans la représentation mentale construite à partir de la perception que le pin, le bouleau, ou le pin que le sapin ? Là encore, on peut sans trop de risque émettre l'hypothèse de différences dans la disponibilité lexicale liée au mode d'acquisition des termes de base, sans doute plus livresque que perceptive, donnant du même coup à la représentation cognitive davantage le statut de représentation stéréotypique, y compris dans sa dimension iconographique, emblématique de la culture (Rastier, 1991) (le chêne, le roi des arbres ; le chêne et le roseau, Saint-Louis sous son chêne etc ...) que celle d'une image mentale issue de la perception individuelle.

Cependant, qu'il s'agisse d'un prototype perceptif individuel ou d'une image stéréotypique (cf. Dubois & Resche-Rigon, 1993), des processus cognitifs purs ne peuvent rendre compte des variations d'ordre, et l'exemple des arbres peut nous permettre de suggérer des hypothèses relevant de la prise en compte de processus discursifs dans la production de ces listes de mots. En effet, si on prend en compte la répartition des termes cités entre les premiers rangs, celle-ci est en partie aléatoire et imputable à la **contrainte de linéarité** inhérente aux listes à produire. En effet, un sujet chez qui le terme inducteur déclenche deux termes ne peut que les produire en un rang différent. L'étude des rangs de citation de *Tanne (sapin)* et *Fichte (épicéa)* dans les listes d'arbres peut servir d'illustration à ce processus.

	Rang 1	Rang 2	Rang 3	Rang 4	Rang 5
Fichte	7	5	10	1	3
Tanne	2	8	6	6	3

Tableau 7a : nombre de sujets (sur 53) citant Tanne et Fichte (5 premiers rangs)

	Rang 1	Rang 2	Rang 3	Rang 4	Rang 5
Fichte	13,21	22,64	41,51	43,40	49,06
Tanne	3,77	18,87	30,19	41,51	47,17
Ecart	9,44	3,77	11,32	1,89	1,89

Tableau 7b : pourcentages cumulés de sujets ayant cité Tanne et Fichte selon le rang

Sur les cinq premiers rangs (nombre de citations cumulées), *Tanne* et *Fichte* font pratiquement jeu égal (cf. tableau 7), mais la répartition de ces deux termes sur chaque rang révèle des variations importantes. *Fichte* domine aux rangs 1 et 3, tandis que *Tanne* l'emporte aux rangs 2 et 4... Et la prise en compte de *Tannenbaum* (de même valeur référentielle que *Tanne*) modifierait l'équilibre global pour les rangs 1-5 au profit de *Tanne/Tannenbaum*... C'est dire que les études quantitatives à la virgule près sont en l'occurrence peu pertinentes, la mesure de l'accessibilité ne peut être que grossière.

5. LA DISTRIBUTION DES TERMES : TYPICALITÉ DES EXEMPLAIRES ?

Les différences dans la fréquence de citation des différents termes ont été majoritairement utilisées comme outil, comme une métrique de la structure interne des catégories et de la typicalité des différents exemplaires au sein de ces catégories. Ces différences sont censées en effet refléter la structure graduelle de la catégorie en tant que structure des représentations mentales en mémoire. Quelques travaux ont certes étudié ces listes pour elles-mêmes, dans le but d'améliorer l'outil (« normes ») en isolant plus précisément les variables confondues dans ces listes. Ainsi en est-il des études de corrélations entre facteurs de la représentativité des catégories, la fréquence de production et le jugement de typicalité, la familiarité de la chose référée, ou du mot, les fréquences associatives (Mervis, Catlin & Rosch, 1976 ; Segalowitz & Poulin-Dubois, 1990 ; Hampton et Gardiner, 1983). Jamais cependant, à notre connaissance, la typicalité issue des listes catégorielles n'a été mentionnée ou n'a suscité l'hypothèse qu'il puisse s'agir également d'un fait de langue (hormis une corrélation avec la fréquence du mot, valeur dont on connaît la difficulté à l'établir) et encore moins d'un phénomène de discours. Seule à nouveau, la recherche de Segalowitz & Poulin-Dubois (1990) mentionne que :

« la structure interne des catégories peut, à l'occasion, être attribuable à des facteurs autres que la similarité perceptive entre les exemplaires de la catégorie. Plus spécifiquement, certaines différences inter linguistiques dans le degré de typicalité peuvent, correspondre à des différences interlinguistiques de familiarité linguistique » (p. 508).

