

BAUDOT, Jean (1992) : *Fréquences d'utilisation des mots en français écrit contemporain*, Montréal. Presses de l'Université de Montréal, 432 p.

Alain Content

Volume 39, Number 3, septembre 1994

URI: <https://id.erudit.org/iderudit/002362ar>

DOI: <https://doi.org/10.7202/002362ar>

[See table of contents](#)

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)

1492-1421 (digital)

[Explore this journal](#)

Cite this review

Content, A. (1994). Review of [BAUDOT, Jean (1992) : *Fréquences d'utilisation des mots en français écrit contemporain*, Montréal. Presses de l'Université de Montréal, 432 p.] *Meta*, 39(3), 482–483. <https://doi.org/10.7202/002362ar>

■ BAUDOT, Jean (1992): *Fréquences d'utilisation des mots en français écrit contemporain*, Montréal, Presses de l'Université de Montréal, 432 p.

La psycholinguistique moderne se donne pour objet de développer des modèles mécanistes des processus mentaux mis en œuvre dans le traitement du langage naturel. Son champ d'investigation inclut aussi bien les processus de compréhension du langage que les aspects productifs et la problématique de l'acquisition. Les travaux concernent tant les langues parlées que les langues écrites, ainsi que d'autres systèmes de communication linguistique comme les langues signées et autres systèmes de communication. Le lexique mental — son organisation, sa constitution, les mécanismes d'accès et de restitution — constitue une composante extrêmement importante du système de traitement langagier, et une très large part des travaux récents y sont consacrés. C'est en effet à partir de 1970 que le concept de lexique mental fait son apparition dans les travaux de psychologie cognitive, initialement à l'occasion de travaux qui ne concernaient que très marginalement le traitement du langage naturel.

Ces recherches ont rapidement fait apparaître la nécessité de disposer d'une connaissance des caractéristiques statistiques de la langue. Ainsi, par exemple, l'information sur la fréquence des mots constitue une variable d'importance cruciale pour de nombreux champs d'étude de la psycholinguistique contemporaine. D'une part, le fait que la performance varie selon la fréquence d'usage des mots dans la langue constitue un indice important de la nature des procédures de traitement utilisées par le système cognitif humain, et cette observation est à l'origine de controverses intenses concernant la nature des mécanismes d'identification des mots parlés et écrits. D'autre part, l'importance et l'ubiquité de l'effet de la fréquence d'usage exige que cette variable soit prise en considération dans la très large majorité des investigations expérimentales, ne serait-ce que pour s'assurer que ce facteur n'interfère pas de manière artefactuelle avec les autres variables étudiées.

L'information sur la fréquence d'usage des mots n'est évidemment pas la seule caractéristique importante qui doit être prise en considération. Mais elle joue un rôle crucial pour de nombreuses raisons. On sait depuis longtemps déjà que la fréquence d'usage des mots est associée à d'autres caractéristiques : les mots plus fréquents ont tendance à être composés de phonèmes plus fréquents et de structure phonologique plus simple. Par exemple, Landauer et Streeter avaient observé, dès 1973, que les mots d'usage fréquent tendaient souvent à se trouver dans des régions plus peuplées de l'espace lexical, si l'on conçoit que tout mot de la langue a des «voisins» du point de vue formel (les mots qui partagent toutes les lettres à une près (cage, nage, rage, pape, pale, pige, etc.) sont des voisins de page) et que le lexique comprend donc des zones plus peuplées et d'autres moins occupées. Cette notion de voisinage a progressivement pris une importance considérable non seulement dans les travaux empiriques, mais également dans les débats théoriques concernant l'organisation des connaissances lexicales et les procédures d'accès, et l'on s'accorde actuellement à admettre que la facilité ou la difficulté à reconnaître un mot familier — et donc à accéder à sa signification — dépend non seulement de ses propres caractéristiques, mais également de celles de son «environnement lexical», et il semble que les deux aspects soient liés de manière complexe à la fréquence d'usage.

Un recensement de la fréquence des mots de l'anglais contemporain, qui a constitué un des outils essentiels de la psycholinguistique en langue anglaise jusqu'à ces dernières années, avait été publié en 1967 par Kucera et Francis. Aucune source comparable n'existait pour le français, et c'est cette lacune que vient combler l'ouvrage de Jean Baudot. D'autres listes de fréquence existent. La plus connue est probablement celle publiée en 1960 ? par le CNRS français (Groupe du Trésor de la langue française à Nancy).

Malheureusement pour les psycholinguistes qui s'intéressent plus à la langue contemporaine qu'à son histoire et son usage littéraire, ces tables ont été réalisées dans une perspective très différente. Elles sont basées sur un corpus important de textes, mais dans lequel seuls des textes littéraires (romans, pièces, essais) datant d'avant 1960 ont été pris en considération. Dans la mesure où elle est basée sur un échantillonnage bien équilibré de documents écrits récents, l'étude réalisée par M. Baudot et ses assistants fournit donc une information complémentaire extrêmement utile.

L'équipe du professeur Baudot ne s'est pas contentée d'un relevé de la fréquence des formes lexicales. Le travail d'analyse extrêmement méticuleux réalisé sur la décomposition des homonymes fournit une information particulièrement riche dont il n'existe à ma connaissance aucun équivalent à ce jour pour la langue française. Comme le souligne M. Gross dans la préface, cette analyse ne peut se faire de manière entièrement automatique, et on comprend donc la limitation à un corpus d'environ 1 000 000 de mots, qui, en regard des bases de données utilisées pour des traitements automatisés, peut apparaître relativement restreint mais constitue cependant un échantillon suffisamment grand pour donner des informations valides, tout au moins pour les formes de fréquence moyenne et élevée.

Il y a donc toutes les raisons de penser que l'ouvrage réalisé par Jean Baudot et son équipe est susceptible de constituer un outil de référence extrêmement précieux pour les recherches en psycholinguistique, ainsi que dans divers autres domaines de l'étude des langues et de leur usage.

ALAIN CONTENT
Université libre de Bruxelles, Bruxelles, Belgique