Probabilités et statistiques en psychologie et en linguistique

Philippe Gréa

▶ **To cite this version:**

Philippe Gréa. Probabilités et statistiques en psychologie et en linguistique : Petit tour d'horizon. Texto ! Textes et Cultures, 2017. halshs-01548454

HAL Id: halshs-01548454
https://shs.hal.science/halshs-01548454

Submitted on 26 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Probabilités et statistiques en psychologie et en linguistique

Petit tour d'horizon*

Philippe Gréa
Université Paris Nanterre & MoDyCo CNRS-UMR7114

* Ce texte est extrait d'un mémoire d'HDR soutenu en 2016.

Table des matières

Table des matières 3

1 Les probabilités et les statistiques dans la cognition 5

1.1 L'approche probabiliste de la catégorisation (fréquentiste) 5

1.2 Le cerveau bayésien (approche subjective) 8

1.3 L'apprentissage profond 17

1.4 Avantages et inconvénients 19

1.4.1 Mots grammaticaux et mots lexicaux 20

1.4.2 La flexibilité sémantique 20

2 Les probabilités et les statistiques en linguistique 24

2.1 La fréquence et ses effets 25

2.2 Sémantique et probabilités 28

3 Affinités et mesures d'association 32

3.1 L'information mutuelle 32

3.2 L'analyse collostructionnelle 34

3.3 L'exception culturelle française : le calcul des spécificités 36

3.4 Sens et collocation 45

Bibliographie générale 49

Index des noms 55

Index thématique 57

1 Les probabilités et les statistiques dans la cognition

Dans les années 80, une polémique a opposé deux conceptions de la cognition : le fonctionnalisme (ou computationnalisme), qui considère, à l'image de ce que fait un ordinateur, que la pensée correspond à un calcul sur des représentations symboliques, et le connexionnisme, qui la conçoit comme une inférence statistique. À cette époque, l'un des arguments contre l'approche connexionniste consistait à minimiser l'importance des processus statistiques dans la cognition :

> Give up on the idea that networks offer (to quote Rumelhart & McClelland [1986: 110]) "... a reasonable basis for modeling cognitive processes in general". It could still be held that networks sustain some cognitive processes. A good bet might be that they sustain such processes as can be analyzed as the drawing of statistical inferences; as far as we can tell, what network models really are is just analog machines for computing such inferences. Since we doubt that much of cognitive processing does consist of analyzing statistical relations, this would be quite a modest estimate of the prospects for network theory compared to what the Connectionists themselves have been offering. (Fodor & Pylyshyn 1988: 68)

Aujourd'hui, la situation s'est inversée et de nombreuses disciplines associées de près ou de loin aux sciences cognitives mettent les probabilités et les statistiques au cœur de la cognition, et considèrent qu'une grande partie de nos facultés s'expliquent grâce aux régularités statistiques de l'environnement.

1.1 L'approche probabiliste de la catégorisation (fréquentiste)

Dans le domaine de la psychologie, on voit émerger une conception probabiliste des notions de concept et de catégorisation dès les années 70, avec la notion de *cue-validity* (Rosch 1973;

Rosch 1978; Rosch & Mervis 1975; Smith & Medin 1981). L'intuition défendue par la théorie du prototype est l'idée que certains attributs (traits, ou encore, propriétés) sont plus fortement associés à certaines catégories que d'autres. Par exemple, « avoir une crinière » est un attribut que seuls quelques animaux possèdent (le cheval, l'âne, le zèbre, le lion mâle, etc.) à l'exclusion des autres. La *cue-validity* est un moyen de quantifier la force de cette association. Elle calcule la probabilité conditionnelle qu'un exemplaire appartienne à une catégorie sachant qu'il a tel attribut :

$$(1) \qquad P(\textit{catégorie} \mid \textit{attribut}) = \frac{P(\textit{catégorie}, \textit{attribut})}{P(\textit{attribut})}$$

Cette égalité se lit de la façon suivante : la probabilité d'appartenir à telle catégorie sachant qu'on a tel attribut *P(catégorie|attribut)* est égale à la probabilité conjointe *P(catégorie, attribut)* – à savoir le fait d'appartenir à telle catégorie et, dans le même temps, d'avoir cet attribut –, divisée par la probabilité d'avoir cet attribut *P(attribut)*. Afin d'illustrer le fonctionnement de (1), prenons un exemple de monde réduit dans lequel il existerait trois catégories d'animaux : cheval, chien, poisson ; et trois attributs distincts : « avoir une crinière », « avoir quatre pattes », « avoir une paire d'yeux ». Partons du principe que chaque catégorie d'animaux possède 10 exemplaires (soit une population totale de 30 individus). Précisons en outre que si les chevaux de notre monde ont tous une crinière, c'est aussi le cas de l'un des dix chiens (il s'agit d'un mastiff tibétain, qui n'est pas un chien parfaitement prototypique, mais qui est tout de même un chien).[1] Dans cet état de choses, la probabilité qu'un nouvel exemplaire soit un cheval sachant qu'il a une crinière est très haute. Pour s'en convaincre, il suffit de résumer la situation à l'aide du Tableau 1 puis d'appliquer la définition de la probabilité conditionnelle donnée en (1) :

Tableau 1.

	« avec crinière »	« sans crinière »	total
cheval	10/30	0/30	10/30
chien	1/30	9/30	10/30
poisson	0/30	10/30	10/30
total	11/30	19/30	30/30

$$(2) \qquad P(\textit{cheval} \mid \textit{crinière}) = \frac{P(\textit{cheval}, \textit{crinière})}{P(\textit{crinière})} = \frac{10/30}{11/30} = 0.90\underline{90}...$$

[1] Pour changer, nous ne prenons pas l'exemple classique des oiseaux, mais le raisonnement aurait été identique : nous aurions ainsi 10 oiseaux, dont 9 volent, tandis que le dixième (par exemple, une autruche) ne vole pas.

La relation d'association entre la catégorie *cheval* et l'attribut « avoir une crinière » est donc très forte (malgré l'existence du mastiff tibétain) : on dira que « avoir une crinière » est une propriété typique de la catégorie *cheval*. En revanche, la probabilité qu'un nouvel exemplaire soit un cheval sachant qu'il a quatre pattes est plus faible (puisque tous les chiens ont aussi quatre pattes). L'attribut « avoir quatre pattes » est donc beaucoup moins typique de la catégorie *cheval* :

(3) $$P(cheval \mid pattes) = \frac{P(cheval, pattes)}{P(pattes)} = \frac{10/30}{20/30} = 0{,}5$$

Quant à la probabilité d'être un cheval sachant que l'attribut « avoir une paire d'yeux » est vérifié, elle est encore plus faible (puisque tous les chiens et les poissons ont aussi une paire d'yeux) :

(4) $$P(cheval \mid yeux) = \frac{P(cheval, yeux)}{P(yeux)} = \frac{10/30}{30/30} = 0{,}3\underline{3}...$$

Une telle mesure est donc un bon moyen de quantifier la typicité d'un attribut pour une catégorie donnée. En outre, elle est utilisée pour définir un « niveau de base » (*basic level*), qui regroupe les catégories qui maximisent la *cue-validity*, c'est-à-dire qui comptent le plus d'attributs (très) typiques. Or, ce n'est pas le cas des catégories superordonnées (*mammifère*, *machine*) car elles ont peu d'attributs typiques qui les distinguent des autres catégories. Ce n'est pas non plus le cas des catégories subordonnées (*mastiff tibétain*, *Mercedes*) puisqu'elles partagent beaucoup de leurs attributs typiques avec la catégorie basique dont elles dépendent (respectivement, *chien* et *voiture*).

Transposée dans le domaine de la sémantique lexicale, cette mesure de typicité a connu un grand succès. Elle permet en effet d'introduire beaucoup de souplesse par rapport au principe d'opposition distinctive chère au structuralisme :

> La prise en compte des propriétés typiques implique un changement radical dans la façon de concevoir la définition sémantique d'un terme. Le modèle des CNS a pour vocation légitime de fournir une définition contrastive, qui indique clairement les traits qui séparent une catégorie des autres. L'approche prototypique ouvre la porte aux traits non contrastifs. Il ne s'agit plus seulement de dire ce qui distingue un chien d'un chat, mais de décrire positivement ce qu'est un chien et ce qu'est un chat. (Kleiber 1990 : 74)

Dans le domaine de la psychologie, en revanche, la *cue-validity* a fait l'objet de discussions dès les années 80, et de nombreux travaux ont montré les limites de cette approche. L'un des

reproches qu'on peut faire à ce genre de modèle est formulé par Smith & Medin (1981), et tient au fait qu'un ensemble d'attributs typiques ne représente en réalité qu'une faible partie des connaissances que les individus ont d'un concept. Par exemple, on sait que certaines conjonctions d'attributs sont beaucoup plus probables que d'autres : la probabilité d' « avoir des ailes » et « avoir des plumes » est beaucoup plus élevé que celle d' « avoir des ailes » et « avoir de la fourrure » et cette information fait aussi partie de nos connaissances générales sur le monde. Les modèles fondés sur la notion *cue-validity* ont donc beaucoup évolués depuis les travaux de Rosch. Certains auteurs ont tentés d'amender le principe. C'est par exemple le cas de Barsalou & Billman (1989) qui mettent en avant la notion de *systematicity* (le degré de cohérence d'une catégorie) en établissant une différence entre l'attribut et sa valeur. L'objectif de ces derniers est d'apporter une solution à l'apparente contradiction qui existe entre la stabilité d'une catégorie (dans la mémoire à long-terme) et sa variation (liée au contexte ou a des effets de récence). D'autres, en revanche, ont remis en question la pertinence de la notion. Ainsi, Murphy (1982) montre que la *cue-validity* ne parvient pas à discriminer les catégories du niveau de base. Dans Murphy & Medin (1985) on trouve une critique plus générale de l'approche probabiliste de la catégorisation au profit d'un autre paradigme qui met en avant les théories d'arrière plan sur lesquelles les individus s'appuient pour catégoriser.

Nous ne pouvons faire ici la synthèse de l'importante littérature consacrée à la question de la catégorisation en psychologie. Ce que nous retiendrons toutefois, c'est que dans le domaine de la psychologie, les *cue-validity models* en particulier, et l'approche prototypique en général, ne constituent aujourd'hui qu'une partie réduite des alternatives théoriques possibles.

1.2 Le cerveau bayésien (approche subjective)

A partir des années 2000 et dans un tout autre domaine, à savoir celui de la neurophysiologie, les chercheurs commencent à utiliser des modèles probabilistes qui s'appuient sur un autre concept mathématique, l'inférence bayésienne. Cette dernière s'avère être d'une grande efficacité pour modéliser et comprendre les fonctions du cerveau (Rao, Olshausen, & Lewicki 2002). Elle se fonde sur la règle de Bayes (pasteur et mathématicien anglais du XVIIIème

siècle), que l'on dérive très simplement à partir de la définition de la probabilité conditionnelle :

(5) a. $P(y \mid x) = \frac{P(y,x)}{P(x)}$ (probabilité de y sachant x, cf. [1-4])

b. $P(y,x) = P(y \mid x) \times P(x)$ (grâce à a.)

c. $P(x \mid y) = \frac{P(x,y)}{P(y)}$ (probabilité de x sachant y)

d. $P(x,y) = P(x \mid y) \times P(y)$ (grâce à c.)

e. puisque $P(y,x) = P(x,y)$ (la probabilité conjointe est symétrique)

f. alors $P(y \mid x) \times P(x) = P(x \mid y) \times P(y)$ (grâce à b. et d.)

g. donc $P(y \mid x) = \frac{P(x \mid y) \times P(y)}{P(x)}$ (règle de Bayes) (grâce à f.)

On peut appliquer sans difficulté cette règle à notre problème de cheval. Ainsi, la probabilité d'être un cheval sachant qu'on a une crinière se calcule avec la règle de Bayes de la façon suivante pour un résultat identique à (2) :

(6) $$P(cheval \mid crinière) = \frac{P(crinière \mid cheval) \times P(cheval)}{P(crinière)} = \frac{\frac{10/30}{10/30} \times 10/30}{11/30} = 0.90\underline{90}...$$

Cependant, l'intérêt premier de la règle de Bayes n'est pas de recalculer une probabilité conditionnelle grâce à une autre méthode. Il tient d'abord et avant tout à la possibilité de relier *P(cheval|crinière)* (d'un côté de l'égalité) et *P(crinière|cheval)* (de l'autre), ou plus généralement, de relier la plausibilité d'une hypothèse ou d'une interprétation sachant qu'on a telle donné, *P(hypothèse|donnée),* et la plausibilité de cette donnée sachant une certaine hypothèse ou interprétation, *P(donnée|hypothèse)*. En d'autres termes, elle est un moyen de mathématiser l'inférence, un processus que les humains réalisent quotidiennement sans même y penser :

(7) $$P(h \mid d) = \frac{P(d \mid h) \times P(h)}{P(d)}$$

Un exemple intuitif, que nous reprenons à Tenenbaum, Kemp, Griffiths, & Goodman (2011 : 1280), permettra d'en illustrer le principe général. Disons que nous sommes au mois de février. Je constate que depuis quelques jours, Pierre tousse. C'est ce qu'on appelle la donnée

(que l'on note *d* dans ce qui suit). En rapport à cette donnée, je formule alors trois hypothèses ou interprétations différentes (notées h_1, h_2 et h_3) :

- h_1 : Pierre a une grippe
- h_2 : Pierre a un cancer du poumon
- h_3 : Pierre a une gastro-entérite

La question à laquelle il s'agit de répondre est alors la suivante : laquelle de ces trois hypothèses est la plus plausible sachant la donnée (Pierre tousse) ? Traduit en termes mathématiques, cette question revient simplement à calculer $P(h_1|d)$ (la plausibilité de h_1 sachant que Pierre tousse), $P(h_2|d)$ et $P(h_3|d)$ puis à sélectionner l'hypothèse dont la plausibilité est la plus forte. Or, c'est justement ce que permet de faire la règle de Bayes. Pour le montrer, appliquons-la à chaque hypothèse (nous n'aborderons pas la question du dénominateur $P(d)$, qu'on appelle l'évidence, et qui est identique dans les trois situations) :

(8) $$P(h_1 \mid d) = \frac{P(d \mid h_1) \times P(h_1)}{P(d)}$$

(9) $$P(h_2 \mid d) = \frac{P(d \mid h_2) \times P(h_2)}{P(d)}$$

(10) $$P(h_3 \mid d) = \frac{P(d \mid h_3) \times P(h_3)}{P(d)}$$

- Dans le cas (8) (plausibilité que Pierre ait une grippe sachant qu'il tousse), il faut tout d'abord calculer le premier terme, $P(d|h_1)$, (la probabilité qu'on tousse sachant qu'on a la grippe), qu'on appelle aussi la vraisemblance. Or, cette dernière est élevée car, quand on a la grippe, généralement, on tousse. Quant au second terme, $P(h_1)$, qu'on appelle la probabilité a priori, elle est aussi très élevée, puisque nous sommes en hiver et qu'à cette époque, il est fréquent d'avoir la grippe. Résultat : $P(h_1|d)$ (qu'on appelle aussi la probabilité a posteriori) est élevée.

- Dans le cas de figure (9), $P(d|h_2)$, (vraisemblance) est élevée (quand on a un cancer du poumon, généralement, on tousse) mais, en revanche, $P(h_2)$ (probabilité a priori) est faible (le cancer du poumon n'est pas une maladie très fréquente, en tout cas, pas aussi fréquente que la grippe). Résultat : $P(h_2|d)$ (probabilité a posteriori) est moins élevée que $P(h_1|d)$.

- Dans le dernier cas de figure, $P(d|h_3)$ (la vraisemblance) est faible (quand on a une gastro-entérite, on ne tousse pas, on a mal au ventre) et $P(h_3)$ (probabilité a priori) est élevé (l'hiver, c'est aussi l'époque de la gastro). Résultat : $P(h_3|d)$ (probabilité a posteriori) est moins élevée que $P(h_1|d)$.

Conclusion : l'hypothèse h_1 est plus plausible que les deux autres.

Bien évidemment, tout le monde arrive à cette conclusion sans effort particulier. La règle de Bayes a toutefois l'intérêt de donner un substrat mathématique à ce genre de raisonnement naturel. En outre, même si ce type de raisonnement n'apporte aucune certitude, la possibilité de le mathématiser montre bien qu'il est de nature rationnelle. Cependant, il semble qu'il faille attendre les travaux de Jaynes (2003) pour que le raisonnement bayésien puisse être conçu comme une extension à part entière de la logique classique, et qu'il acquiert le statut de raisonnement valide. Dans son ouvrage, Jaynes montre que l'inférence bayésienne formalise un certain type de raisonnement, le raisonnement plausible, qui se caractérise par une connaissance incomplète d'une situation donnée. Contrairement au raisonnement déductif, il ne peut fournir aucune certitude, mais il n'en reste pas moins rationnel. Afin de nous faire une idée plus précise de la différence entre inférence bayésienne et déduction logique, comparons les deux modes de raisonnement.