Nous donnerons ici quelques indications qui suggèrent précisément que ces listes peuvent être considérées de manière productive comme des séquences discursives, et susciter ainsi des hypothèses sur le jeu complexe des structures cognitives (de la typicalité comme variable psychologique) et des structures lexicales et langagières.

5.1. Rang et dispersion des exemplaires cités : typicalité ou divergence discursive ?

Si on calcule (cf. tableau 8 ci-dessous), pour toutes les catégories, le nombre de termes cités à chacun des rangs, on constate une progression régulière du nombre de termes différents cités en fonction du rang de citation (calcul effectué sur les cinq premiers rangs). Ce qui n'a rien d'étonnant, certes, mais quelle est la nature de ce phénomène : est-il psychologique et lié aux structures des représentations en mémoire, ou psycholinguistique sous l'influence des contraintes de production de mots isolés en séquence ?

Rangs	Blumen fleurs	Sportarten sports	Werkzeuge outils	Kleidung vêtements	Obst fruits	Getränke boissons	Gemüse légumes	Berufe professions	Bäume arbres
1	18	19	7	12	17	10	16	27	17
2	23	18	12	15	13	15	18	31	16
3	28	21	17	20	20	17	25	38	19
4	38	17	21	21	25	24	25	38	28
5	38	26	24	26	32	28	25	35	26

Tableau 8 : Nombre de termes cités pour chacun des cinq premiers rangs

Si on tente de systématiser la seconde hypothèse, on peut considérer que, au niveau de chaque protocole individuel :

- a) chaque terme inducteur T_0 déclenche chez le sujet une « association » avec un terme (T_1) ;
- b) la production du terme T_2 peut être alors :
 - soit demeurer connectée au terme inducteur (ce qui serait le pur respect de la consigne et n'interviendrait que sous l'hypothèse de l'indépendance linguistique et psycholinguistique des termes dans le processus séquentiel de production) ;
 - soit au terme T_1 , qui interférerait dans la déclinaison du paradigme, ;

- de même pour T₃, qui pourrait être relié soit au terme T₂, soit au terme T₁, soit au terme inducteur (cf. figure 1 ci-dessous).

Différents chemins possibles à partir du même terme inducteur T₀

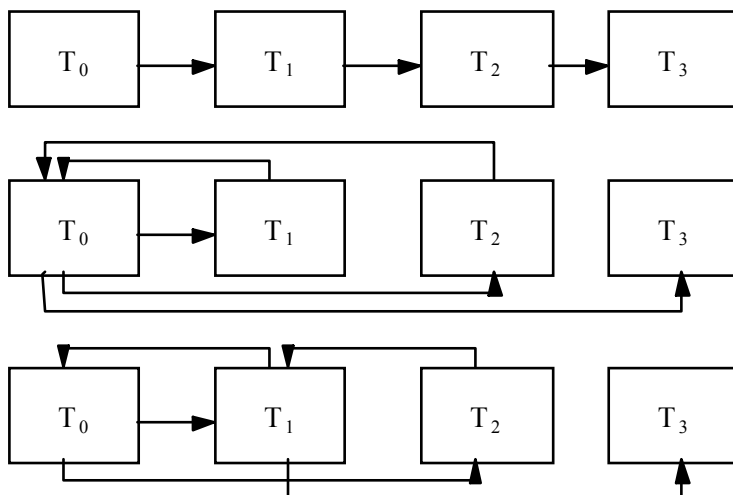


Figure 1 : Exemples de passages directs plausibles de T₁ à T₂ sans retour à T₀.

