Un modèle de sélection de ressources multimodales pour l'apprentissage de la Langue française Parlée Complétée

Magali Norré Brigitte Bigi Núria Gala
Laboratoire Parole et Langage, Aix Marseille Univ, 5 Avenue Pasteur, 13100, France
magali.norre@univ-amu.fr, brigitte.bigi@cnrs.fr, nuria.gala@univ-amu.fr

Contexte – La Langue française Parlée Complétée (LfPC) est un système de communication multimodale qui combine la lecture labiale à des gestes pour désambiguïser les sons que l'on ne peut pas distinguer en se basant que sur la lecture des lèvres. La LfPC est l'adaptation française du *Cued Speech*, un code inventé aux États-Unis pour les personnes sourdes ou malentendantes (Cornett, 1967). Bien que son utilité soit reconnue pour améliorer la compréhension, il existe peu de recherches sur son apprentissage et peu de ressources disponibles (Gala et al., 2024).

Objectif – Cet article a pour but de proposer des variables pour un modèle de sélection de ressources multimodales enregistrées au format de capsules vidéo accessibles librement en ligne. Les textes proviennent du corpus Alector (Gala et al., 2020). Il s'agit de 200 textes littéraires ou scientifiques, qui s'adressent à des enfants entre 7 et 11 ans. La moitié sont des textes originaux ; l'autre moitié sont des versions simplifiées manuellement. L'extraction de leurs caractéristiques phonémiques et leur analyse permettront de proposer un premier modèle de la complexité d'apprentissage de la LfPC qui sera évalué par des personnes sourdes au sein du projet VizLector.

Méthodologie – Le logiciel d'annotation open source SPPAS (Bigi, 2015) a été utilisé pour extraire automatiquement les clés LfPC du corpus Alector. Chaque phonème français est codé avec une clé LfPC qui désigne l'une des 5 positions du visage pour représenter une voyelle et l'une des 8 formes de la main représentant une consonne (figure 1). Pour l'analyse du corpus, 10 variables phonémiques ont été extraites : le nombre de clés LfPC, le nombre de transitions entre des positions éloignées (côté ou pommette vs gorge et inversement), le nombre de chaque semi-voyelle ou semi-consonne (w, ɥ, j) et le nombre total, le nombre de liaisons obligatoires, le nombre de séquences CV (Consonne Voyelle) et C-C (Consonne-Consonne) dans un mot. Nous faisons ainsi l'hypothèse que ces phénomènes pourraient être plus complexes à (dé)coder en LfPC.

Résultats – Au total, 129 608 clés ont été extraites. Les 10 variables phonémiques étudiées sont peu corrélées entre elles. La position la plus fréquente est le côté du visage qui peut représenter le plus de voyelles et l'absence de voyelle (table 1). La forme de la main la plus fréquente est la forme 5, qui permet aussi de coder l'absence de consonne. La forme 7 est la moins fréquente car elle ne code que le phonème /g/. Ces premiers résultats sont proches de ceux de la littérature. Les mêmes ratios de séquences C, V et CV sont trouvés dans un corpus de 4 143 clés produits par des codeurs sur des textes lus (Bigi, 2023). Ces variables seront combinées à des variables de lisibilité dans un modèle non supervisé pour classer automatiquement les textes par niveau de complexité.

Remerciements – Ce travail a bénéficié d'une aide du gouvernement français au titre de France 2030, dans le cadre de l'Initiative d'Excellence d'Aix Marseille Université (AMIDEX - AMX-22-RE-AB-022), ainsi que de la FIRAH dans le cadre de son appel à projet 2022 (APa2022_022).

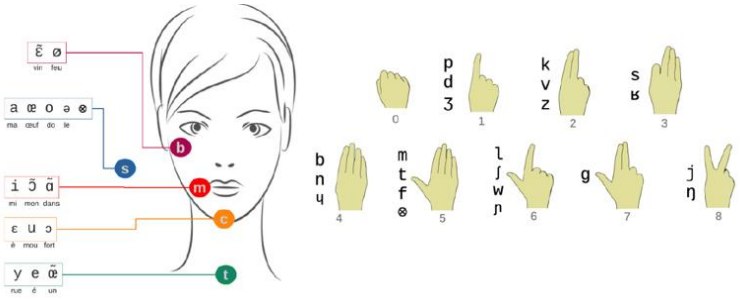


FIGURE 1 : Clés, positions (voyelles) et formes de la main (consonnes) en LfPC

Position	#	%	Voyelles	Forme de la main	#	%	Consonnes
Côté (s)	65 625	50,60	a œ o ə ⊕	5	30 596	23,61	m t f ⊕
Menton (c)	22 869	17,60	ɛ u ɔ	3	30 135	23,25	s ʁ
Lèvres (m)	22 354	17,20	i ɔ̃ ɑ̃	1	20 359	15,70	p d ʒ
Gorge (t)	15 523	12,00	y e œ̃	6	19 673	15,17	l ʃ w ɲ
Pommette (b)	3 237	2,50	ɛ̃ ø	2	14 289	11,02	k v z
				4	9 043	6,97	b n ɥ
				8	3 941	3,04	j ŋ
				7	1 572	1,21	g

TABLE 1 : Nombre total de clés et pourcentage par position et par forme de la main

Références

BIGI, B. (2015). SPPAS - Multi-Lingual Approaches to the Automatic Annotation of Speech. *The Phonetician*, 111-112, p. 54-69. HAL : hal-01417876.

BIGI, B. (2023). An analysis of produced versus predicted French Cued Speech keys. In *10th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics,* p. 24-28. HAL : hal-04081282.

CORNETT, R. O. (1967). Cued Speech. *American Annals of the Deaf*, 112, p. 3-13.

GALA, N., BIGI, B., & BAUER, M. (2024). Automatically Estimating Textual and Phonemic Complexity for Cued Speech: How to See the Sounds from French Texts. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, p. 1817-1824. HAL : hal-04580180.

GALA, N., TACK, A., JAVOUREY-DREVET, L., FRANÇOIS, T., & ZIEGLER, J. C. (2020). Alector: A parallel Corpus of Simplified French Texts with Alignments of Misreadings by Poor and Dyslexic Readers. In *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC 2020),* p. 1353-1361. HAL : hal-02503986.