La déduction logique consiste à utiliser les syllogismes, tel que le modus ponens (Jaynes 2003 : 2) :

(11) Modus Ponens

Si P est vrai alors Q est vrai, P est vrai, donc Q est vrai

Exemple

S'il pleut à midi alors le ciel se couvre à 11h00

Il pleut à midi

Donc le ciel se couvre à 11h00

Cependant, dans leur vie quotidienne, les humains utilisent un autre genre de raisonnement, le raisonnement plausible qui prend la forme suivante :

(12) Raisonnement plausible

Si P est vrai alors Q est vrai, Q est vrai, donc P est plus plausible

Exemple

S'il pleut à midi alors le ciel se couvre à 11h00

Le ciel se couvre à 11h00

Donc il est plus plausible qu'il pleuve à midi

Dans cette optique, le raisonnement plausible est conçu comme un schéma d'inférence à part entière, au même titre que les syllogismes (modus ponens, modus tollens, syllogisme hypothétique). La différence est qu'un syllogisme déductif apporte une certitude (la

conclusion est vraie ou fausse), tandis qu'une inférence statistique permet d'estimer la plausibilité d'une conclusion en fonction des prémisses. À la suite du mathématicien Pólya et de R. T. Cox, Jaynes (2003) élabore une axiomatisation de ce type de raisonnement que nous ne développerons pas ici. À terme, l'objectif de Jaynes est d'obtenir un système formel de façon à ce qu'un robot puisse soit en mesure de calculer la plausibilité d'une proposition en fonction d'un contexte donné.

Pour illustrer l'intérêt que peut avoir ce genre de démarche concernant la question de la catégorisation, revenons une dernière fois à notre monde composé de trois espèces d'animaux. Imaginons que je m'apprête à rencontrer 10 nouveaux exemplaires de chiens et que je me pose la question de savoir s'ils auront ou non des crinières, et surtout, dans quelles proportions. Dans cette situation, plusieurs hypothèses sont possibles. Nous en choisissons trois :

- h_1. Une première hypothèse consiste à penser qu'il y aura un chien sur dix qui aura une crinière (crinière = 0.1).

- h_2. Une seconde hypothèse possible consiste à penser qu'il y aura un chien sur deux avec une crinière (crinière = 0.5).

- h_3. Une dernière consiste à penser que neuf chiens sur dix auront une crinière (crinière = 0.9).

Au vu de ce que je connais du monde réduit présenté plus haut, j'ai toutefois un certain nombre d'a priori fondés sur ce que je connais déjà de ce monde. L'un d'entre eux consiste à penser que si je venais à rencontrer de nouveaux exemplaires de chien, il y aurait peu de chance qu'ils aient une crinière, sans que cela ne soit impossible. Je suis donc capable de quantifier la plausibilité de chacune de trois hypothèses ci-dessus : la première est de loin la plus probable, tandis que la troisième est très improbable. Quant à seconde, elle n'est pas totalement improbable mais sa plausibilité reste cependant assez faible. Pour rendre compte de ces différents degrés de croyance, je peux associer à chaque hypothèse ci-dessus un degré de confiance sous la forme, à nouveau, d'une probabilité. Par exemple :

(13) $P(\text{crinière} = 0.1) = 0.7$
$P(\text{crinière} = 0.5) = 0.25$
$P(\text{crinière} = 0.75) = 0.05$

Ces probabilités permettent de quantifier mon a priori sur la catégorie des chiens et leur possibilité d'avoir ou non une crinière (on parle alors d'approche subjective des probabilités,

opposée à l'approche fréquentiste). Nous représentons cet a priori graphiquement dans le premier panneau de la Figure 1.

Imaginons maintenant que parmi les dix nouveaux exemplaires de chiens, il s'avère que trois d'entre eux ont une crinière. Cette observation n'est pas vraiment celle à laquelle je pouvais m'attendre. Devant ces nouvelles données, je suis donc obligé de réviser mes croyances sur les chiens. La règle de Bayes est justement le moyen de mathématiser ce processus d'actualisation des croyances. Pour cela, il suffit d'appliquer la règle de Bayes à chacune des trois hypothèses.

La difficulté principale tient au fait qu'il faut calculer chaque terme de (7) (vraisemblance, a priori et évidence).Commençons par le calcul du numérateur, $P(d|h)\,P(h)$, pour chacune des trois hypothèses. Nous le détaillons dans le Tableau 2 où chaque colonne correspond à une hypothèse. La première cellule du tableau correspond à la probabilité conjointe que j'observe ces données et que h_1 soit vraie : $P(d,h_1)$. La seconde est la probabilité que j'observe ces données et que h_2 soit vraie $P(d,h_2)$. La troisième correspond à $P(d,h_3)$. Grâce à (5b), je sais que $P(d,h) = P(d|h)P(h)$. La valeur de $P(h)$ ne pose aucun problème puisqu'il s'agit de mon a priori que j'ai intuitivement fixé en (13). La valeur de la vraisemblance, $P(d|h)$, est plus délicate à calculer. Détaillons par exemple le calcul de $P(d|h_1)$, c'est-à-dire la probabilité d'observer 3 chiens avec crinières parmi un groupe de dix chiens, sachant qu'on fait l'hypothèse qu'il y a une chance sur dix (0,1) pour qu'un chien ait une crinière et neuf chance sur dix (0,9) qu'il n'en ait pas. Il suffit de calculer la probabilité pour chaque chien puis de les multiplier (chaque chien est conçu comme un événement indépendant des autres). La vraisemblance $P(d|h_1)$ est alors : $0{,}1\times0{,}1\times0{,}1\times0{,}9\times0{,}9\times0{,}9\times0{,}9\times0{,}9\times0{,}9\times0{,}9$ ou plus simplement $0{,}1^3\times0{,}9^7$. Nous la représentons dans le second panneau de la Figure 1. Il suffit ensuite de multiplier ce résultat par notre a priori $p(h_1)$, c'est-à-dire le degré de croyance que j'accorde à l'hypothèse h_1 (cf. [13]) à savoir 0,7. J'obtiens alors : $0{,}1^3\times(0{,}9)^7\times0{,}7 = 0{,}0003348078$. Le raisonnement est identique pour h_2 et h_3 et nous ne le détaillerons donc pas. L'étape suivante consiste à calculer le dénominateur $p(d)$. Ce dernier s'obtient facilement, en additionnant les trois numérateurs de chaque hypothèse (cf. Tableau 2). Nous avons alors tous les termes pour faire une dernière division et obtenir $P(h_1,d)$, l'a posteriori, c'est-à-dire la plausibilité de mon hypothèse h_1 au regard des nouvelles données (3 chiens sur 10 avec une crinière). Le résultat est représenté graphiquement dans le troisième panneau de la Figure 1 (et reprend les résultats calculés en [14]).

Tableau 2.

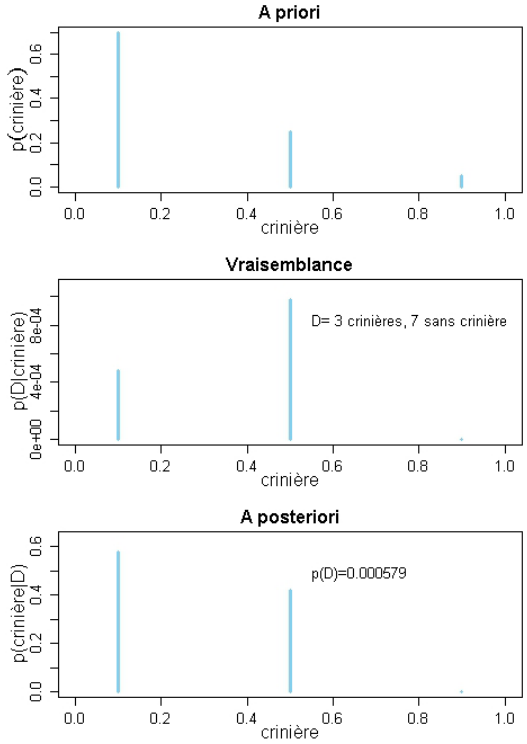
	h_1 : crinière = 0.1	h_2 : crinière = 0.5	h_3 : crinière = 0.9	total
d	$p(d,h_1)$ $= p(d\|h_1)\times p(h_1)$ $= 0{,}1^3\times(0{,}9)^7\times 0.7$ = 0,0004782969×0,7 = 0,0003348078	$p(d,h_2)$ $= p(d\|h_2)\times p(h_2)$ $=0{,}5^3\times(0{,}5)^7\times 0.25$ =0,0009765625×0.25 =0,0002441406	$p(d,h_3)$ $= p(d\|h_3)\times p(h_3)$ $= 0{,}9^3\times(0{,}1)^7\times 0.05$ = 7,29e-08×0,05 = 3,645e-09	$p(d)$ = 0,0005789486

(14) $$P(\textit{crinière} = 0.1 \mid d) = \frac{P(d \mid h_1)\times P(h_1)}{P(d)} = \frac{0.1^3(0.9)^7\times 0.7}{0.0005789486} = 0.5783032$$

$$P(\textit{crinière} = 0.5 \mid d) = \frac{P(d \mid h_2)\times P(h_2)}{P(d)} = \frac{0.5^3(0.5)^7\times 0.25}{0.0005789486} = 0.4216965$$

$$P(\textit{crinière} = 0.9 \mid d) = \frac{P(d \mid h_3)\times P(h_3)}{P(d)} = \frac{0.9^3(0.1)^7\times 0.05}{0.0005789486} = 6.295896^{\text{e-06}}$$

Figure 1.



On constate alors que si h_1 est toujours la plus plausible des trois hypothèses, elle est toutefois moins plausible qu'avant la présentation des données (0.58 au lieu de 0.7). La plausibilité de h_2, quant à elle, augmente pour atteindre une valeur de 0.42 alors que nous l'avions initialement fixée à 0.25 (soit presque le double). Enfin, h_3 se trouve encore moins probable qu'avant.

Entre le moment qui précède la présentation des données et celui qui suit, notre concept de *chien* à donc évolué, de sorte que l'attribut « avoir une crinière » est désormais un attribut relativement plausible du chien (sans toutefois être prototypique). La règle de Bayes constitue donc un bon moyen de formaliser cette évolution d'un concept en fonction des données dans la mesure où elle est itérative et peut s'appliquer autant de fois que nécessaire. Rien n'empêche, en effet, de prendre les résultats calculés en (14) pour établir un nouvel a priori sur la catégorie des chiens. Il devient alors possible de refaire le calcul lorsqu'on est confronté à de nouvelles données (de nouveaux exemplaires de chiens). En fonction de ces données, une nouvelle plausibilité des hypothèses sera calculée. Par exemple, h_2 (un chien sur deux a une crinière) pourrait se renforcer et devenir la plus plausible, ou encore, la plausibilité de h_3 pourrait remonter à son tour tandis que h_1 deviendrait moins probable, etc.

Il s'agit d'un exemple « jouet », qui n'a évidemment pas d'autre ambition que pédagogique (dans le monde réel, la crinière n'est définitivement pas un attribut associé à la catégorie des chiens, malgré l'existence des mastiffs tibétains), mais qui nous permet cependant de donner une petite idée de ce que propose Anderson (1990). Dans cet ouvrage, en effet, l'objectif de l'auteur consiste, entre autres, à modéliser la capacité de prédiction liée à la faculté de catégorisation, par exemple prédire la dangerosité d'une nouvelle entité après avoir vu *n* entités présentant certaines valeurs dans différentes dimensions (taille, poids, couleur, etc.) :

> Basically, our system has certain priors about the world that correspond to its assumptions about the structure of the environment; our system has experience with certain objects; and, from this, our system makes predictions about features of new objects. (Anderson 1990 : 100)
> Assume that the person has seen n objects, is presented with a n + 1st object with some dimensions specified, and wants to estimate the probabilities of various values on other dimensions — for example, we want to predict whether the next object we encounter has a positive value on the "dangerous" dimension. (Anderson 1990 : 101)

A notre connaissance, cet ouvrage est le premier à faire une utilisation explicite des statistiques bayésiennes pour modéliser le processus de catégorisation.[2] A la suite de cet ouvrage, de nombreux chercheurs en psychologie ont eu recours aux statistiques bayésiennes pour décrire avec succès des phénomènes aussi variés que la perception, la catégorisation, l'apprentissage, ou la prise de décision. Grâce à la règle de Bayes, certaines notions qui manquaient jusque là de précision trouvent désormais une formulation explicite. Par exemple,

[2] Précisons en outre que dans le même ouvrage, l'auteur rend aussi compte de la mémoire et de l'inférence causale à l'aide de la règle de Bayes.

Tenenbaum & Griffiths (2001) proposent une mathématisation, dans le cadre bayésien, de la notion de représentativité (*representativeness*, i.e. prototypicalité) plus élaborée que la simple probabilité conditionnelle (cf. section précédente). Les statistiques bayésiennes permettent en outre d'apporter une solution au problème classique de l'induction, que Quine illustre à l'aide du célèbre *gavagaï* (Quine 1960), et qui peut être interprété comme une sorte de métaphore de l'acquisition du sens des mots par un enfant :

> Consider a typical dilemma faced by a child learning English. Upon observing a competent adult speaker use the word dog in reference to Max, a particular Dalmatian running by, what can the child infer about the meaning of the word dog? The potential hypotheses appear endless. The word could refer to all (and only) dogs, all mammals, all animals, all Dalmatians, this individual Max […] (Xu & Tenenbaum 2007: 245)

Xu & Tenenbaum (2007) et Tenenbaum, Kemp, Griffiths, & Goodman (2011) proposent ainsi un modèle de l'acquisition du sens d'un mot qui n'est pas très éloigné de notre exemple des chiens avec (ou sans) crinière, puisqu'il se fonde sur le calcul de la plausibilité des différentes hypothèses possibles :

> Our key innovation is the use of a Bayesian inference framework. Hypotheses about word meanings are evaluated by the machinery of Bayesian probability theory rather than deductive logic: Hypotheses are not simply ruled in or out but scored according to their probability of being correct. (Xu & Tenenbaum 2007 : 246)

Toute la difficulté de l'entreprise consiste à trouver une formulation mathématique satisfaisante de la vraisemblance et de l'a priori dans ce genre de situation. C'est ce que les auteurs réussissent à faire (i) en posant que la vraisemblance diminue lorsque la taille de la catégorie augmente et (ii) en ramenant l'a priori à un « hierarchical clustering ("average linkage"; [Duda & Hart 1973]) on human participants' similarity ratings ».

L'ensemble de ces travaux a donné naissance à un cadre théorique général connu sous le nom de « théorie du cerveau statisticien », et que certains considèrent comme une véritable révolution dans les sciences cognitives.[3] Dans cette approche, le cerveau est conçu comme un calculateur bayésien qui fait des hypothèses sur le monde et qui en actualise en permanence la plausibilité en fonction des données, de la vraisemblance et de ses a priori.

[3] Cf. par exemple S. Dehaene qui en fait une présentation détaillée dans ses conférences au Collège de France, accessibles aux liens suivants :
http://www.college-de-france.fr/site/stanislas-dehaene/course-2011-2012.htm
http://www.college-de-france.fr/site/stanislas-dehaene/course-2012-2013.htm

1.3 L'apprentissage profond

Dans le domaine de l'intelligence artificielle, le cas de l'apprentissage profond (*deep learning*) nous semble aujourd'hui difficilement contournable. Étant donné le succès actuel de cette approche, on pourrait nous reprocher de céder à la mode du moment. Cependant, il est indiscutable que le deep learning signe aujourd'hui le retour en force des réseaux de neurones dans l'intelligence artificielle, et constitue une réponse à la citation de Fodor & Pylyshyn qui ouvre la section 1.

Depuis les années 40, le connexionnisme connaît un parcours pour le moins étrange, oscillant entre de grands moments de gloire et des périodes d'indifférence plus ou moins prolongées. Cette évolution débute avec l'article de McCulloch & Pitts (1943) qui lance l'idée qu'un réseau de neurones est en mesure de reproduire une porte logique. Avec le perceptron, élaboré par Rosenblatt, le connexionnisme tient alors lieu d'alternative au modèle computationnel « orthodoxe », avant de connaître une première phase d'oubli après la publication de Minsky & Papert (1988) qui démontrent les limites du modèle. Il renaît de ses cendres avec deux ouvrages issus d'un célèbre collectif scientifique, le *Parallel Distributed Processing (PDP) group* (McClelland & Rumelhart 1989; Rumelhart & McClelland 1986) puis retombe dans l'indifférence au tournant des années 90 – 2000. À partir du milieu des années 2000, le connexionnisme, sous sa nouvelle étiquette de « deep learning », va peu à peu soulever une nouvelle vague d'enthousiasme dans tous les laboratoires d'informatique. Cette soudaine notoriété est d'autant plus étonnante qu'il n'y a aujourd'hui rien de fondamentalement nouveau par rapport aux connaissances mathématiques que nous avions des réseaux de neurones dans les années 90. En particulier, la règle d'apprentissage utilisée dans le *deep learning* est la même qu'à l'époque : la rétropropagation du gradient d'erreur.[4] La seule chose qui change, c'est le fait que nous disposons désormais de bases d'apprentissage colossales et de moyens de calculs astronomiques.