La dispersion croissante (= le nombre croissant de termes au fur et à mesure que le rang de citation augmente, c'est-à-dire que l'on s'éloigne du terme inducteur) peut donc être analysée comme le résultat de ces possibilités différentes d'association, là encore sous l'effet de contraintes diverses (cognitives **et langagières**), plus précisément analysées dans Poitou, Dubois 1999, et dont nous ne soulignerons ici que la nature psycholinguistique en réinterprétant ces listes comme un **discours**, certes très particulier, même si sa pratique apparaît comme familière dans notre culture.

Cette analyse ne peut cependant intervenir que sur des séquences discursives **individuelles**, seul plan où pourraient se manifester les jeux des contraintes morphologiques, morpho-phoniques, des contiguïtés des rimes, sur des processus psycholinguistiques individuels et qui se trouvent effacés dans le calcul d'un protocole moyen. Ce dernier, en regroupant les données individuelles " casse " précisément la séquentialité discursive du protocole individuel.

Ainsi, pour la catégorie des arbres, en français, comme en allemand, observe-t-on les séquences suivantes :

Protocole collectif

Protocoles individuels

1	Chêne	Chêne	Chêne	<i>Pommier</i>
2	<i>Pommier</i>	Sapin	Tilleul	<i>Poirier</i>
3.	<i>Sapin</i>	Pin	Erable	<i>Prunier</i>
4.	<i>Cerisier</i>	Orme	Frêne	Sapin
5.	<i>Poirier</i>	Hêtre	Olivier	Chêne
6.	<i>Hêtre</i>	Bouleau	Araucaria	Noyer
7.	<i>Bouleau</i>	Eucalyptus	Yeuse	Aubépine
8.	<i>Pin</i>	<i>Peuplier</i>	Sapin	Platane
9.	<i>Peuplier</i>	<i>Pommier</i>	Cyprès	Hêtre
10.	<i>Prunier</i>	<i>Prunier</i>	Cèdre	Pin
11	<i>Saule</i>	<i>Poirier</i>	Platane	Mélèze
12	<i>Erable</i>	<i>Figuier</i>	Marronnier	Cognassier
13	<i>oranger</i>	Pêcher	Noyer	Sureau

1	<i>Eiche</i>	Birke	Kastanie	<u><i>Apfelbaum</i></u>
2	<i>Tanne</i>	Magnolia	Buche	<u><i>Kirschbaum</i></u>
3.	<i>Fichte</i>	Buche	Birke	<u><i>Olivenbaum</i></u>
4.	<i>Birke</i>	Eiche	Pappeln	<u><i>Birnenbaum</i></u>
5.	<i>Buche</i>	Tanne	Eiche	<u><i>Ahornbaum</i></u>
6.	<i>Kastanie</i>	Fichte	Apfelbaum	<u><i>Kastanienbaum</i></u>
7.	<i>Apfelbaum</i>	<u><i>Apfelbaum</i></u>		Blätter
8.	<i>Ahorn</i>	<u><i>Birnenbaum</i></u>		Baumstamm
9.	<i>Birnenbaum</i>	<u><i>Pflaumenbaum</i></u>		Ast
10.	<i>Kirschbaum</i>	<u><i>Kastanienbaum</i></u>		Buche
11	<i>Pappel</i>	<u><i>Mirabellenbaum</i></u>		Eiche
12	<i>Blätter</i>	<u><i>Haselnussbaum</i></u>		

Tableau 9 : Protocole collectif (fréquence de citation) et trois exemples de protocole individuel

On note ainsi, en particulier, sur certains protocoles individuels, l'émergence de séquences structurées par des similitudes morphologiques des signifiants : celle des suites déclinées avec le suffixe *-ier* en français ou *-baum* en allemand, séquences qui se trouvent effacées lors du moyennage sur le protocole collectif. Ces séquences peuvent certes correspondre à l'énumération d'une sous-catégorisation sémantique, celle de la sous-catégorie des arbres fruitiers, mais l'intrusion, par exemple, du *peuplier* dans les listes en français suggère qu'il s'agit davantage d'un phénomène lié aux propriétés des signifiants qu'à une propriété sémantique.

