
actes n° 2022 | 2023

LE DISCOURS HORS-NORME(S) OU L'ATYPIE DU DISCOURS

La norme et la variation dans le cadre du Traitement Automatique du Langage

Représenter la diversité des parcours d'acquisition du langage à l'aide de
modèles informatiques

Andrea BRIGLIA *Chercheur PostDoc*

GIPSA

CRISSP

Grenoble Alpes University

Massimo MUCCIARDI

Pirrotta GIOVANNI

Édition électronique :

URL :

<https://cjc-praxiling.numerev.com/articles/actes-2022/2407-la-norme-et-la-variation-dans-le-cadre-du-traitement-automatique-du-langage>

DOI : numerev_1937

Date de publication : 16/10/2023

CertiScience® *Certifié évalué par les pairs*

Cette publication est sous licence **CC BY-NC-ND** (Attribution - No commercial - No derivatives).

Pour **citer cette publication** : BRIGLIA, A., MUCCIARDI, M., GIOVANNI, P. (2023) La norme et la variation dans le cadre du Traitement Automatique du Langage. *CJC-Praxiling*, (actes n°2022).

https://doi.org/10.34745/numerev_1937

Résumé : Cet article pose la problématique du statut de la norme et de la variation en TAL en proposant des exemples tirés des recherches précédentes concernant des modèles informatiques employés pour représenter l'acquisition de langue française. Deux cas d'étude exemplifient le choix autour de l'axe norme-variation : le calcul automatique d'une distribution de fréquence et la reconnaissance de motifs séquentiels. Que le niveau d'analyse soit le mot (premier exemple) ou le phonème (deuxième exemple), des obstacles et compromis reviennent d'une manière analogue. Le choix - souvent difficile et contraint - entre la précision de la description du langage et la nécessité d'avoir des données uniformes pour que la machine puisse les traiter aisément. Les biais évitables et inévitables, les précautions à prendre en amont, ainsi que les avantages et les inconvénients de ce type de modèles seront discutés. L'article se termine en dessinant les contours des futures complémentarités possibles entre méthodes qualitatives et quantitatives.

Abstract : This article deals with the problem of the status of norm and variation in NLP by proposing examples drawn from previous research concerning computer models used to represent French language acquisition. Two case studies illustrate the choice around the norm-variation axis: the automatic computation of a frequency distribution and the recognition of sequential patterns in words containing specific syllable sequences that are hard to learn due to their inner phonetic difficulty. Whether the level of analysis is the word (first example) or the phoneme (second example), obstacles and trade-offs come up in a similar way. The choice - often difficult and constrained - between the accuracy of the language description and the need to have uniform data for the machine to be easily handled. The avoidable and unavoidable biases, the precautions to be taken beforehand, as well as the advantages and disadvantages of these types of NLP models will be discussed. The article ends by outlining the possible future complementarities between qualitative and quantitative methods in current linguistics.

Keywords : first language acquisition; NLP, French; variation; norm