Les réseaux de neurones sont conçus et développés, entre autres applications, pour répondre au genre de question posée dans les sections précédentes : quelle est la catégorie de telle entité sachant qu'elle possède tel attributs ; quelle est sa dangerosité sachant ses attributs (cf. Anderson, section 1.2) ? Mais ils sont aussi susceptibles de répondre à beaucoup d'autres questions qui n'ont aucun rapport avec les préoccupations de la psychologie (ou de la linguistique), comme par exemple, la détection de fraudes sur les cartes bancaires ou de

[4] Que nous évoquons en 1996 dans notre DEA de Sciences du langage.

problèmes médicaux (sur la base de différents symptômes). En réalité, un réseau de neurones est toujours capable d'apporter une réponse à n'importe quelle question, pour peu qu'on lui donne suffisamment d'exemples (on parle d'apprentissage supervisé). Mathématiquement, cela revient à donner au réseau un vecteur d'entrée *x* (par exemple, une liste d'attributs liés à une entité, une liste d'attributs liés à un retrait bancaire [montant, date, heure, etc.], une liste de symptômes) et à lui donner les moyens d'apprendre, à l'aide de la rétropropagation de l'erreur et sur la base de très nombreux exemples, une fonction *f(x)* qui donne un résultat en sortie (une catégorie, une probabilité d'avoir à faire à une fraude, à une maladie, etc.).

Cependant, même si les résultats obtenus actuellement par le deep learning dépassent de loin tout ce qu'on pouvait imaginer il y a encore dix ans, les reproches formulés contre les réseaux de neurones dans les années 80-90 sont toujours d'actualité. Tout d'abord, si le réseau de neurones est capable d'apprendre la fonction *f(x)* et de faire des généralisations à partir de cette dernière, il est impossible de savoir quelles sont les variables qui ont été extraites lors du processus d'apprentissage. En d'autres termes, le réseau de neurones est une sorte de boite noire capable de reproduire un processus (de catégorisation, par exemple), mais pas de l'expliquer. Deuxièmement, l'optimisation d'un réseau de neurones, et tout particulièrement, celle des réseaux en cascade utilisés dans le *deep learning*, relève d'une pratique intuitive qui n'est actuellement contrôlée par aucune théorie mathématique. Pour qu'un réseau converge vers la fonction *f(x)*, il faut essayer, plus ou moins au hasard, différentes architectures (et dans le cas du *deep learning*, cette question peut devenir très compliquée en raison des nombreux sous-réseaux mis en cascade), différentes fonction d'activation (les réseaux profonds ont renoncés à la tangente hyperbolique et à la sigmoïde au profit, par exemple, de la fonction *Rectified Linear Unit* [ReLU]), différentes techniques d'apprentissage (par exemple, l'utilisation du *drop-out*, qui consiste à éliminer la moitié des unités d'une couche pour mieux distribuer la décision sur toutes les unités), sans que l'on puisse trop savoir à l'avance ce qui va marcher et si ça va marcher.[5] L'expérience acquise par le chercheur après plusieurs années d'essais et d'erreurs est donc un élément important de la discipline. Pour cette raison, les réseaux de neurones sont souvent considérés comme relevant de l'ingénierie plutôt que d'une discipline fondamentale mathématiquement bien circonscrite.

[5] Montavon, Orr, & Müller (2012) est ainsi entièrement consacré aux différents « trucs » qui permettent au réseau de converger vers une solution.

1.4 Avantages et inconvénients

Ce rapide panorama, qui ne prétend aucunement à l'exhaustivité, nous aura donc au moins permis d'esquisser le rôle central que jouent les probabilités et les statistiques dans la cognition, et ce, depuis les années 80 jusqu'à nos jours. Nous avons aussi pu constater que la notion de plausibilité, que l'on mathématise à l'aide de l'inférence bayésienne, n'entre pas fondamentalement en contradiction avec la logique formelle. De telles avancées sont donc d'une grande importante pour la psychologie, puisqu'elles permettent d'expliciter, dans un cadre rationnel, des notions centrales comme la catégorisation, la prototypicalité, la généralisation, etc. Du côté de l'intelligence artificielle, nous avons aussi évoqué le cas du *deep learning*, qui présente un cadre théorique et mathématique peut-être un peu moins raffiné que les modèles bayésiens issus de la psychologie cognitive, mais qui donne aujourd'hui des résultats impressionnants.

Cependant, du point de vue de la linguistique, et en particulier, de la sémantique, ces analyses ne sont pas totalement satisfaisantes. L'extraction d'une liste d'attributs (ou un réseau d'attributs, ou quoi que ce soit d'autre) à partir de la présentation d'un objet puis sa généralisation à d'autres objets (que ce soit dans une approche fréquentiste ou subjective) est une capacité que l'on peut certes modéliser de façon satisfaisante au moyen des probabilités bayésiennes. De même, un réseau de neurones est aujourd'hui capable de traiter un problème non linéairement séparable dans un espace à très grande dimension et, par exemple, de détecter des régularités sur la base de milliers d'images différentes avec un taux de réussite spectaculaire. Mais encore faut-il s'assurer que le sens d'un mot est bien réductible à ces attributs ou ces régularités que l'on infère à partir du ou des référent(s). Pour le dire de façon simple, les analyses qui précédent posent un problème au sémanticien parce que la notion de concept ou de catégorie est élaborée en étroite relation avec son référent. Cette situation est propice à un glissement terminologique entre sens et référence qui n'est pas toujours explicite dans les travaux mentionnés plus haut, mais qu'on peut facilement mettre à jour. Par exemple, dans le travail de Xu & Tenenbaum (2007), l'a priori, qui correspond à l'espace des « sens » possibles pour un mot donné (et dont le modèle doit choisir le plus probable), prend en réalité la forme d'une simple structure hiérarchique construite sur la base des similarités que les sujets perçoivent entre différents objets (forme, texture, taille, etc.). Dans cette situation, la question du sens est entièrement repliée sur celle du référent. Or, il existe de nombreux arguments qui montrent que le sens est, dans une très large mesure, indépendant du référent,

et qu'il fonctionne selon un régime qui lui est propre. Nous en retiendrons deux dans ce qui suit.

1.4.1 Mots grammaticaux et mots lexicaux

Le premier argument tient à l'existence d'une classe de mots qui s'oppose aux mots lexicaux (ou encore, mots « pleins »), et que l'on retrouve dans toutes les langues naturelles, mais qui ne se réduisent pas facilement à une liste d'attributs. Ce sont les mots grammaticaux ou « vides » (par exemple, les déterminants *quelques* et *plusieurs*, ou les prépositions *parmi* et *entre*). Ces derniers ne sont généralement pas pris en compte dans les expériences de psychologie. La raison en est simple : s'ils ont un sens, il ne s'agit pas d'un sens référentiel mais d'un sens fonctionnel (on parle aussi de mot-outils) dont la nature semble assez différente de celle des mots pleins, et dont la sémantique doit évidemment rendre compte.

1.4.2 La flexibilité sémantique

Deuxièmement, le contexte est susceptible de faire varier les attributs associé à un mot lexical dans des proportions importantes. Les psychologues ont bien conscience de ce phénomène, qu'ils désignent par le nom de « flexibilité sémantique » (*semantic flexibility*) et dont ils tentent de rendre compte dès le milieu des années 70 (Anderson & Ortony 1975; Barclay, Bransford, Franks, McCarrell, & Nitsch 1974; Barsalou 1982; Barsalou 1993; Greenspan 1986; Roth & Shoben 1983; Schoen 1988; Tabossi & Johnson-Laird 1980). L'exemple princeps (pour la psychologie) est proposé par Barclay, Bransford, Franks, McCarrell, & Nitsch (1974) :

(15) *The man {lifted / tuned} the piano.*

Dans le cas où l'homme soulève le piano, *piano* présente l'attribut « être lourd », tandis que dans le cas où il en joue, c'est « instrument de musique » qui est activé. À la suite de cet article, de nombreux autres exemples vont être discutés par les psychologues :

(16) a. *The accountant pounded {the stake / the desk}.* [hammer vs. fist] (Anderson & Ortony 1975: 170)

b. *{The goldsmith cut the glass with the / The mirror dispersed the light from} the diamond.* [hard vs. brilliant] (Tabossi & Johnson-Laird 1980: 597)

c. *{The fresh meat was protected by / Robert fell on} the ice.* [frozen vs. slipper] (Greenspan 1986: 545)

De tels exemples ont permis, en particulier, d'examiner l'articulation entre mémoire de travail et mémoire à long-terme, entre propriétés dépendantes ou indépendantes du contexte, etc.

Du côté de la linguistique, ce phénomène occupe évidemment une place centrale. Weinreich l'illustre, deux ans avant l'article de Barclay, Bransford, Franks, McCarrell, & Nitsch (1974), à l'aide d'un exemple célèbre :

> When one considers the phrases *eat bread* and *eat soup*, one realizes that *eat* has a slightly different meaning in each phrase: in the latter expression, but not in the former, it covers the manipulation of a spoon. (Weinreich 1972 : 35)

Au plan sémantique, la question de la flexibilité est cruciale mais difficile à circonscrire. Dans une conception large, elle est en effet susceptible de couvrir un continuum qui va de la polysémie (beaucoup d'exemples de flexibilité correspondent en fait à des polysémies systématiques) à la simple variation contextuelle.

Dans le cadre de la philosophie analytique, la flexibilité sémantique est discutée dans la mesure où elle peut, en première approche, remettre en cause le principe de compositionnalité. Par exemple, le verbe *couper* semble prendre un sens différent selon qu'il apparaît dans *couper le gâteau* ou dans *couper les cheveux*, alors que le principe de compositionnalité exige que la contribution sémantique de *couper* soit la même dans les deux syntagmes (Searle 1980). En s'appuyant sur la distinction entre deux processus contextuels différents, saturation vs. modulation, Recanati (2004 ; 2009) montre que ce genre d'exemple se distingue des indexicaux. La forme logique de ces derniers, en effet, contient une variable qui doit être saturée par le contexte. En revanche, il n'y a pas de telle variable dans les cas de flexibilité sémantique, mais un processus de modulation (le terme est repris à [Cruse 1986]) déterminée, selon lui, par une fonction pragmatique.

Dans le cadre de la Grammaire Cognitive, Langacker rend compte d'une partie des phénomènes de flexibilité sémantique à l'aide de la notion de zone active (*active zone*, désormais ZA). Pour l'illustrer, revenons à l'exemple (16a) tiré de Anderson & Ortony (1975 : 170), auquel nous ajoutons quelques variations supplémentaires :[6]

(17) a. *The accountant pounded the desk.* [fist]

b. *The accountant pounded the stake.* [hammer]

[6] A noter que le dernier exemple, (17d), est sans doute un cas de polysémie. Il montre que la séparation entre flexibilité et polysémie relève plus d'un continuum que d'une séparation tranchée.

c. *The accountant pounded the garlic and salt together.* [pestle]

d. *The accountant pounded the Serbian position.* [gun]

Dans ces exemples, nous spécifions l'entité qui subit le processus à l'aide du complément d'objet. En revanche, l'élément au moyen duquel le processus est réalisé n'est pas spécifié. Comme les exemples (17) le montrent, ce dernier peut correspondre à une partie de l'entité dénotée par le sujet de la proposition (le poing du comptable dans [17a]) ou bien à une très grande variété d'instruments (ex. [17b-d]) impliquant des scénarios différents. Dans tous ces exemples, ce n'est pas le comptable à proprement parler qui réalise l'action mais ce que Langacker (2009) appelle la zone active. Se pose alors la question de savoir comment cette zone active, justement, est activée. Anderson & Ortony, pour leur part, ne disposent pas de la notion de ZA et proposent de régler le problème en se fondant sur la notion de voisinage sémantique (*semantic neighborhood*). Pour Langacker, en revanche, l'explication est à rechercher du côté de la notion d'association, association qui doit être légitimée par nos connaissances encyclopédiques :

> the only requirement is that the active zone be *associated* with the nominal referent in some evident fashion. […] The reason is that we are able to make sense of discrepant expressions by exploiting general knowledge. For example, every expression in [(17)][7] evokes a basic scenario, a familiar aspect of everyday life in our culture. This encyclopedic cultural knowledge – not any narrow, dictionary-type definitions of the component lexical items – gives us what we need to properly understand the expressions. (Langacker 2009 : 50)

Cette explication a toutefois le défaut d'être trop puissante, et c'est sous cet angle que Kleiber attaque la question de la flexibilité. Son travail se distingue des autres sur deux points. Premièrement, il prend en compte et discute des travaux issus d'horizons très différents : la notion de forme schématique (Franckel & Paillard 1998; Franckel & Lebaud 1992; Victorri 1997), de transfert de sens (Nunberg 1995), de zone active (que nous venons d'évoquer), de facette sémantique (Croft & Cruse 2004 : 116-140), de coercition de type (Pustejovsky 1993, 1995), etc. Deuxièmement, il porte une attention toute particulière aux problèmes de sur-généralisation et aux contraintes linguistiques qui pèsent sur la flexibilité. Si le sens est flexible, en effet, il ne l'est pas au point d'accepter n'importe quel type de modulation. Par exemple, un énoncé tel que (18) peut porter sur la couleur de l'encre ou la surface du stylo,

[7] Les exemples auxquels se réfère Langacker portent sur la préposition *in*, mais le raisonnement s'applique dans les mêmes termes à (17).

mais des énoncés tels que (19) ne sont pas valides s'ils désignent le vernis des ongles de Marie ou les yeux de Marie.

(18) *Le stylo est rouge.*

(19) a. *Marie est rouge.*
b. *Marie est bleue.*

Kleiber est alors amené à proposer un principe cognitif général apte à rendre compte à la fois de la flexibilité et de ses contraintes : le *principe de métonymie intégrée*, selon lequel « certaines caractéristiques de certaines parties peuvent caractériser le tout » (Kleiber 1995 : 123), du moins, sous certaines conditions de saillance et de pertinence des parties en question (conditions qui ne sont pas respectées dans les exemples [19]). À ce premier principe s'en ajoute un second, le *principe d'intégration méronomique* ou *méronomisation*, selon lequel « le rapport de contiguïté entre deux entités X et Y peut être dans certaines situations transformé en rapport de partie (X)-tout (Y) » (Kleiber 1995 : 128). Il permet de prendre en compte les cas de figure qui ne mettent pas en jeu une relation partie-tout à proprement parler, mais une relation de contiguïté (par exemple, la relation entre le conducteur et sa voiture, dans *Paul est garé en bas*).

La Sémantique Interprétative se pose aussi la question de la flexibilité sémantique, mais adopte une position très radicale en donnant à la flexibilité une extension universelle (bien au-delà de la seule relation partie / tout ou de la contiguïté). Dans ce cadre, la flexibilité est conçue comme étant la règle générale. L'exemple le plus parlant est sans doute celui de Rastier (1991 : 211) à propos du lexème *poisson*. Dans la terminologie de Rosch, il appartient au niveau de base (cf. section 1.1). Cependant, ses attributs sont très différents selon qu'il apparaît dans la conjonction *le canari et le poisson*, où il est inclus dans la classe des animaux domestique (et où on imagine plutôt un poisson rouge), ou bien dans la conjonction *le cormoran et le poisson*, où il fait partie du domaine marin (et où, par exemple, on imagine un hareng). Si, dans les deux cas, le poisson est bien doté d'écailles et de branchies (mais pas d'une crinière…), cela paraît finalement assez secondaire par rapport à ce qu'il convient d'appeler (d'après Rastier) le sens du mot *poisson*.[8] La Sémantique Interprétative est connue pour mener le raisonnement jusqu'au bout, au point d'inverser le rapport qui s'établit habituellement entre le sens d'un mot et son contexte. Dans ce cadre théorique, c'est le contexte qui détermine entièrement le sens des mots, et non l'inverse. Ce point de vue est

[8] A noter que Kleiber ne considère pas qu'il s'agit là d'un changement de sens, mais d'une simple variation interprétative (Kleiber 2008).

résumé dans un slogan célèbre, selon lequel « le global (le texte) détermine le local (les mots) ».

2 Les probabilités et les statistiques en linguistique

Parallèlement à la critique de Fodor & Pylyshyn (1988) qui ouvre la Section 1, et dans laquelle il s'agissait de minimiser le poids de l'inférence statistique dans la cognition, on trouve chez Chomsky un rejet des statistiques lorsqu'elles sont utilisées pour rendre compte de la grammaire. L'argument est bien connu :

> […] la notion de « grammatical en anglais » ne peut être assimilée à celle « d'ordre élevé d'approximation statistique ». On peut dire que ni [*Colorless green ideas sleep furiously*] ni [**Furiously sleep ideas green colorless*] (ni en vérité aucune partie de ces phrases) ne sont jamais apparues dans un discours anglais. Partant, dans tout modèle statistique de la grammaticalité, ces phrases seront mises sur le même plan et considérées comme également « étrangères » à l'anglais. Cependant, [*Colorless green ideas sleep furiously*], bien que dépourvue de sens, est grammaticale, tandis que [**Furiously sleep ideas green colorless*] ne l'est pas. […] Pour prendre un autre exemple, dans le contexte : « I saw a fragile ____ », les mots « whale » et « of » peuvent avoir une fréquence identique (=zéro) dans l'expérience linguistique antérieure d'un locuteur : il reconnaîtra immédiatement que l'une de ces substitutions, mais non l'autre, donne une phrase grammaticale. […] De toute évidence, l'aptitude de quelqu'un à produire et à reconnaître des énoncés grammaticaux n'est pas fondée sur des notions d'approximation statistique ou d'autres de même nature. (Chomsky 1969 : 18-19)

Dans cette optique, la grammaire est conçue comme un ensemble de règles syntaxiques a priori qui génère toutes les phrases grammaticale d'une langue et seulement elles.[9]

[9] Comme le souligne Gardies (1975), il est difficile de ne pas voir le rapport avec la « grammaire pure » (ou « morphologie pure des significations ») que Husserl élabore dans la 4ème Recherche Logique. Le but de cette dernière, en effet, est de garantir l'unité de sens (la grammaticalité, chez Chomsky) de n'importe quelle proposition qui respecte les règles de la grammaire. Chez Husserl, en effet, c'est à ce niveau grammatical que s'opère la séparation entre sens (*cet arbre est vert*) et non-sens (*cet arbre est ou*).