Conclusion : des « normes » aux listes ou d'une évidence à un phénomène à expliquer

Au terme de ces quelques repères identifiés dans le fonctionnement des « normes catégorielles », nous sommes conduits à remettre en cause, non pas la validité de cet outil de mesure¹², mais, davantage les présupposés théoriques dont ils relèvent, et dont l'efficacité comme instrument entretient l'illusion de vérité des représentations du monde ou d'un reflet de structures cognitives sous-jacentes à des phénomènes linguistiques sans épaisseur.

En résumé :

- les termes inducteurs ne peuvent être considérés comme renvoyant de manière homogène à des catégories univoquement définies *a priori*, et identiques d'un locuteur à un autre, d'une langue et d'un groupe culturel à un autre, en rendant possible l'identification un référent « vrai », universel, à partir de ces listes.
- les prototypes, définis comme items les plus saillants, sont diversement dépendants de différents principes d'organisation, de la similitude perceptive, à la disponibilité référentielle, mais aussi à des contraintes de langue (répertoire lexical de la langue, disponibilité dans le groupe et les sujets interrogés).
- la structure interne de la catégorie, elle aussi, dépend certes de contraintes cognitives que l'on peut assimiler à la typicalité mais aussi des spécificités de la tâche d'énumération de mots en séquence, où interviennent de ce fait des contraintes liées à cette séquentialité sur les processus psycholinguistiques de production, comme la diversité des liaisons d'un terme à un autre avec une relation plus ou moins directe au terme inducteur au fur et à mesure qu'on avance dans la liste, les dynamiques de rime liées aux propriétés morphosyntaxiques des formes lexicales, uniquement repérables dans les protocoles individuels.

Les « normes », dès qu'elles sont considérées en tant que listes de formes lexicales, voire de discours d'un genre particulier,

12 Même dans l'univers des sciences physiques des étalons de mesure peuvent avoir une portée opérationnelle, sans que leurs fondements théoriques ne soient clairement explicités.

apparaissent comme le résultat d'un grand nombre de processus et de facteurs hétérogènes qui en font à la fois l'intérêt et en tracent les limites. Intérêt comme instrument de mesure, efficace parce que corrélé avec nombre de variables, et du même coup, limité dans la possibilité d'identifier les déterminations ou des contraintes spécifiques (liées soit aux structures cognitives, soit aux structures des langues). La complexité et la diversité des processus qui régissent les listes catégorielles conduisent ainsi à récuser l'idée que ces « normes catégorielles » puissent refléter les structures cognitives directement reliées à des mots comme labels référant à l'évidence des choses, sans qu'interviennent des propriétés spécifiques des systèmes linguistiques et des phénomènes langagiers. L'adéquation des mots aux choses apparaît ainsi comme le résultat de conditions de productions très particulières, établies par consensus dans une communauté linguistique particulière, même si c'est la nôtre. Et donc, plutôt que d'être les prémisses à l'évaluation d'une hypothèse cognitive, les listes apparaissent comme le résultat, le produit d'un type très particulier de production langagière, qui **contraint** le sujet dans une stratégie de simple mise en correspondance des mots et des choses.

Outre la contribution à une sémantique lexicale, cette réflexion permet de clarifier le travail pluridisciplinaire au sein des sciences cognitives, en permettant de susciter divers questionnements occultés par quelques évidences concernant l'adéquation des mots aux choses :

Ainsi,

- dans quelle mesure, un même terme, celui de lexique, renvoie-t-il à un même concept relevant de la science cognitive et transcendant les domaines psychologique et linguistique ?
- ou
- dans quelle mesure, les modes d'objectivation du lexique dans chacun des domaines conduisent-ils à élaborer différents objets distincts sans correspondance bi-univoque nécessaire entre eux ?
- ou encore,
- dans quelle mesure, le nomadisme du concept de lexique impose-t-il ses caractéristiques, issues de sa conceptualisation dans un domaine en linguistique, dans le champ de la psychologie, ou inversement ?