2.1 La fréquence et ses effets

Si, comme l'a montré la Section 1, la fréquence joue un rôle décisif dans le domaine de la psychologie, en est-il de même dans le traitement du langage ? Du point de vue de Chomsky, nous venons de le voir, il n'en est rien. Mais si ce devait être le cas, en quoi la fréquence (d'un phonème, d'un morphème, d'un lexème, d'une catégorie grammaticale, d'une structure syntaxique, etc.) pourrait-elle avoir un impact sur le langage, sa compréhension et / ou sa production ? Une façon simple de répondre à cette question consiste à reprendre la règle de Bayes en (7). Parmi les termes qui la composent, il y a ce qu'on appelle l'a priori, *P(h)*. Nous avons utilisé ce dernier dans les sections précédentes pour modéliser la croyance d'un individu concernant la catégorie des chiens, croyance selon laquelle les chiens portent très rarement une crinière (cf. [13] et [14]). Cette croyance se fondait sur une observation selon laquelle seulement un chien sur dix portait une crinière et avait dû être modifiée lors de la présentation de dix nouveaux chiens (parmi lesquels il s'avérait que trois avaient des crinières).

Au plan linguistique, à quoi pourrait bien correspondre un tel a priori ? Une interprétation possible consiste tout simplement à poser que l'a priori correspond à la fréquence relative d'une unité dans « l'expérience linguistique antérieure d'un locuteur » (pour reprendre les termes de Chomsky). Ainsi, les expériences de décision lexicale (décider si telle chaîne de caractère est un mot un non) montrent de façon très robuste que les mots les plus fréquents sont reconnus plus rapidement. Plus encore, il s'avère que la vitesse de reconnaissance est directement corrélée à la fréquence. Traduit en termes bayésiens, on peut voir les choses de la manière suivante : la plausibilité de l'hypothèse que nous ayons affaire à un mot sachant telle chaîne de caractère, *P(h|d)*, dépend directement de l'a priori, *P(h)*, c'est-à-dire la fréquence de ce mot dans l'expérience linguistique antérieure du locuteur. Prenons un autre exemple portant cette fois sur des questions d'ambiguïté. Le fait que nous ayons observé, à de multiples reprises, que le mot *table* est plus souvent utilisé comme nom (*la table est dans le salon*) plutôt que comme verbe (*je table sur une augmentation*) nous amène à avoir un a priori sur ce mot-forme (qu'on pourrait formuler de la façon suivante : « la probabilité que *table* soit employé comme nom est plus haute que la probabilité qu'il soit employé comme verbe »). Dès lors, dans le cas d'une nouvelle occurrence de *table*, nous considèrerons, à cause de cet a priori, comme plus plausible l'hypothèse selon laquelle il s'agit d'un nom plutôt que d'un verbe. De nombreuses expériences en psycholinguistique permettent de montrer l'existence

d'un tel a priori (par exemple, le temps de lecture est plus long lorsque *table* est utilisé comme verbe).[10]

Les probabilités conditionnelles (cf. (1)) sont elles aussi très utilisées pour rendre compte des effets de fréquence sur les relations entre mots ou entre catégories syntaxiques. Pour illustrer cela, partons de l'exemple classique d'un dès à six faces. Les probabilités nous permettent de calculer les chances de tirer une suite particulière de deux nombres, disons, 3 et 5. Il suffit de calculer la probabilité de chaque sous-événement (en l'occurrence, un sixième) et de les multiplier : un sixième au carré égal 0.028. Imaginons que nous nous dotions d'une langue fictive que nous appellerons, manière de plaisanter, le Schmolblock. Elle contient 50 mots, parmi lesquels 25 sont des noms, 20 sont des verbes, et 5 sont des adverbes. Dans ce contexte, il est tout aussi facile de calculer la probabilité de tirer un verbe et un nom (c'est-à-dire d'avoir un syntagme verbal) :

(20) $P(Verbe, Nom) = P(Verbe) \times P(Nom) = \frac{20}{50} \times \frac{25}{50} = 0.4 \times 0.5 = 0.2$

Cela n'est cependant valable que si nous partons du principe que tirer un verbe et tirer un nom sont deux événements indépendants en Schmolblock, de sorte que le fait de tirer un verbe n'a pas d'influence sur le fait de tirer un nom, et inversement. Or, notre (bonne) connaissance du Schmolblock nous indique que ce n'est pas le cas. En effet, nous disposons d'un corpus de Schmolblock numérisé de dix milliards de mots (disons, le SchTenTen), et lorsque nous l'analysons (après un étiquetage laborieux), nous constatons que dans 90 % des cas, un verbe est précédé d'un nom. En d'autres termes, nous savons que la probabilité d'avoir un verbe sachant que nous avons un nom, c'est-à-dire, $P(Verbe|Nom)$, est de 0.9. Les deux événements (tirer un nom et tirer un verbe) sont deux événements dépendants et pour cette raison, il faut donc partir de la définition de la probabilité conditionnelle :

(21) a. $P(Verbe \mid Nom) = \frac{P(Verbe, Nom)}{P(Nom)}$

b. $P(Verbe, Nom) = P(Nom) \times P(Verbe \mid Nom) = 0.5 \times 0.9 = 0.45$

Le résultat obtenu est ainsi plus élevé que celui calculé en (20), parce que nous intégrons une information supplémentaire, à savoir le rapport de dépendance qui existe en Schmolblock entre le verbe et le nom. Bien entendu, nous aimerions pouvoir étendre ce genre de calcul à des suites contenant plus de deux mots. Par exemple, j'aimerais savoir quelle est la

[10] Dans le même esprit, une autre interprétation possible de la règle de Bayes pourrait consister à faire correspondre l'a priori à l'*entrenchment* (Gréa 2006).

probabilité, en Schmolblock, d'avoir un nom suivi d'un verbe lui-même suivi d'un adverbe. Pour cela, il suffit d'étendre la règle des probabilités conditionnelles à trois événements. Cette opération n'est pas spécialement compliquée, si l'on se rappelle qu'une probabilité conjointe de trois événements est, entre autres, équivalente à la probabilité conjointe du premier événement d'un côté, et du second et du troisième (pris ensemble) de l'autre :

(22) $P(Adv, Verbe, Nom) = P(Adv, (Verbe, Nom)) = P(Verbe, \underbrace{Nom) \times P(Adv}\,|\, Verbe, Nom)$
$= P(Nom) \times P(Verbe \mid Nom) \times P(Adv \mid Verbe, Nom)$ } (Cf. [21b])

La formule (22) est un exemple de ce qu'on appelle une chaîne de Markov et peut se généraliser à n'importe quel nombre d'événements :

(23) $P(A_1, A_2, ..., A_n) = P(A_1) \times P(A_2 \mid A_1) \times ... \times P(A_n \mid A_1, A_2, ..., A_{n-1})$

Une chaîne de Markov est donc un bon moyen de calculer la probabilité d'une séquence (en l'occurrence, une séquence de catégories grammaticales) en fonction d'un lexique (les cinquante mots que contient le Schmolblock), de la fréquence, et surtout, d'un corpus de Schmolblock dont on fait l'hypothèse qu'il reflète correctement les propriétés combinatoires de la langue (ces mêmes propriétés que la grammaire générative de Chomsky a pour tâche de fixer sous la forme d'un système a priori). Elle permet aussi de calculer la probabilité d'une transition, par exemple, *P(Adv|Nom,Verbe)*, c'est-à-dire la probabilité d'avoir un adverbe, sachant qu'on a un nom et un verbe.

Une critique que l'on pourrait adresser à ce genre de raisonnement tient au fait qu'il s'applique à des séquences de mots sans tenir compte de la structure grammaticale des phrases. Mais aujourd'hui, les grammaires probabilistes (ou grammaires stochastiques) intègrent cette information supplémentaire en s'appuyant non plus sur un corpus de textes, mais sur un corpus d'arbres syntaxiques (on désigne ce genre de corpus par le terme de *treebank* « banque d'arbres »).[11]

D'une manière générale, les chaînes de Markov peuvent être utilisées pour n'importe quel type d'unité : catégories grammaticales (comme c'est le cas dans notre exemple de Schmolblock), mais aussi, phonèmes, syllabes, morphèmes, lexèmes, etc. En cela, elles s'avèrent être un modèle psycholinguistique pertinent, par exemple, pour l'acquisition du lexique par l'enfant, en permettant de détecter les frontières de mots. C'est ce qu'ont permis de démontrer de nombreux travaux en psychologie cognitive (Aslin, Saffran, & Newport 1998; Saffran, Aslin, & Newport 1996; Saffran, Newport, & Aslin 1996), travaux que la

[11] Concernant le français, cf. Abeillé, Clément, & Toussenel (2003).

linguistique cognitive et la grammaire de construction se sont réappropriées pour justifier leur propre approche :

> In the past decade, we have witnessed major discoveries concerning children's ability to extract statistical regularities in the input. Children are able to extract word forms from continuous speech based on transitional probabilities between syllables (Saffran, Aslin, & Newport 1996). For example, the phrase bananas with milk, contains four transitional probabilities across syllables (*ba* to *na*; *na* to *nas*; *nas* to *with*; and *with* to *milk*). The probability that *ba* will be followed by *na*, and the probability that (*ba*)*na* will be followed by *nas* is higher than the probability that *nas* will be followed by *with*. That is, transitional probabilities are generally higher within words than across words. Eight-month-old infants are sensitive to these statistical cues (Saffran, Aslin, & Newport 1996) and treat these newly acquired words as part of their lexical inventory (Saffran 2001). (Goldberg 2006 : 70)

2.2 Sémantique et probabilités

Les conceptions probabilistes de la sémantique, plus proches de nos préoccupations, ne datent pas non plus d'hier. On trouve par exemple dès la fin des années 70, sous la plume de Nunberg, une tentative intéressante pour rendre compte de la flexibilité sémantique à l'aide de la *cue-validity* (Nunberg 1979). Son analyse se fonde sur la notion de fonction (*referring function*, désormais RF) et dans l'article de 1979, le problème de la flexibilité se pose de la façon suivante :

> Suppose a language-learner encounters two uses of a word *w*, to refer to both *a* and *b*. And suppose also that he has good reason for supposing that *w* is not simply homonymous - that there is only one convention governing the use of *w* - say by applying some version of the tests we offered earlier in sorting out polysemy and homonymy. And finally, suppose he has no reason to assign *a* and *b* to the same category, no matter how generously he figures things, and so could not argue that *w* was simply 'vague.' […] What theory about the relation between the uses will he arrive at, and why? (Nunberg 1979 : 166-167)

La réponse de Nunberg est que le locuteur utilise la *cue-validity* pour déterminer lequel des deux usages dérive de l'autre (lequel des deux est l'argument d'une RF qui donne le second usage en sortie). Nous allons illustrer le principe à l'aide d'un exemple repris à Nunberg, le mot *stardust* qui désigne à la fois un disque et une chanson. Le locuteur doit faire deux hypothèses distinctes :

- h_1, selon laquelle *stardust* désigne à l'origine un disque (un objet concret)

- h_2, où *stardust* désigne à l'origine une chanson, c'est-à-dire un nom d'idéalité (pour reprendre la terminologie de Flaux [2011] et de Stosic & Flaux [2012], qui la reprennent eux-mêmes à Husserl).

Dans le cas de h_1, le locuteur doit juger la plausibilité que *stardust* désigne une chanson sachant qu'il désigne déjà un disque : *P(chanson|disque)*. Dans le cas de h_2, ce même locuteur doit juger la plausibilité que *stardust* désigne un disque sachant qu'il désigne déjà une chanson (h_2) : *P(disque|chanson)*. Intuitivement, la réponse est relativement simple (précisons que Nunberg n'effectue aucun calcul dans son article). Les attributs de la chanson (son rythme, sa mélodie, etc.) ne dépendent pas du disque, et cette dernière pourrait très bien exister sans même avoir été jamais enregistrée. Par conséquent *P(chanson|disque)* est faible. À l'inverse, les propriétés du disque (le nom imprimé, les sons qui s'y trouvent gravés) sont entièrement dépendantes de la chanson et le disque ne serait pas ce qu'il est si la chanson était différente. En d'autres termes, *P(disque|chanson)* est élevée. La conclusion est donc que la plausibilité de h_2 est plus grande car il est plus rationnel d'utiliser la chanson pour désigner le disque, plutôt que d'utiliser le disque pour désigner la chanson.

Tableau 3.

h_1 : *stardust* désigne un disque	h_2 : *stardust* désigne une chanson
cue-validity de la RF *f(disque) = chanson* ?	*cue-validity* de la RF *g(chanson) = disque* ?
P(chanson|disque) est faible	*P(disque|chanson)* est haute
Conclusion : utiliser le disque pour désigner la chanson est moins rationnel	Conclusion : utiliser la chanson pour désigner un disque est plus rationnel.

Le principe est identique avec un nom tel que *Baudelaire*, qui désigne à la fois un individu (l'auteur) et son œuvre. Selon Nunberg, en effet, *P(œuvre|écrivain)* est élevée dans la mesure où si l'œuvre de Baudelaire est sombre et dépressive, c'est parce que l'écrivain lui-même l'était. À l'inverse, *P(écrivain|œuvre)* est beaucoup plus faible :

> It is hard to say just what criteria are most important for establishing personal identity in folk metaphysics, but they seem to be wrapped up with circumstances of birth, lineage, and physical properties. Caedmon [Baudelaire, dans notre exemple] - that very man - could have died in infancy, taken a vow of silence, or written *David Copperfield* without our ever being tempted to say that he was not the same person. And since in description we have free choice among all possible ways of identifying the man, we could do better here than to pick him out by a property so contingent or accidental as his having written these poems. (Nunberg 1979 : 168)

Tableau 4.

h_1 : *Baudelaire* désigne un écrivain	h_2 : *Baudelaire* désigne une œuvre
cue validity de la RF *h(écrivain) = œuvre* ? *P(œuvre|écrivain)* est haute Conclusion : utiliser l'écrivain pour désigner l'œuvre est plus rationnel	cue validity de la RF *k(œuvre) = écrivain* ? *P(écrivain|œuvre)* est faible Conclusion : utiliser l'œuvre pour désigner l'écrivain est moins rationnel.

L'intérêt d'une telle approche tient en particulier au fait que « that calculations of cue-validity are made against a set of background assumptions that may vary from context to context, from place to place, or from time to time. » (p. 170). Par la suite, cependant, Nunberg (1995) abandonne cette approche probabiliste au profit d'un tout autre point de vue, selon lequel ce n'est plus le référent de *stardust* qui change (où la RF permet de passer d'une chanson à un disque) mais le prédicat sous la portée duquel *stardust* se trouve.

Aujourd'hui, sous l'impulsion de la psychologie et, plus précisément, de la théorie du cerveau statisticien évoquée dans les sections précédentes, nous constatons un retour en force des approches probabilistes de la sémantique. En particulier, on note une mise en œuvre assez récente des probabilités bayésiennes dans le cadre de problèmes classiques issus de la sémantique et de la pragmatique formelle. C'est le cas par exemple, de la question du vague abordée par Lassiter & Goodman (2015), qui proposent un traitement bayésien du paradoxe Sorite (qui se rapproche de la proposition que faisait déjà le mathématicien Borel en 1907 [Égré & Barberousse 2014]), ou encore de l'articulation entre sémantique et pragmatique (Goodman & Lassiter 2015).[12] D'une manière générale, il semble que les probabilités bayésiennes commencent aujourd'hui à être très utilisées pour rendre compte de la flexibilité sémantique et il y a de fortes chances pour que la sémantique formelle, dans les prochaines années, soit confrontée à un changement de paradigme relativement important (un basculement de la sphère logico-déductive vers le raisonnement plausible).

La règle de Bayes est aussi un outil qui pourrait être d'une grande pertinence et d'un grand intérêt dans d'autres cadres théoriques moins formels. Dans Rastier (1987 ; 1991), par exemple, la Sémantique Interprétative s'avère entretenir de profondes affinités avec l'approche bayésienne, comme en témoigne les citations suivantes :

> Cette hétérogénéité entraîne que le type de « vérité » auquel peut parvenir une sémantique interprétative n'est pas de l'ordre du vrai même relatif, mais du plausible, entendu comme un compromis entre les prescriptions et les licences de

[12] C'est à S. Loiseau que nous devons de connaître l'existence du laboratoire CoCoLab, du département de psychologie de Stanford, qui est très actif sur ce segment émergent. Le lecteur intéressé peut consulter la liste des publications au lien suivant :
https://cocolab.stanford.edu/index.html

> divers systèmes. Dès lors, lire un texte ne consiste pas seulement à énoncer une ou plusieurs isotopies, mais encore à évaluer leur plausibilité relative. Cela suppose aussi qu'on puisse produire des critères purement sémantiques pour récuser des lectures, et rejeter la thèse post-moderne que pour tout texte digne de ce nom il existe une infinité de lectures possibles. (Rastier 1987 : 12)

> Pour une sémantique interprétative l'équivocité est une donnée fondamentale. Dans le meilleur des cas, on peut établir qu'une interprétation est préférable à toutes les autres. En d'autres termes, et bien que toute notre tradition herméneutique milite contre cette conclusion, le sens d'un texte n'est pas de l'ordre du vrai, mais du plausible. Plutôt que de révoquer les interprétations jugées impropres, il convient donc de les hiérarchiser, en graduant leur plausibilité relativement à une stratégie donnée. (Rastier 1991 : 160)

Transposée au niveau mathématique, la question ne porte plus sur la probabilité d'un attribut sachant une entité donnée, *P(attribut|entité)*, comme c'est le cas pour la psychologie, mais sur la probabilité d'un sens sachant un contexte donné, *P(sens|contexte)*. Cela change considérablement la donne, mais la complique aussi beaucoup. Pour calculer l'a posteriori *P(sens|contexte)* dans un cadre bayésien, il faudrait savoir comment modéliser l'espace des a priori, *P(sens)*, la vraisemblance, *P(contexte|sens)*, et enfin l'évidence *P(contexte)*, ce qui n'est pas une mince affaire. Cela constitue toutefois un axe de recherche promis à un grand avenir. Certains termes de l'équation bayésienne posent moins de problèmes que d'autres. Par exemple, l'a priori *P(sens)* pourrait être représenté à l'aide d'un graphe de cooccurrence construit à partir d'un dictionnaire monolingue (Loiseau, Gréa, & Magué 2010). La vraisemblance *P(contexte|sens)*, quant à elle, pourrait être calculée à partir d'un graphe de cooccurrence construit à partir d'un contexte donné. Ce genre d'approche permettrait alors d'identifier les usages peu probables d'un mot (une métaphore innovante, par exemple, où l'a posteriori *P(sens|contexte)* serait significativement faible). Une approche similaire pourrait permettre d'associer une plausibilité à des isotopie et réaliser l'objectif formulé par Rastier, à savoir « hiérarchiser les interprétations ». Bien évidemment, à cette étape de notre réflexion, il ne s'agit là que d'intuitions qui demandent à être concrétisées, testées et le cas échéant, amendées.

3 Affinités et mesures d'association

3.1 L'information mutuelle

Dans un article célèbre, Church & Hanks (1990) proposent d'utiliser la notion d'information mutuelle dans le cadre des statistiques textuelles. Celle-ci va nous permettre d'introduire la dernière section de ce travail. Avant d'entrer dans le détail de la méthode et de la mettre en relation avec ce qui précède, le point de départ du raisonnement mérite d'être rappelé. Le type de phénomène que les auteurs ont initialement en tête est l'amorçage, un facteur fondamental de la psychologie cognitive. Dans une tâche de décision lexicale, par exemple, les locuteurs réagissent plus vite que la normale si le mot cible (par exemple, *beurre*) est précédé d'un autre mot qui lui est « sémantiquement associé » (*pain*), et moins vite que la normale lorsque le mot cible est précédé par un mot « sémantiquement éloigné » (*infirmière*). Les auteurs proposent alors de modéliser cette relation d'association entre deux mots *x* et *y* à l'aide des probabilités, en comparant deux mesures distinctes : la probabilité conjointe *P(x,y)*, c'est-à-dire la probabilité d'observer les mots *x* et *y* ensemble dans une fenêtre de *w* mots (*w* pouvant varier de 0 – *x* et *y* sont alors contigus – jusqu'à la dimension d'un paragraphe, par exemple), et la probabilité d'observer *x* et *y* indépendamment *P(x)P(y)*. C'est l'information mutuelle :

(24) $$I(x,y) = \log_2 \frac{P(x,y)}{P(x)P(y)}$$

Le logarithme binaire (log_2) de 1 est égal à 0 de sorte que lorsque $P(x,y)=P(x)P(y)$ alors l'information mutuelle $I(x,y)$ est égale à 0. Dès lors, trois situations sont possibles :

> Informally, mutual information compares the probability of observing x and *y together* (the joint probability) with the probabilities of observing x and *y independently* (chance). If there is a genuine association between x and y, then the joint probability $P(x,y)$ will be much larger than chance $P(x)\ P(y)$, and consequently $I(x,y) >> 0$. If there is no interesting relationship between x and y, then $P(x,y) \approx P(x)\ P(y)$, and thus, $I(x,y) \approx 0$. If x and y are in complementary distribution, then $P(x,y)$ will be much less than $P(x)\ P(y)$, forcing $I(x,y) << 0$. (Church & Hanks 1990 : 23)

Dans un second article, Church, Hanks, Hindle, & Gale (1991) montrent tout l'intérêt de cette méthode pour les mots quasi-synonymes en l'appliquant au cas de *strong* et *powerfull*.

Prenons cet exemple afin d'en illustrer le fonctionnement. Dans le corpus *1988 Associated Press newswire* (dont la taille, notée *N*, est de 44,3 millions de mots), les auteurs observent 1 984 occurrences de *powerfull* et 388 occurrences de *legacy*. En outre, *powerfull* et *legacy* apparaissent ensemble (*powerfull legacy*) 7 fois (pour *w*=0). Sur la base de ces données, il est possible de calculer l'information mutuelle de *powerfull* et *legacy* :

$$(25)\quad I(powerfull, legacy) = \log_2 \frac{P(powerfull, legacy)}{P(powerfull)P(legacy)} = \log_2 \frac{\frac{7}{44300000}}{\frac{1984}{44300000} \times \frac{388}{44300000}} = 8.654$$

En d'autres termes, *powerfull* et *legacy* entretiennent une relation d'association significative. Par la suite, les auteurs présentent plusieurs autres méthodes complémentaires (*t-score*, qui mesure non pas l'association mais la répulsion entre deux mots, *Maximum Likelihood Estimator*, *Expected Likelihood Estimator*, *Good-Turing Estimator*, etc.) que nous ne détaillerons pas ici. Ce qui nous importe, ce sont les avancées apportées par ces dernières pour l'analyse linguiste :

> How can a lexicographer make use of statistics of this kind? Two possibilities are immediately apparent. In the first place, they might encourage lexicographers to sharpen the focus of definitions, highlighting salient facts and omitting the remote possibilities that occur only to nervous lexicographers, anxious to cover all possible eventualities. In the second place, they might be used to formulate explicit rules for choosing among near synonyms. When is it better to talk about *strong support*, and when is *powerful support* more appropriate? (Church, Hanks, Hindle, & Gale 1991)

La seconde possibilité est celle sur laquelle nous allons nous focaliser dans la suite de ce travail. De telles mesures d'association permettent en effet d'extraire des données qui constituent la voie royale pour une caractérisation sémantique des synonymes. Par exemple, dans le cas de *strong* et *powerfull*, le calcul de l'information mutuelle permet aux auteurs d'arriver à la conclusion suivante :

> An important criterion for differentiation seems to be that *strong* tends to denote an intrinsic quality, whereas *powerful* appears to be extrinsic, referring more to the effect on others or on the external world. Any worthwhile politician or cause can expect *strong supporters*, who are enthusiastic, convinced, vociferous, etc. But far more valuable are *powerful supporters*, who will brings others with them. (Church, Hanks, Hindle, & Gale 1991)

Une partie importante de notre propre travail consiste, sur le même modèle que celui que nous venons d'exposer, quoique à l'aide d'une méthode différente (cf. Section 3.3) et d'un cadre sémantique précis (la Grammaire Cognitive de Langacker), à examiner un certain nombre de synonymes proches et à en tirer des conséquences plus générales sur la nature, par exemple, du pluriel ou de la localisation spatiale.

3.2 L'analyse collostructionnelle

A partir du début des années 2000, une série de méthodes, réunies sous le terme de *Collostructional Analysis* (« Analyse Collostructionnelle », *collostructional* étant un mot-valise construit sur *construction* and *collocational*) voient le jour à l'intérieur du cadre des *Construction Grammars* (« Grammaires de constructions »), principalement sous l'impulsion de Gries et Stefanowitsch. Cette famille de méthodes se subdivise en trois sous-classes (Gries 2015 : 2; Stefanowitsch 2013) : *collexeme analysis* (Stefanowitsch & Gries 2003), *distinctive collexeme analysis* (Gries & Stefanowitsch 2004b), and *co-varying collexeme analysis* (Gries & Stefanowitsch 2004a; Stefanowitsch & Gries 2005).

L'idée centrale des Grammaires de Constructions consiste à poser qu'il n'y a pas de séparation entre le niveau lexical (les unités lexicales) et le niveau grammatical (les patrons syntaxiques). Les premiers comme les seconds doivent être conçus comme des signes, c'est-à-dire des paires forme / sens. La seule différence entre les deux niveaux tient à leur degré de spécificité. Une unité lexicale comme *maison* est très spécifique, tandis qu'un patron syntaxique tel que <Adj comme GN> (par exemple, *malin comme un singe*) l'est beaucoup moins. Dans ce contexte théorique, l'analyse collostructionnelle se distingue de l'information mutuelle par le type d'unité pris en compte. Si, dans les travaux de Church, il s'agit de quantifier le degré d'association (ou de répulsion) entre lexèmes, l'analyse collostructionnelle se donne plusieurs possibilités supplémentaires :

- Analyse collexémique : il s'agit de comparer une unité lexicale et un patron syntaxique. Par exemple Stefanowitsch & Gries (2003) s'intéresse aux verbes préférentiellement associés à la construction ditransitive en anglais (*John sent Mary the book*).

- Analyse collexémique distinctive : contrairement aux travaux de Church, on compare des constructions proches plutôt que des mots.[13] Par exemple, Gries & Stefanowitsch (2004b) s'intéressent au degré d'association existant entre un verbe et la construction ditransitive par opposition à une autre construction proche, la construction dative (*John sent the book to Mary*).

- Analyse collexémique covariante : il s'agit de comparer deux unités fonctionnellement dépendantes en raison de leur appartenance à une même construction. Par exemple, Desagulier (2015) s'intéresse au degré d'attraction et de répulsion entre les Adj et les N dans la construction <Adj comme GN>.

Depuis Gries (2013), une autre mesure vient s'ajouter à ces différentes méthodes : Δ*P* (Ellis 2006). Elle permet de prendre en compte le caractère asymétrique de la relation d'association. Par exemple, dans l'expression *mad as a hatter* (« fou comme un chapelier »), « Il apparaît que *mad* n'est pas un bon indice de *hatter* tandis que *hatter* est un excellent indice de *mad*. » (Desagulier 2015 : 116).[14]

Ces différentes méthodes entretiennent toutes un lien étroit avec les concepts que nous avons évoqués dans première section :

> Ultimately, collostruction strengths are based on (i) the conditional probabilities *p*(word|construction) and *p*(construction|word), which are related to notions of cue-validity, cue reliability (cf. Goldberg [2006 : Ch. 5-6] and Stefanowitsch [2013]), associative learning measures such as Δ*P*, and prototype formation, and (ii) the frequencies that give rise to the probabilities, which are correlated with entrenchment. Put yet another way : "it is assumed […] that the statistical associations found in the data are reflected in psychological associations in the mind of the language user" (Stefanowitsch 2006: 258). (Gries 2012)

Il y a donc bien une convergence d'intérêt entre psychologie cognitive et linguistique cognitive sur la question des probabilités et des statistiques, et de leur capacité à rendre compte, respectivement, de la cognition et du langage, par opposition au point de vue computationel (cf. citation de Fodor & Pylyshyn [1988: 68]) et génératif (Chomsky 1969 : 18-19).

[13] « We propose a similar method for the analysis of alternating pairs, differing from Church et al. in that we look at near-synonymous (or functionally near-equivalent) constructions rather than words, and that we focus on words appearing in particular slots in these constructions rather than at all words within a given span (we refer to such words as *collexemes* of the construction(s) in question). » (Gries & Stefanowitsch 2004b)

[14] A noter que ce problème était déjà évoqué dans le premier article de Church : « Technically, the *association ratio* is different from *mutual information* in two respects. First, joint probabilities are supposed to be symmetric: *P(x,y) = P(y,x)*, and thus, mutual information is also symmetric: *I(x,y) = I(y, x)*. However, the association ratio is not symmetric, since f(x, y) encodes linear precedence. (Recall that f(x, y) denotes the number of times that word x appears *before y* in the window of w words, not the number of times the two words appear in either order.) » (Church & Hanks 1990 : 24)

3.3 L'exception culturelle française : le calcul des spécificités

Dix ans avant Church & Hanks (1990) et plus de vingt avant Stefanowitsch & Gries (2003), Lafon (1980) élabore une mesure d'association connue sous le nom de « calcul des spécificités » (Habert 1985; Labbé & Labbé 2001; Lafon 1980, 1984; Lebart & Salem 1994; Salem 1987). Elle s'inscrit dans la tradition de l'analyse du discours et son domaine d'application initial est le texte. Dans son article de 1980, par exemple, Lafon applique le calcul des spécificités à un corpus composé de dix discours prononcés par Robespierre à la Convention nationale entre les mois de novembre 1793 et juillet 1794. Ce corpus se divise donc très naturellement en 10 sous-parties (chaque discours) d'inégale longueur. La question posée par Lafon consiste à savoir si la fréquence d'une forme à l'intérieur d'une sous-partie du corpus (un discours pris parmi les dix) est proche de la fréquence attendue au regard de sa fréquence totale (sa fréquence dans les dix discours), de la taille du corpus total, et de la taille du sous-corpus, ou bien si, au contraire, elle est plus fréquente ou moins fréquente qu'attendu.

> Par exemple : peuple (296), forme commune [i.e. relativement fréquente], figure 53 fois dans D5 (longueur 7 896 items), mais 14 fois seulement dans D4 (6 903) qui est à peine plus court. Ce partage nous pousse à dire que peuple apparaît beaucoup dans D5, et peu dans D4. Patriotes (81) apparaît 2 fois dans Dl (8 395) mais 23 fois dans D4 (6 903), et nous formulons volontiers un jugement analogue à propos de cette forme. Sur quoi se fondent de telles appréciations ? Sur une idée spontanée mais qui n'en est pas moins précise : le partage de la fréquence totale en parts proportionnelles aux longueurs des parties. Dans ce cas, la forme nous paraît équitablement partagée entre les parties. (Lafon 1980 : 136)

Le calcul des spécificités permet de quantifier notre intuition sur la répartition des formes à l'intérieur d'un corpus. Par la suite, elle est mise en œuvre dans le domaine de l'analyse du discours pour identifier le vocabulaire sur- ou sous-représenté dans différents genres discursifs : politique, littéraire, philosophique, etc. Habert (1985 : 129) fait une revue des différents corpus auxquels cette méthode a été appliquée :[15]

> -R. Benoît (1981) étudie un échantillonnage d'éditoriaux des Cahiers du bolchevisme (devenus Cahiers du communisme), revue du comité central du PCF, divisé en trois sous-parties de textes groupés autour d'un événement leur donnant homogénéité : congrès de Paris (mars 1932), congrès de Villeurbanne (janvier 1936) et congrès de Paris (juin 1945).

[15] Pour les références complètes, nous renvoyons à l'article de Habert, accessible au lien suivant : http://www.persee.fr/web/revues/home/prescript/article/mots_0243-6450_1985_num_11_1_1207

-S. Bonnafous (1980) prend pour corpus les sept motions des divers courants au congrès de Metz du Parti socialiste (1979), mais s'attache aux motions principales (Mitterrand, Rocard, CERES).
-M. Boutrolle-Caporal (1982) compare 12 récits pour enfants de P. Gripari, M. Tournier, J. Held, M. et R. Farré, C. Grenier et G. Chaulet.
-A. Geffroy (1980) examine les articles de journalistes Hébert, Roux et Leclerc qui veulent prendre la succession politique de Marat, assassiné le 13 juillet 1793.
-M.-R. Guyard (1981) s'attache aux spécificités des principaux auteurs du Surréalisme au service de la Révolution.
-B. Habert (1982) compare les 19 Résolutions générales des congrès confédéraux de la CFTC de 1945 à 1964 et de la CFDT de 1964 à 1979.
-P. Lafon (1980) utilise pour sa démonstration 10 discours prononcés par Robespierre à la Convention nationale entre novembre 1793 et juillet 1794.
-D. Peschanski (1981 a) contraste 9 regroupements d'articles-leaders de L'Humanité autour de dates porteuses de forte expression politique, entre février 1934 et août 1936.
-D. Peschanski (1981 b) met en regard 326 articles-leaders de L'Humanité de 1934 à 1936, regroupés par mois, soit 32 parties.
-M. Tournier (1975) étudie les Pétitions ouvrières de 1848 au sein d'un ensemble comprenant à la fois d'autres textes de la même période (« Romantiques », « L'Atelier », Louis Blanc) et des textes révolutionnaires (Robespierre et Hébert).