L'hypothèse de déterminismes universels produisant des catégories universelles pourrait ainsi conduire à une théorie intéressante, à condition qu'elle ne soit pas un dogme auquel les données empiriques devraient se conformer. « La simplicité de nos théories – dont nous sommes les auteurs – n'implique nullement la

simplicité intrinsèque du monde. » K. Popper ; *Une défense de l'indéterminisme*, p. 37.

Liste de Références

- Battig, W. F., Montague, W. E. (1969) Category norms for verbal items in 56 categories : a replication and extension of the Connecticut Category Norms, *Journal of experimental Psychology Monograph*, 80, 1-46.
- Berlin, B., Kay, P. (1969) *Basic color terms*. Berkeley, University California Press.
- Dubois, D. (1983). Analyse de vingt-deux catégories sémantiques du Français : Organisation catégorielle, lexicale et représentation, *L'Année Psychologique*, 83, 465-489.
- Dubois, D. (1986) *La compréhension de phrases : représentations sémantiques et processus*, Thèse d'État, Paris VIII, Juin 1986, 335 p.
- Dubois, D. (1991) Les catégories sémantiques naturelles : prototype et typicalité, in D. Dubois (Éd.), *Sémantique et cognition : Catégories, concepts et typicalité*, Paris, Éditions du CNRS.
- Dubois, D. (1993) Lexique et catégories naturelles : représentations ou connaissances, *Cahiers de praxématique*, 21, 105-124.
- Dubois, D. (1997) *Catégorisation et cognition : de la perception au discours*, Paris, Kimé.
- Dubois, D., Poitou, J. (2002) « Normes catégorielles » (listes de termes) pour vingt-deux catégories sémantiques, *Cahiers du LCPE N° 5*.
- Dubois, D., Grinevald, C. (1999) Pratiques de la couleur et dénominations, *Faits de langues*, Ophrys, 14. (English version "Denominations of colors in practices" the XXVI LACUS forum proceedings, Edmonton, 1999).
- Dubois, D., Hilaire, G. (2002) Tables de propriétés pour des exemplaires typiques et non typiques appartenant à vingt-deux catégories sémantiques, *Cahiers du LCPE n° 5*.
- Dubois, D., Lavigne-Tomps F. (1994) Anaphore associative et traitements cognitifs, in C. Schmedecker et al. (eds.), *L'anaphore associative. Aspects linguistiques, psycholinguistiques et automatiques*, Paris, Klincksiek, 273-296.
- Dubois, D., Resche-Rigon, P. (1993). Prototypes ou stéréotypes: productivité et figement d'un concept, in C. Plantin (éd.) *Lieux communs, topoï, stéréotypes, clichés*, Paris, Kimé.
- Dubois, D., Resche-Rigon, P. (1995). De la « naturalité » des catégories sémantiques: des catégories d'objets naturels aux catégories lexicales, *Intellectica*, 20, 1, 217-245.
- Dubois, D., Resche-Rigon, P., Tenin, A. (1997) Des couleurs et des formes: catégories perceptives ou constructions cognitives, in D. Dubois (éd.), *Catégorisation et cognition*, Paris, Kimé, 7-40.
- Flavell, J. (1961) Meaning and meaning similarity : II The semantic differential and co-occurrence as predictors of judged similarity of meaning, *Journal of General Psychology*, 64, 321-335.
- Foley, W.A. (1997). *Anthopological linguistics*, Malden, Blackwell.