-La parole syndicale étudie les Résolutions votées par les congrès confédéraux et les instances intermédiaires (comités ou conseils confédéraux ou nationaux) de la CGT, de la CFDT, de la CFTC et de FO de 1971 à 1976.

C'est peut-être en raison de son fort ancrage discursif (qui, pour le coup, est une spécificité française) que cette méthode n'a jamais été diffusée au-delà frontières de l'espace francophone. Pourtant, sa philosophie générale est extrêmement proche de celle qui motive l'analyse collostructionnelle (à quelques détails près sur lesquels nous revenons ci-dessous) et elle s'avère parfaitement transposable dans le domaine sémantique. Depuis Gréa (2008), nous utilisons systématiquement cette méthode pour étayer nos analyses. Dans ce qui suit, nous proposons de l'illustrer à l'aide d'un exemple tiré de Gréa (2015), les prépositions *entre* et *parmi* qui sont habituellement considérés comme de proches synonymes (Ashino 2007; Franckel & Paillard 2007; Guentchéva 2003; Hilgert 2007, 2009, 2010, 2013; Kwon-Pak 2006; Van Goethem 2009).

La première étape du processus consiste à extraire les occurrences qui nous intéressent à partir d'un corpus déterminé (en l'occurrence, Frantext, désormais FR, et Le Monde, LM). La plupart du temps, nous avons recours au logiciel Nooj (Silberztein 2003, 2004), qui, dans le cas présent, nous a permis d'extraire (entre autres choses) l'ensemble des N_{pl} qui apparaissent

dans les constructions <entre SN_{pl}> et <parmi SN_{pl}>. Le Tableau 5 présente le nombre d'occurrences des deux constructions dans les deux corpus fusionnés (FR+LM).[16]

Tableau 5. Tableau Lexical Entier

	FR+LM
<entre SN_{pl}>	120 502
<parmi SN_{pl}>	55 330
Total	175 832

La seconde étape du processus fait intervenir le calcul des spécificités.[17] Ce dernier repose sur une loi probabiliste, la loi hypergéométrique. Il s'agit d'une loi de distribution permettant de décrire les probabilités d'un résultat de tirage sans remise. Elle se caractérise par trois paramètres : T est la longueur totale du corpus (dans le cas présent, T = 175 832, cf. Tableau 5) ; t est la longueur d'une sous-partie de ce corpus (par exemple, pour la sous-partie correspondant à <entre SN_{pl}>, t = 120 502) ; f est la fréquence d'une forme particulière (un N_{pl} pris parmi l'ensemble des N_{pl} qui entrent dans les deux constructions). Pour comprendre comment fonctionne cette loi, prenons l'exemple du lexème *parties*. Ce nom a 1 751 occurrences dans FR+LM (f = 1 751). Le Tableau 6 présente la répartition de ces 1 751 occurrences à l'intérieur des deux constructions.

Tableau 6. Répartition des occurrences de *parties*

N_{pl}	<entre SN_{pl}>	<parmi SN_{pl}>	Total
parties	1 738	13	1 751

Nous voulons savoir si cette distribution est conforme à la situation d'indépendance statistique (la situation attendue si la distribution de *parties* obéissait au hasard) ou si, au contraire, *parties* est sur-employé (ou sous-employé) dans une construction donnée, en tenant compte du nombre d'occurrences de chaque construction dans le corpus (cf. Tableau 5). Plus précisément, nous voulons savoir quelle est la probabilité d'avoir 1 738 occurrences de *parties* après *entre* sachant que, premièrement, ce nom apparaît aussi 13 fois après *parmi*, deuxièmement, que la construction <entre SN_{pl}> a une taille de 120 502 occurrences et troisièmement, que le corpus total à une taille de 175 832 occurrences.

Intuitivement, nous nous doutons que la répartition de *parties* est ici très déséquilibrée et que ce nom est sur-employée après la préposition *entre*, même si nous prenons en compte la

[16] Une erreur s'est glissée dans le Tableau Lexical de (Gréa 2015). Le tableau correct qui a été utilisée pour calculer les spécificités est celui qui figure ici (Tableau 5).

[17] La présentation qui suit reprend sans grands changements, mais à l'aide d'exemples différents, celles qui sont faites dans (Gréa 2017; Gréa & Haas 2015).

différence de tailles des deux sous-corpus. La loi hypergéométrique est un moyen de traduire cette intuition en termes mathématiques. En effet, notre problème est équivalent à un problème de tirage sans remise dans lequel une urne contiendrait 175 832 boules (T, la taille totale du corpus). 1 751 d'entre elles sont blanches (les 1 751 occurrences de *parties*, i.e. le paramètre *f*) et toutes les autres sont noires (tous les N_{pl} autres que *parties*). Notre question se formule alors de la façon suivante : en tirant 120 502 boules dans cette urne (soit *t*, le nombre d'occurrences de la construction <entre SN_{pl}>), quelle probabilité avons-nous de tomber sur 1 738 boules blanches (la valeur effectivement observée) ?

C'est ce que la loi hypergéométrique nous permet de calculer. Pour cela, nous utilisons un script R réalisé par B. Desgraupes (Modal'X, Paris Ouest Nanterre la Défense) et S. Loiseau (Université de Paris 13).[18] La Figure 2a donne la probabilité (en ordonnée) de tirer un certain nombre *k* de boules blanches (en abscisse). Dans la situation que nous venons de décrire, il apparaît que le résultat le plus probable est de tirer 1 200 boules blanches (c'est ce qu'on appelle la valeur modale), tandis que la probabilité d'en tirer 1 738 est proche de zéro. Comme le souligne Lafon (1980 : 141), il faut toutefois préciser notre intuition. La valeur observée est 1 738. Elle semble donc élevée et elle aurait pu l'être davantage encore (1 739, 1 740, …). Ce qui nous intéresse réellement, ce n'est donc pas la probabilité que *parties* apparaisse *exactement* 1 738 fois dans la construction <entre SN_{pl}>, mais que *parties* y apparaisse *au moins* 1 738 fois. En d'autres termes, ce qui nous intéresse, c'est la somme des probabilités : $Prob_{1738}$ + $Prob_{1739}$ + $Prob_{1740}$, etc.[19] Pour obtenir cette mesure, il faut passer de la fonction de masse (Figure 2a) à la fonction de répartition (Figure 2b). La figure ainsi obtenue n'est cependant pas le meilleur moyen de représenter notre problème : nous nous situons dans la « queue » (droite) de la courbe. Une façon « d'épaissir » cette queue consiste à utiliser le logarithme népérien. Cette fonction, en effet, a cette caractéristique de tendre vers -∞ lorsqu'une valeur se rapproche de 0, et d'être égale à 0 lorsque sa valeur est égale à 1 (rappelons qu'une probabilité est toujours comprise entre 0 et 1).[20] Nous obtenons ainsi la Figure 2c. Une dernière opération consiste en un simple changement de signe (Figure 2d). Les résultats supérieurs à la valeur modale (il y a plus de boules blanches qu'attendu) seront désormais positifs (on prend la valeur absolue) et seront désormais appelées des « spécificités

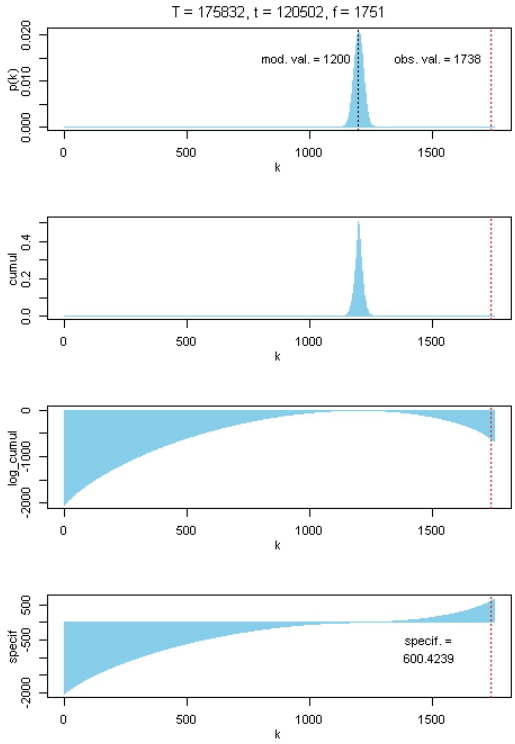
[18] Accessible au lien suivant : https://r-forge.r-project.org/R/?group_id=1547

[19] Ce point constitue la principale différence entre le « calcul des spécificités » et la *multiple distinctive collexeme analysis*.

[20] A ne pas confondre avec le logarithme binaire utilisé par Church pour le calcul de l'information mutuelle, cf. (24).

positives » (Lafon 1980 : 142). Dans cette optique, on dira que *parties* est une spécificité positive de la construction <entre SN_{pl}> (par opposition à la construction <parmi SN_{pl}>).

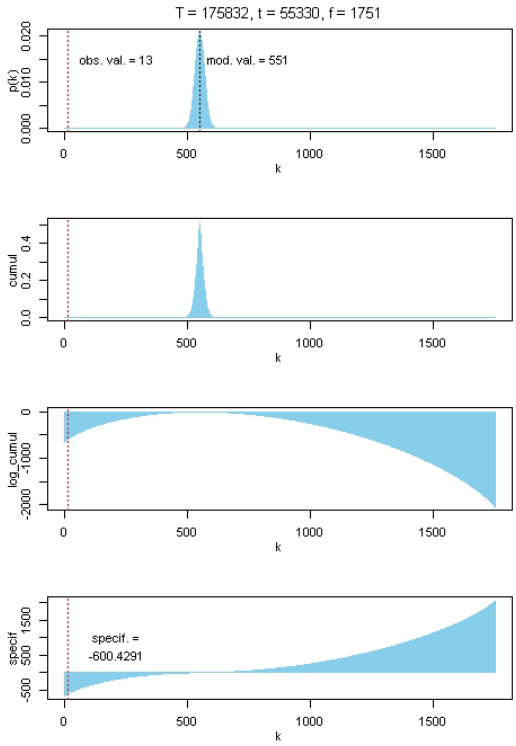
Figure 2. Spécificités de *parties* pour <entre SN_{pl}>



Qu'en est-il, maintenant, de *parties* pour la construction <parmi SN_{pl}> ? Comme on peut le voir dans la Figure 3, nous sommes dans la situation inverse où le résultat est inférieur à la valeur modale (il y a moins de boules blanches qu'attendu). Dans ce second cas de figure, nous calculons la somme des probabilités : $Prob_0$ + $Prob_1$ + $Prob_2$, etc. jusqu'à la valeur observée (en l'occurrence, 13). L'événement qui nous intéresse, en effet, n'est pas que « *parties* apparaisse *exactement* 13 fois après *parmi* », mais que « *parties* apparaisse *au plus* 13 fois après *parmi* ».[21] Dans ce second cas de figure, on laisse leur valeur négative aux résultats et nous appelons ces derniers des « spécificités négatives » (Lafon 1980 : 142). Dans cette optique, le lexème *parties* est une spécificité négative pour la construction <parmi SN_{pl}>.

[21] Il est à noter que les spécificités positives (relation d'association) et négatives (relation de répulsion) sont obtenues à l'aide de la même méthode (le calcul des spécificités). Dans le cas de l'information mutuelle, en revanche, il est nécessaire de recourir à une seconde méthode pour retrouver les exemples de répulsion significatifs, à savoir le t-score, cf. Church, Hanks, Hindle, & Gale (1991 : Section 2.2).

Figure 3. Spécificités de *parties* pour <parmi SN_{pl}>



On interprète l'ensemble de ces résultats de la façon suivante : il y a une grande probabilité pour que du point de vue du lexème *parties*, les constructions <entre SN_{pl}> et <parmi SN_{pl}> ne s'apparentent pas à un tirage aléatoire. On constate au contraire une forte surreprésentation de *parties* dans la construction <entre SN_{pl}> (en d'autres termes, une forte attraction ou association entre *parties* et *entre*) et inversement, une forte sous-représentation de *parties* dans <parmi SN_{pl}> (une forte répulsion entre *parties* et *parmi*).

Portons maintenant notre attention sur un nouveau lexème : *experts*, dont la répartition est donnée dans le Tableau 6.

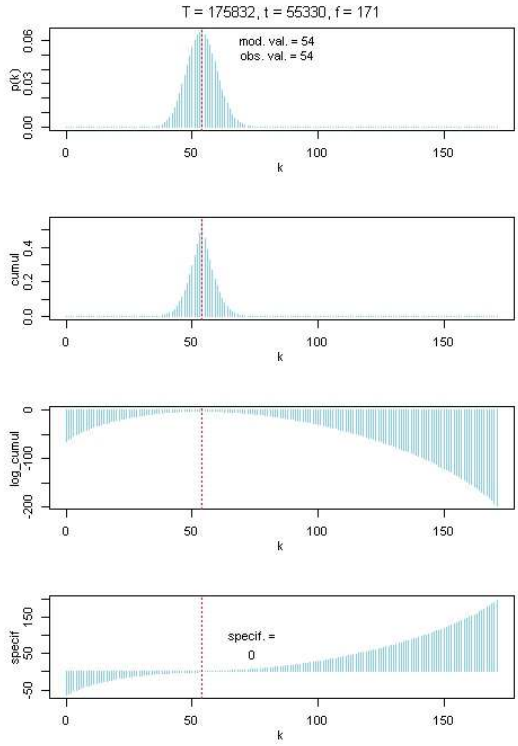
Tableau 7. Répartition des occurrences de *experts*

N_{pl}	<entre SN_{pl}>	<parmi SN_{pl}>	Total
experts	117	54	171

Contrairement aux cas précédents, nous avons l'intuition que la répartition est ici mieux équilibrée. La Figure 4 permet de confirmer ce jugement. Elle démontre que la valeur observée est exactement égale à la valeur modale, c'est-à-dire la valeur attendue dans le cas d'un tirage au hasard (Figure 4a). Le nombre d'occurrences de *experts* dans les constructions <entre SN_{pl}> et <parmi SN_{pl}> est donc celui qui est attendu dans la situation d'indépendance statistique. On dira que *experts* est une forme « banale » pour *entre* et *parmi*. Le cas où la valeur observée est exactement égale à la valeur modale est toutefois exceptionnel. La plupart

du temps, une forme banale se rapproche plus ou moins de la valeur modale. Par conséquent, la séparation entre les formes banales et les formes spécifiques est une question de seuil. Il est fixé de façon arbitraire. Dans la plupart de nos travaux, nous considérons qu'une forme est banale lorsqu'elle se situe dans l'intervalle de spécificité [-15, 15].[22]

Figure 4. Spécificités de *experts* pour <parmi SN_{pl}>



Venons en maintenant aux limites de la méthode. Elles concernent les formes de faible fréquence. Pour l'illustrer, intéressons-nous au N_{pl} *échelons*. Il a seulement 20 occurrences dans notre corpus et ces dernières sont toutes après *entre* (cf. Tableau 8). Malgré cette répartition déséquilibrée, *échelons* apparaît comme étant une forme banale pour les constructions <entre SN_{pl}> et <parmi SN_{pl}>. Il a en effet une spécificité positive de 7,1 pour <entre SN_{pl}> et une spécificité négative de -6,96 pour <parmi SN_{pl}> (cf. Figure 5). Or, notre intuition de locuteur nous indique que *entre les échelons* est plus naturel que *parmi les échelons*, sans que cela se traduise par une spécificité négative inférieure à -15. Cette situation s'explique par la faible fréquence de *échelons*. Comme le montre la Figure 5, le nombre d'occurrence de *échelons* dans notre corpus n'est pas suffisamment élevé pour conférer une réelle significativité à l'absence d'occurrences après *parmi*. Cette limite théorique de la

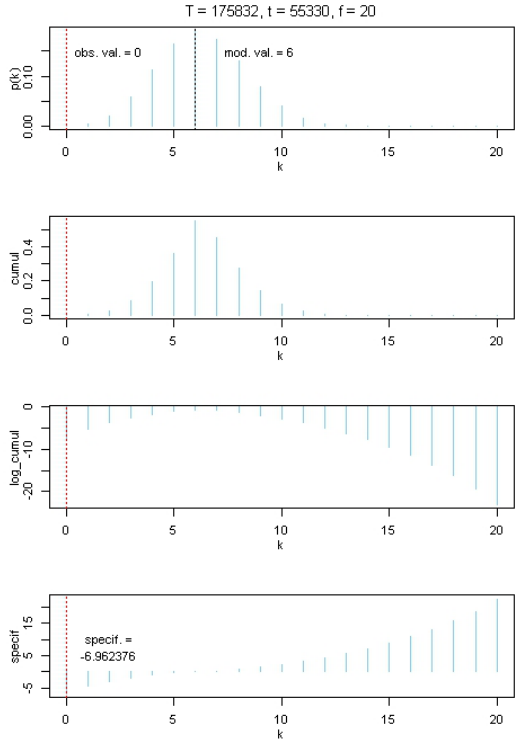
[22] Il s'agit d'un intervalle prudent qui permet de ne pas augmenter artificiellement le nombre de formes typiques d'une construction. Nous avons toutefois fait deux exceptions à cette règle, dans (Gréa 2012; Gréa & Moline 2013), où certains sous-corpus ont des tailles trop réduites pour donner lieu à des mesures de spécificité importantes.

méthode est examinée par Salem (1987) qui propose de poser un « seuil d'absence spécifique » et Labbé & Labbé (2001) qui proposent « d'associer [à cette méthode] un seuil minimal de fréquence en dessous duquel le calcul ne sera pas effectué ».

Tableau 8. Répartition des occurrences de *échelons*

N_{pl}	<entre SN_{pl}>	<parmi SN_{pl}>	Total
échelons	20	0	20

Figure 5. Spécificités de *échelons* pour <parmi SN_{pl}>



Une fois le calcul est réalisé pour tous les N_{pl} du corpus, nous présentons le résultat sous forme de listes appelées « tables de spécificités », c'est-à-dire la liste des N_{pl} les plus spécifiques de chaque construction (par opposition à l'autre construction). À titre d'exemple, nous reproduisons dans ce qui suit une partie des tables de spécificité tirées de (Gréa 2015). Le Tableau 9 donne la liste des 10 N_{pl} les plus spécifiques de la construction <entre SN_{pl}> (par opposition à <parmi SN_{pl}>), et le Tableau 10 donne les 10 N_{pl} les plus spécifiques de <parmi SN_{pl}> (par opposition à <entre SN_{pl}>).

Tableau 9. Liste des 10 N_{pl} les plus spécifiques de <entre N_{pl}> (FR et LM confondus)

Rang	Mot-forme	Fréq. après *entre*	Fréq. après *parmi*	Fréq. tot. (f)	Spécificités
1	mains	7122	0	7122	2758.96009657467
2	pays	7280	783	8063	1132.11483816603
3	parties	1738	13	1751	600.423935784374
4	tours	1266	5	1271	454.597286305256
5	doigts	1144	1	1145	427.439929488243
6	partenaires	1511	98	1609	319.645641923876
7	communautés	994	18	1012	308.909285138729
8	bras	839	4	843	298.129251334098
9	dents	775	1	776	287.469320619896
10	murs	766	3	769	274.850982960915

Tableau 10. Liste des 10 N_{pl} les plus spécifiques de [parmi N_{pl}] (FR et LM confondus)

Rang	Mot-forme	Fréq. après *entre*	Fréq. après *parmi*	Fréq. tot. (f)	Spécificités
1	autres	312	3343	3655	2985.28416224684
2	premiers	27	762	789	778.938517620608
3	meilleurs	8	403	411	431.80679477279
4	victimes	33	440	473	404.597242381022
5	derniers	21	375	396	361.929043228305
6	mesures	25	278	303	246.777501617963
7	nouveautés	4	203	207	217.696133659673
8	invités	25	245	270	211.565278376753
9	priorités	7	202	209	207.055395964336
10	clients	38	242	280	185.229284050551

La troisième et dernière étape du processus, qui constitue en réalité l'étape la plus difficile, mais qui se trouve souvent délaissée dans les travaux contemporains issus de la linguistique de corpus, est l'interprétation sémantique de ces résultats. Les tables de spécificités ci-dessus constituent une sorte de synthèse des grandes tendances distributionnelles de *entre* et *parmi*, mais elles ne donnent aucune information sur les principes sémantiques qui dirigent et conduisent ces tendances. Il reste donc à les découvrir, ce qui n'est pas toujours très facile. Dans le cas de figure qui nous intéresse, nous avons montré que le principe sémantique qui détermine les distributions de *entre* et de *parmi* tient à la question de l'interdépendance ou de la non interdépendance fonctionnelle des N_{pl}. Les N_{pl} spécifiques de *entre* sont en majorité des noms qui dénotent les parties d'un tout : *mains*, *parties*, *doigts*, mais aussi *barreaux* ou *échelons*, c'est-à-dire des constituants fonctionnellement dépendants les uns des autres du fait de leur appartenance à un même tout.

Depuis 2008, nous avons pu mesurer les avantages qu'une telle méthode présente en termes d'aide à l'analyse sémantique. Au-delà de l'opposition *entre* / *parmi* que nous venons d'évoquer, nous avons ainsi eu l'occasion de l'appliquer à plusieurs types de phénomènes :

- La pluralité
    - avec les déterminants *quelques* et *plusieurs* (Gréa 2008)
- La relation d'inclusion
    - avec les expressions *être une partie de* et *faire partie de* (Gréa 2012)

- avec les prépositions *au milieu de, au centre de, au sein de, au cœur de* et *parmi* (Gréa 2017)

L'opposition action / objet dans le domaine nominal :

- avec les constructions <un mode de N> et <une manière de N> (Gréa & Moline 2013)

- les constructions <mode de N> et <type de N> (Gréa & Haas 2015)

- où l'opposition entre action (*travail*) et événement (*affrontement*) (Haas & Gréa 2015)

3.4 Sens et collocation

Il existe beaucoup d'autres méthodes permettant de quantifier le degré d'association (ou de répulsion) entre différents éléments linguistiques (construction grammaticale, mots, etc.). On trouvera dans Wiechmann (2008) une liste de plus d'une quarantaine de méthodes ainsi que leur évaluation. Au-delà de son utilité pratique, en tant qu'elle donne accès à une synthèse de la tendance distributionnelle d'une unité, se pose toutefois la question de la justification de cette méthode par rapport à l'analyse sémantique. En quoi l'étude des tendances distributionnelles d'une unité a-t-il un lien avec le sens de cette unité ?

L'idée qui motive ce genre d'étude est assez ancienne et se trouve résumée par un slogan célèbre qu'on trouve sous la plume de Firth (1957) : « You shall know a word by the company it keeps ! ». Le meilleur moyen d'accéder au sens d'une unité lexicale consiste à s'intéresser à son entourage préférentiel, c'est-à-dire ce qu'on appelle, dans la tradition contextualiste anglaise, ses collocations ou ses colligations (Legallois 2006). Firth donne ainsi l'exemple des phrases suivantes : *Don't be such an ass !, You silly ass !, What an ass he is !* où *ass* « is in familiar and habitual company, commonly collocated with *you silly-*, *he is a silly-* , *don't be such a-*. » et prend un sens bien particulier (qu'on traduirait par « âne » en français, plutôt que par « cul »). D'autres exemples, toujours de Firth, illustrent bien la démarche contextualiste de l'auteur :

> It can safely be stated that part of the ' meaning ' of *cows* can be indicated by such collocations as *They are milking the cows, Cows give milk.* The words *tigresses* or *lionesses* are not so collocated and are already clearly separated in meaning at the *collocational level*.

> Situations of calendrical reference in which, for example, the names of the days of the week and of the month are a feature would attest the systematic use of the series of seven and twelve. But that is not by any means the complete cultural picture. In English, for instance, typical collocations for the words Sunday, Monday, Friday, and Saturday, furnish interesting material and would certainly separate them from the corresponding words in Chinese, Hebrew, Arabic or Hindi. The English words for the months are characteristically collocated : March hare, August Bank Holiday, May week, May Day, April showers, April fool, etc. (Firth 1957)

Ce point de vue, qu'on retrouve dans les travaux de Harris (1968: 12), et sous la plume duquel il a pris l'étiquette de *distributional hypothesis*, constitue le sol théorique sur lequel reposent les travaux évoqués dans les Sections 3.1 et 3.2, comme en attestent les citations suivantes :

> Our approach has much in common with a position that was popular in the 1950s. It was common practice to classify words not only on the basis of their meanings but also on the basis of their co-occurrence with other words. Running through the whole Firthian tradition, for example, is the theme that "You shall know a word by the company it keeps" (Firth 1957). Harris's "distributional hypothesis" dates from about the same period. He hypothesized that ''the meaning of entities, and the meaning of grammatical relations among them, is related to the restriction of combinations of these entities relative to other entities'' (Harris 1968: 12). (Church, Hanks, Hindle, & Gale 1991)

> As mentioned above, CA [*Collostructional Analysis*] is basically little more than the extension of the quantitative study of collocation (co-occurrences of words) with association measures (AMs) in corpus linguistics to the study of colligation (co-occurrences of words and grammatical patterns or constructions, hence *collostruction*) in Construction Grammar. Why would one want to study such co-occurrence phenomena? Because of the so-called *distributional hypothesis*, the assumption and finding that the similarity linguistic elements exhibit in terms of functional characteristics (semantic, pragmatic,…) will be reflected in their distributional patterning in language (corpora) (cf. Firth [1957]; Harris [1970: 785-786]), i.e., the frequencies with which linguistic elements of interest co-occur with other linguistic/contextual elements.

Au-delà de cette tradition bien installée, la légitimité de cette hypothèse nous paraît toutefois susceptible d'être encore discutée. Pour justifier son point de vue, Firth fait par exemple appel à un philosophe célèbre : « As Wittgenstein says, ' the meaning of words lies in their use.' » Mais quitte à se prévaloir d'une philosophie, on pourrait tout aussi bien faire appel à la *material relevancy* de Gurwitsch (1957), ou encore s'inspirer de la psychologie en transposant dans ce contexte la notion de frange introduite par James (1890).

L'idée sous-jacente à ces deux derniers points de vue, et que nous allons discuter succinctement dans ce qui suit, est la notion d'affinité. Bien qu'intuitivement intéressante,

cette notion, au même titre que les atomes crochus de Démocrite, est en l'état trop imprécise pour acquérir le statut de concept opératoire. La notion d'affinité, en effet, peut correspondre à des cas de figure et des types de relation très différents. Une version rigide de l'affinité pourrait ainsi correspondre à la sorte de relation qui existe entre deux pièces d'un puzzle. Si ces dernières en viennent à s'emboiter parfaitement, c'est en raison de leurs formes complémentaires qui s'ajustent les unes aux autres. C'est ce genre d'affinité qui est mise en avant dans le cadre de la Grammaire Cognitive de Langacker (ou en tout cas, de notre utilisation de la Grammaire Cognitive). Ainsi, la relation d'affinité qui existe entre la préposition *entre* et le N_{pl} *parties* s'explique par le dispositif schématique porté par *entre* auquel se conforme parfaitement le schéma associé au pluriel de *partie* (Gréa 2015). À l'inverse, la relation de répulsion qui existe entre *parmi* et *parties* s'explique par le fait que le schéma porté par *parties* ne se conforme pas à celui de *parmi*, ou alors exige un reprofilage pour pouvoir s'y conformer, et par conséquent, donne lieu à un emploi marqué (et donc moins fréquent). Par exemple, il s'avère que la majorité des 13 occurrences de *parmi les parties* correspond en réalité à des *parties civiles*, c'est-à-dire à des éléments d'un ensemble (et non des parties d'un tout).[23]

Cette possibilité de reprofiler un mot nous amène alors à une version un peu moins rigide de la relation d'affinité, en tant qu'elle intègre une certaine élasticité propre aux éléments combinés. L'opposition entre ces deux approches est commentée par Cohen (1986), dans le cadre de la philosophie analytique, lorsqu'il oppose l'approche interactionniste à l'approche isolationniste :

> According to the insulationist account the meaning of any one word that occurs in a particular sentence is insulated against interference from the meaning of any other word in the same sentence. On this view the composition of a sentence resembles the construction of a wall from bricks of different shapes. The result depends on the properties of the parts and the pattern of their combination. But just as each brick has exactly the same shape in every wall or part of a wall to which it is moved, so too each standard sense of a word or phrase is exactly the same in every sentence or part of a sentence in which it occurs…
> Interactionism makes the contradictory assertion: in some sentences in some languages the meaning of a word in a sentence may be determined in part by the word's verbal context in that sentence… On this view the composition of a sentence is more like the construction of a wall from sand-bags of different kinds. Though the size, structure, texture and contents of a sand-bag restrict the range of shapes it can take on, the actual shape it adopts in a particular situation depends to

[23] Si notre corpus était uniquement construit dans le domaine juridique, un tel emploi deviendrait probablement très fréquent et n'apparaîtrait plus comme marqué. Mais cela ne remettrait pas en question notre analyse : en deçà de la question du genre textuel, il y a de toute façon une relation d'affinité « formelle » entre *parties civiles* et *parmi*.

> a greater or lesser extent on the shapes adopted by other sand-bags in the wall, and the same sandbag might take on a somewhat different shape in another all or in a different position in the same wall. (Cohen [1986 : 223], cité par Recanati [2009])

Dans cette seconde version, donc, les pièces d'un puzzle (ou d'un mur, selon Cohen) sont simplement susceptibles de se déformer pour s'ajuster les unes aux autres, tout en gardant un noyau invariant (qui correspondrait par exemple à la base conceptuelle dans la Grammaire Cognitive).[24] A noter que l'absence totale d'affinité serait assez bien rendu par la métaphore d'un mur de briques identiques : chaque brique n'entretiendrait pas plus (ou pas moins) d'affinité avec la brique qui lui est immédiatement contigüe qu'avec n'importe quelle autre brique du mur.

Une version encore plus souple de la notion d'affinité correspond à la sorte de relation qui s'instaure dans un système électrostatique où l'on rapproche deux corps conducteurs chargés, de sorte que leurs charges respectives se déplacent dans un processus d'ajustement mutuel. C'est le genre d'affinité qui se trouve mis en avant dans le cadre de la Sémantique Interprétative ou dans la notion de compositionalité gestaltiste de Victorri & Fuchs (1996) et Col, Aptekman, Girault, & Victorri (2010).

Nous ne pensons pas que ces trois exemples épuisent tous les cas de figure possibles, et il existe sans doute d'autres configurations qui pourraient relever de ce qu'on entend intuitivement par « affinité ». L'utilisation de ce terme permet toutefois de poser le problème sous un angle intéressant, et qui a l'avantage de se placer en léger décalage par rapport à la linguistique de corpus contemporaine, pour laquelle, en somme, l'affinité est plutôt conçue comme le résultat d'une sorte de règle de Hebb.

[24] Cette idée d'un noyau qui resterait invariant sous de simples variations de surface est très critiquée par Cadiot & Visetti (2001).

Bibliographie

ANDERSON, J.R., 1990, *The Adaptive Character of Thought*, Hillsdale, Lawrence Erlbaum Associates.

ANDERSON, R.C., & ORTONY, A., 1975, « On Putting Apples into Bottles - A Problem of Polysemy », *Cognitive psychology*, 7, p. 167-180.

ASHINO, F., 2007, « A propos de la préposition parmi : étude contrastive avec entre », *Cahiers d'Études Françaises*, 12, p. 1-16.

ASLIN, R.N., SAFFRAN, J.R., & NEWPORT, E.L., 1998, « Computation of conditional probability statistics by 8-month-old infants », *Psychological Science*, *9*, p. 321–324.

BARCLAY, J.R., BRANSFORD, J.D., FRANKS, J.J., MCCARRELL, N.S., & NITSCH, K., 1974, « Comprehension and Semantic Flexibility », *Journal of Verbal Learning and Verbal Behavior*, 13, p. 471-481.

BARSALOU, L.W., 1982, « Context-independent and context-dependent information in concepts », *Memory & Cognition*, *10*, 1, p. 82-93.

BARSALOU, L.W., 1993, « Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols », *Theories of memory*, *1*.

BARSALOU, L.W., & BILLMAN, D., 1989, « Systematicity and semantic ambiguity », in D. S. GORFEIN, *Resolving semantic ambiguity*, p. 146-203. New York, Springer.

CADIOT, P., & VISETTI, Y.-M., 2001, *Pour une théorie des formes sémantiques : motifs, profils, thèmes*. Paris, PUF.

CHOMSKY, N., 1969, *Structures syntaxiques [trad. Braudeau, M.]*, Éditions du Seuil.

CHURCH, K.W., & HANKS, P., 1990, « Word association norms, mutual information, and lexicography », *Computational linguistics*, *16*, 1, p. 22-29.

CHURCH, K.W., HANKS, P., HINDLE, D., & GALE, W., 1991, « Using Statistics in Lexical Analysis », in ZERNIK, *Lexical Acquisition: Using On-line Resources to Build a Lexicon*, Lawrence Erlbaum, p. 115-164.

COHEN, L.J., 1986, « How is Conceptual Innovation Possible ? », *Erkenntnis*, 25, p. 221-238.

COL, G., APTEKMAN, J., GIRAULT, S., & B. VICTORRI, 2010, « Compositionnalité gestaltiste et construction du sens par instructions dynamiques », *CogniTextes*, [En ligne].

CROFT, W., & CRUSE, D.-A., 2004, *Cognitive Linguistics*, Cambridge, Cambridge University Press.

CRUSE, D.-A., 1986, *Lexical Semantics*. Cambridge, Cambridge University Press.

DESAGULIER, G., 2015, « Le statut de la fréquence dans les grammaires de constructions: simple comme bonjour? », *Langages*, *197*, p. 99-128.

DUDA, R.O., & HART, P.E., 1973, *Pattern classification and scene analysis*. New York, Wiley.

ÉGRÉ, P., & BARBEROUSSE, A., 2014, « Borel on the heap », *Erkenntnis*, 79:5, p. 1043-1079.

ELLIS, N., 2006, « Language acquisition as rational contingency learning », *Applied Linguistics*, *27*, 1, p. 1-24.