- Geeraerts, D. (1991) Grammaire cognitive et sémantique lexicale, *Communications*, 53, 17-50.
- Hampton, J., Dubois, D. (1993) Psychological models of concepts: introduction, in I. Van Mechelen, J. Hampton, R. Michalski, P. Theuns (éds.) *Categories and concept : theoretical views and inductive data analysis*, Londres, Academic Press, 11-22.
- Hampton, J., Gardiner, M. (1983) Measures of internal category structure : a correlational analysis of normative data, *British Journal of Psychology*, 74, 491-516.
- Hampton, J., Dubois, D., Yeh (soumis) The effects of pragmatic context on classification in natural categories.
- Hilaire, G. (2000) *Approche psycholinguistique de la dénomination d'objets naturels et manufacturés à partir d'images, de bruits et d'odeurs : le cas des patients Alzheimer*, Thèse, Université de Lyon 2.
- Howard, D.V. (1980) Category norms : a comparison of the Battig and Montague (1969) norms with the responses of adults between the age of 20 and 80, *Journal of gerontology*, 35, 225-231.
- Jodelet, F. (1965) L'association verbale, in F. Bresson, F. Jodelet, G. Mialaret (éds.) *Traité de psychologie expérimentale*, Tome VIII, 93-137.
- Kleiber, G. (1997a) Sens, référence et existence : que faire de l'extralinguistique ? *Langages*, 127, 9-37.
- Kleiber, G. (1997b) Les catégories de base donnent-elles lieu à des termes de base ? Hoinkes, U., Dietrich, W. (éds.) *Kaliedoskop der lexikalischen Semantik*, Tübingen, Narr, 71-90.
- Kucera, H., Francis, W. *Computational analysis of present day American English*, Providence, Brown University press, 1967.
- Lambert, W., Moore, N. (1966) Word association responses: comparisons of American and French monolinguals with Canadian monolinguals and bilinguals, *Journal of personality and Social Psychology*, 3, 313-320.
- Lieury, A., Iff, M., Duris, P. (1976) *Normes d'associations verbales*, Document miméo, Laboratoire de psychologie expérimentale et comparée Paris V.
- Lucy, J. (1992) *Language diversity and Thought*, Cambridge, Cambridge University Press.
- Mervis, C., Catlin, J., Rosch, E. (1976) Relationships among goodness-of-example, category norms, and word frequency, *Bulletin of the Psychonomic Society*, 7, 283-284.
- Poitou, J. (2000) Prototypes, saillance et typicalité, *Terminologie nouvelles*, 21, 16-26.
- Poitou, J., Dubois, D. (1999) Catégories sémantiques et cognitives : une étude expérimentale en sémantique lexicale, *Cahiers de lexicologie*, 74, 15-27.
- Rastier, F. (1991) *Sémantique et recherches cognitives*, Paris P.U.F.
- Rosch, E. (1973) Natural Categories, *Cognitive Psychology*, 7, 573-605.
- Rosch, E. (1973) On the internal structure of perceptual and semantic categories, in T.E Moore (éd.) *Cognitive development and the acquisition of language*, New York, Academic press.

- Rosch, E. (1975) Cognitive representations of semantic categories, *Journal of experimental psychology*, 104, 192-233.
- Rosch, E. (1978) Principles of categorization, in E. Rosch, B. Lloyd (éds.) *Cognition and categorization*. New York, L. Erlbaum.
- Segalowitz, N., Poulin-Dubois, D. (1990) The structure of categories: typicality gradients, perceived linguistic familiarity and cross-linguistic comparisons, *Cahiers de psychologie cognitive*, 10, 491-512.
- Selosse, P. (2000) *Un aspect de l'épistémè de la renaissance : méthode et nomenclature dans l'œuvre botanique de Caspar Bauhin*, Thèse, Université de Paris IV.
- Taylor, J. (1995) *Linguistic categorization*, Oxford, Clarendon Press.
- Van Mechelen, I., Hampton, J., Michalski, R., Theuns, P. (éds.) (1993) *Categories and concepts: theoretical views and inductive data analysis*, Londres, Academic Press.
- Vandeloise, C. (1991) Autonomie du langage et cognition, *Communications*, 53, 69-102.
- Waxman, S. (1999) The dubbing ceremony revisited: Objects naming and categorization in Infancy and early Childhood, in D. Medin, S. Atran (éds.) *Folkbiology*, Cambridge, M.I.T. Press.
- Wierzbicka, A. (1990) The meaning of color terms: semantics, culture and cognition, *Cognitive Linguistics*, 1-1, 99-150.
- Wierzbicka, A. (1996) *Semantic primes and universals*, Oxford, Oxford University Press.