FIRTH, J., 1957, « A Synopsis of Linguistic Theory 1930-1955 », in *Studies in Linguistic Analysis, Philological Society. Reprinted in Palmer, F. (ed. 1968) Selected Papers of J.R. Firth, Longman, Harlow.*, Oxford, Basil Blackwell, p. 1-32.

FLAUX, N., 2011, « A propos du verbe traduire et du nom traduction », *Studii de Lingvistică*, 1:1, p. 85-104.

FODOR, J.A., & PYLYSHYN, Z.W., 1988, « Connectionism and cognitive architecture: A critical analysis », *Cognition, 28*, 1, p. 3-71.

FRANCKEL, J.-J., & PAILLARD, D., 1998, « Aspects de la théorie d'Antoine Culioli », *Langages, 129*, p. 52-63.

FRANCKEL, J.J., & LEBAUD, D., 1992, « Lexique et opération. Le lit de l'arbitraire », in *La théorie d'Antoine Culioli. Ouvertures et incidences*, Ophrys, Paris, p. 89-105.

FRANCKEL, J.J., & PAILLARD, D., 2007, *Grammaire des prépositions, tome 1*, Paris, Ophrys.

GARDIES, J.L., 1975, *Esquisse d'une grammaire pure*. Paris, J. Vrin.

GOLDBERG, A.E., 2006, *Constructions at Work: The Nature of Generalization in Language*, New York, Oxford University Press.

GOODMAN, N.D., & LASSITER, D., 2015, « Probabilistic Semantics and Pragmatics: Uncertainty in Language and Thought », in S. LAPPIN & C. FOX, *The Handbook of Contemporary Semantic Theory*, Oxford, Wiley Blackwell, p. 655-686.

GREA, P., 2006, « La notion d'*entrenchment* dans le cadre des grammaires cognitives », in D. LEGALLOIS & J. FRANÇOIS, *Autour des grammaires de constructions et de patterns*, Cahier du Crisco, 21, p. 18-26.

GREA, P., 2008, « Quelques et plusieurs », in J. DURAND, B. HABERT & B. LAKS, *Congrès Mondial de Linguistique Française - CMLF'08*, Paris, p. 2031-2050.

GRÉA, P., 2012, « "Faire partie de" : not a piece of cake », in M. BOUVERET & D. LEGALLOIS, *Constructions in French*, Amsterdam, John Benjamins, p. 73-97.

GREA, P., 2015, « Entre et parmi : deux perspectives sur la pluralité », *Travaux de linguistique*, 70, p. 7-38.

GREA, P., 2017, « Inside in French », *Cognitive Linguistics, 28*, 1, p. 77-130.

GREA, P., & HAAS, P., 2015, « Mode de N et type de N, de la synonymie à la polysémie », *Langages*, 197, p. 69-98.

GREA, P., & MOLINE, E., 2013, « "Une manière de construction / un mode de construction" Classification floue et classification hyperonymique », *Le Français moderne*, 2, p. 215-229.

GREENSPAN, S.L., 1986, « Semantic flexibility and referential specificity of concrete nouns », *Journal of Memory and Language, 25*, 5, p. 539-557.

GRIES, S.T., 2012, « Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics: Some necessary clarifications », *Studies in Language, 36*, 3, p. 477-510.

GRIES, S.T., 2015, « More (old and new) misunderstandings of collostructional analysis: On Schmid and Küchenhoff (2013) », *Cognitive Linguistics, 26*, 3, p. 505–536.

GRIES, S.T., & STEFANOWITSCH, A., 2004a, « Co-varying collexemes in the into-causative », in M. A. S. KEMMER, *Language, Culture, and Mind*, Stanford (CA), CSLI, p. 225-236.

GRIES, S.T., & STEFANOWITSCH, A., 2004b, « Extending collostructional analysis: A corpus-based perspective on <alternations> », *International Journal of Corpus Linguistics, 9*, 1, p. 97-129.

GUENTCHEVA, Z., 2003, « Entre : préposition et préfixe », in P. BLUMENTHAL & J.-E. TYVAERT, *La cognition dans le temps* Tübingen, Max Niemeyer Verlag, p. 59-74.

GURWITSCH, A., 1957, *Théorie du champ de la conscience*. Bruges, Desclée de Brouwer.

HAAS, P., & GREA, P., 2015, « Action et événement, deux types nominaux distincts ? », *Langue française*, 185, p. 85-98.

HABERT, B., 1985, « L'analyse des formes « spécifiques » [bilan critique et propositions d'utilisation] », *Mots*, 11, p. 127-154.

HARRIS, Z.S., 1968, *Mathematical Structures of Language*, New York, Wiley.

HILGERT, E., 2007, « Étude de parmi. Le cas des tours partitifs », *Scolia* 21, p. 37-66.

Bibliographie

HILGERT, E., 2009, « Retour sur les prépositions ensemblistes à interprétation spatiale », *SCOLIA*, 24, p. 53-69.

HILGERT, E., 2010, *La partition et ses constructions en français*, Genève Librairie Droz.

HILGERT, E., 2013, « Les prépositions ensemblistes et la question de leur emploi spatial », *CORELA - Numéros thématiques | Langue, espace, cognition.*

JAMES, W., 1890, *The principles of Psychology*, Vol. I. New York, Henry Holt and Company.

JAYNES, E.T., 2003, *Probability Theory: The Logic of Science*, Cambridge, Cambridge University Press.

KLEIBER, G., 1990, *La sémantique du prototype : catégories et sens lexical*, PUF ed., Paris, PUF.

KLEIBER, G., 1995, « Polysémie, transferts de sens et métonymie intégrée », *Folia linguistica, 29*, 1-2, p. 105-132.

KLEIBER, G., 2008, « Petit essai pour montrer que la polysémie n'est pas un sens interdit », in J. DURAND, B. HABERT & B. LAKS, *Congrès Mondial de Linguistique Française - CMLF'08*, Paris, p. 87-101.

KWON-PAK, S.-N., 2006, « Entre vs. parmi : deux prépositions au centre de la partition », in G. KLEIBER, C. SCHNEDECKER & A. THEISSEN, *La Relation partie-tout*, Leuven, Peeters, p. 651-668.

LABBE, C., & LABBE, D., 2001, « Que mesure la spécificité du vocabulaire ? », *Lexicometrica, 3*, p. [on line].

LAFON, P., 1980, « Sur la variabilité de la fréquence des formes dans un corpus », *Mots*, 1, p. 127-165.

LAFON, P., 1984, *Dépouillements et statistiques en lexicométrie*, Genève-Paris, Slatkine-Champion.

LANGACKER, R.W., 2009, « Metonymic grammar », in K.-U. PANTHER, L. L. THORNBURG & A. BARCELONA, *Metonymy and Metaphor in Grammar*, Amsterdam, John Benjamins, p. 45–71.

LASSITER, D., & GOODMAN, N.D., 2015, « Adjectival vagueness in a Bayesian model of interpretation », *Synthese*, p. 1-36.

LEBART, L., & SALEM, A., 1994, *Statistique textuelle*, Paris, Dunod.

LEGALLOIS, D., 2006, « La Grammaire de Construction », in D. LEGALLOIS & J. FRANÇOIS, *Autour des grammaires de constructions et de patterns*, Cahier du Crisco, 21, p. 5-27.

LOISEAU, S., GREA, P., & MAGUE, J.-P., 2010, « Dictionnaires, théorie des graphes et structures lexicales », *Revue de Sémantique et de Pragmatique*, 27, p. 51-78.

MCCLELLAND, J.L., & RUMELHART, D.E., 1989, *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises*, Cambridge, MIT Press.

MCCULLOCH, W.S., & PITTS, W., 1943, « A logical calculus of the ideas immanent in nervous activity », *The bulletin of mathematical biophysics*, 5:4, p. 115-133.

MINSKY, M.L., & PAPERT, S., 1988, *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MIT Press.

MONTAVON, G., ORR, G., & MÜLLER, K.R., 2012, *Neural Networks: Tricks of the Trade*. Berlin, Heidelber, Springer.

MURPHY, G.L., 1982, « Cue validity and levels of categorization », *Psychological Bulletin*, 91:1, p. 174-177.

MURPHY, G.L., & MEDIN, D.L., 1985, « The role of theories in conceptual coherence », *Psychological review*, 92:3, p. 289-316.

NUNBERG, G., 1979, « The non-uniqueness of semantic solutions: Polysemy », *Linguistics and philosophy, 3*, 2, p. 143-184.

NUNBERG, G., 1995, « Transfers of meaning », *Journal of semantics*, *12*, 2, p. 109-132.

PUSTEJOVSKY, J., 1993, « Type Coercion and Lexical Selection », in J. PUSTEJOVSKY, *Semantics and the Lexicon*, Dordrecht, Kluwer, p. 73-94.

PUSTEJOVSKY, J., 1995, *The Generative Lexicon*, MIT Press ed.

QUINE, W.V.O., 1960, *Word and Object*. Cambridge, MIT Press.

RAO, R.P.N., OLSHAUSEN, B.A., & LEWICKI, M.S., 2002, *Probabilistic Models of the Brain: Perception and Neural Function*, MIT Press.

RASTIER, F., 1987, *Sémantique interprétative*, Paris, Presses Universitaires de France.

RASTIER, F., 1991, *Sémantiques et recherches cognitives*, PUF ed., Paris, Presses Universitaires de France.

RECANATI, F., 2004, *Literal Meaning*. Cambridge, Cambridge University Press.

RECANATI, F., 2009, « Compositionality, Semantic Flexibility, and Context-Dependence », in W. HINZEN, E. MACHERY & M. WERNING, *Oxford Handbook of Compositionality* [en ligne]. Oxford, Oxford University Press.

ROSCH, E., 1973, « Natural categories », *Cognitive psychology*, *4*, p. 328-350.

ROSCH, E., 1978, « Principles of categorization », in E. ROSCH & B. B. LLOYD, *Cognition and categorization*, Hillsdale, NJ: Erlbaum,

ROSCH, E., & MERVIS, C., 1975, « Family resemblances : Studies in the Internal Structure of Categories », *Cognitive Psychology*, *7*, p. 573-605.

ROTH, E.M., & SHOBEN, E.J., 1983, « The effect of context on the structure of categories », *Cognitive psychology*, *15*, 3, p. 346-378.

RUMELHART, D.E., & MCCLELLAND, J.L., 1986, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Cambridge, MIT Press.

SAFFRAN, J.R., 2001, « Words in a sea of sounds: The output of infant statistical learning », *Cognition*, *81*, 2, p. 149-169.

SAFFRAN, J.R., ASLIN, R.N., & NEWPORT, E.L., 1996, « Statistical learning by 8-month-old infants », *Science*, *274*, p. 1926-1928.

SAFFRAN, J.R., NEWPORT, E.L., & ASLIN, R.N., 1996, « Word segmentation: the role of distributional cues », *Journal of Memory and Language*, *35*, p. 606–621.

SALEM, A., 1987, *Pratique des segments répétés: essai de statistique textuelle*, Paris, Klincksieck.

SCHOEN, L.M., 1988, « Semantic Flexibility and Core Meaning », *Journal of Psycholinguistic Research*, *17*, 2, p. 113-123.

SEARLE, J., 1980, « The Background of Meaning », in J. SEARLE, F. KIEFER & BIERWISCH, *Speech Act Theory and Pragmatics*, Dordrecht, Reidel, p. 221–232.

SILBERZTEIN, M., 2003. *NooJ manual*. http://www.nooj4nlp.net.

SILBERZTEIN, M., 2004, « NooJ : an Object-Oriented Approach », in C. MULLER, J. ROYAUTE & M. SILBERZTEIN, *INTEX pour la Linguistique et le Traitement Automatique des Langues*, Cahiers de la MSH Ledoux. Presses Universitaires de Franche-Comté, p. 359-369.

SMITH, E.E., & MEDIN, D.L., 1981, *Categories and concepts*, Harvard University Press.

STEFANOWITSCH, A., 2006, « Distinctive collexeme analysis and diachrony: A comment », *Corpus linguistics and linguistic theory*, *2*, 2, p. 257-262.

STEFANOWITSCH, A., 2013, « Collostructional Analysis », in T. HOFFMANN & G. TROUSDALE, *The Oxford Handbook of Construction Grammar*, New York, Oxford University Press.

STEFANOWITSCH, A., & GRIES, S.T., 2003, « Collostructions: On the interaction between verbs and constructions », *International Journal of Corpus Linguistics*, *8*, 2, p. 209–243.

STEFANOWITSCH, A., & GRIES, S.T., 2005, « Covarying collexemes », *Corpus Linguistics and Linguistic Theory*, *1*, 1, p. 1-43.

Bibliographie

STOSIC, D., & FLAUX, N., 2012, « Les noms d'idéalités sont-ils polysémiques? », in L. SAUSSURE & A. RIHS, *Etudes de sémantique et pragmatique françaises*, p. 167-190. Bern, Peter Lang.

TABOSSI, P., & JOHNSON-LAIRD, P., 1980, « Linguistic context and the priming of semantic information », *The Quarterly Journal of Experimental Psychology, 32*, 4, p. 595-603.

TENENBAUM, J.B., & GRIFFITHS, T.L., 2001, « The rational basis of representativeness », *Proceedings of the 23rd annual conference of the Cognitive Science Society*, p. 1036-1041.

TENENBAUM, J.B., KEMP, C., GRIFFITHS, T.L., & GOODMAN, N.D., 2011, « How to Grow a Mind: Statistics, Structure, and Abstraction », *Science*, 331:6022, p. 1279-1285.

VAN GOETHEM, K., 2009, *L'emploi préverbal des prépositions en français: Typologie et grammaticalisation*, Bruxelles, De Boeck Duculot.

VICTORRI, B., 1997, « La polysémie : un artefact de la linguistique ? », *Revue de sémantique et pragmatique, 2*, p. 41-62.

VICTORRI, B., & FUCHS, C., 1996, *La polysémie : construction dynamique du sens*. Paris, Hermès.

WEINREICH, U., 1972, *Explorations in Semantic Theory*, La Hague, De Gruyter.

WIECHMANN, D., 2008, « On the computation of collostruction strength: Testing measures of association as expressions of lexical bias », *Corpus Linguistics and LinguisticTheory*, 4, 2, p. 253-290.

XU, F., & TENENBAUM, J.B., 2007, « Word Learning as Bayesian Inference », *Psychological Review, 114*, 2, p. 245-272.

Index des noms

Abeillé 29
Anderson 16, 18, 23
Barclay 21, 22
Barsalou 8
Cadiot 51
Chomsky 25, 26, 29
Church 34, 35, 36, 37, 38, 42, 43
Cohen 50, 51
Col 51
Cruse 22
Dehaene 17
Démocrite 49
Desagulier 37
Desgraupes 41
Duda 17
Égré 32
Firth 48, 49
Flaux 30
Fodor 18, 25, 38
Gardies 26
Goldberg 37
Goodman 10, 17, 32
Gries 36, 37, 38
Gurwitsch 49
Habert 39
Harris 49
Husserl 26, 30
James 49
Jaynes 11, 12
Kleiber 23, 24, 25
Labbé 45
Lafon 38, 39, 41
Langacker 22, 23, 36, 50
Lassiter 32
Loiseau 32, 41
McClelland 5
McCulloch 18
Medin 8
Minsky 18
Montavon 19
Murphy 8
Nunberg 30, 31
Pylyshyn 18, 25, 38
Quine 16
Rastier 24, 25, 32, 33
Recanati 22, 51
Robespierre 38, 39
Rosch 8, 24
Rumelhart 5
Salem 45
Smith 8
Stefanowitsch 36, 37, 38
Stosic 30
Tenenbaum 10, 16, 17, 20
Victorri 51
Weinreich 22
Wiechmann 48
Wittgenstein 49
Xu 17, 20

Index thématique

Catégorisation.....5, 8, 12, 16, 19, 20
Chaînes de markov.....28, 29
Construction (grammaire de) 29, 36
  dative.....37
  ditransitive.....37
Cue-validity.....6, 7, 8, 20, 30, 31, 37
Deep learning 5, 17, 18, 19, 20
Entrenchment 27, 37
Flexibilité sémantique 21, 22, 23, 24, 30, 32
Grammaire cognitive.....22, 36, 50, 51
Loi hypergéométrique 40, 41
Niveau de base 7, 8, 24
Paradoxe sorite 32
Plausibilité9, 10, 11, 12, 13, 14, 15, 17, 20, 27, 30, 32, 33
Principe d'intégration méronomique.....24
Principe de métonymie intégrée 24
Probabilité conditionnelle... 6, 9, 16, 27, 28
Prototype..... 6, 7, 8, 15, 16, 20, 37
Règle de Bayes 9, 10, 11, 13, 15, 16, 17, 20, 26, 27, 32, 33
  A posteriori 11, 14, 33
  A priori 11, 13, 14, 15, 17, 20, 26, 27, 33
  Evidence 10, 13, 33
  Vraisemblance 10, 11, 13, 14, 17, 33
Règle de Hebb..... 51
Relation partie / tout 24
Schmolblock 27, 28, 29
Sémantique Interprétative..... 24, 32, 51
Table de spécificités 46, 47
Valeur modale..... 41, 43, 44
Zone active..... 23, 24