

Modéliser l'acquisition du langage chez l'enfant

Contexte :

Comment avez-vous appris à parler ? Savez-vous par quelles étapes vous êtes passés avant d'être capable de construire une phrase complète ? Durant les premières années de sa vie, l'enfant apprend à parler et à former des phrases de plus en plus complexes (e.g. il va passer de « manger » à « je voudrais manger, s'il te plait »). Il peut apprendre n'importe quelle langue, sans effort et sans s'en rendre compte. Tout le monde parle sa langue maternelle, personne ne sait expliquer comment il a appris à le faire. Or, il existe une dynamique d'apprentissage très élaborée que nous allons chercher à modéliser dans ce stage.

Données :

Nous allons nous baser sur le corpus CoLaJE¹, comportant des suivis longitudinaux de productions spontanées de 7 enfants, de la naissance jusqu'à l'âge de 7 ans. Ces enfants ont été filmés une heure par mois. Les données ont été retranscrites dans un langage écrit, en utilisant une codification spécifique, dans trois formats différents : orthographe standard, « pho » ce que l'enfant dit et « mod » ce que l'enfant aurait dû prononcer selon la norme adulte.

Mission :

Nous allons rechercher s'il existe des régularités dans le développement typique du langage, notamment grammaticales, qui pourraient indiquer de possibles parcours préférentiels ou des conditions particulières qui favoriseraient ou non l'émergence de certaines variations observées dans un nombre représentatif de cas. En particulier, nous allons utiliser des méthodes de recherche de motifs, c'est-à-dire de récurrences intéressantes, utiles et parfois inattendues. En particulier, nous allons extraire des sous-séquences appelées motifs séquentiels et définies par Srikant et Agrawal (1995).

Tâches principales :

1. Pré-traitement des données : Les textes des retranscriptions vont être nettoyés, structurés puis étiquetés selon différentes formes grammaticales d'intérêt pour les chercheurs en science du langage.
2. Extraction de motifs séquentiels : Nous extrairons des sous séquences fréquentes caractérisant l'évolution de l'utilisation des formes grammaticales par les enfants.
3. Visualisations : Enfin, les connaissances extraites seront disponibles dans un outil d'analyse visuelle permettant aux chercheurs en science du langage de visualiser et d'explorer les données.

Candidatures :

Les candidatures à ce poste (CV, lettre de Motivation, bulletin de notes, références) seront reçues dans un seul document PDF accessible en téléchargement et envoyé par email à sandra.bringay@univ-montp3.fr.

Encadrement :

- UMR 5506 LIRMM : BRINGAY Sandra sandra.bringay@univ-montp3.fr
- Dipralang EA 739 : jeremi.sauvage@univ-montp3.fr

Localisation : LIRMM

Gratification : 6 mois

Références :

Agrawal, Rakesh and Ramakrishnan Srikant. "Mining sequential patterns." Proceedings of the Eleventh International Conference on Data Engineering (1995): 3-14.

¹ <http://colaje.scicog.fr/index.php/corpus>

Béchet, Nicolas & Cellier, Peggy & Charnois, Thierry & Crémilleux, Bruno. (2012). Discovering Linguistic Patterns Using Sequence Mining. 7181. 154-165. 10.1007/978-3-642-28604-9_13.

Colletta, J.M., and Guidetti, M., eds, (2012). *Gesture and Multimodal Development*. Amsterdam: John Benjamins.

Leroy-Collombel, M. & Morgenstern, A. (2011). Rising grammatical awareness in a French-speaking child from 18 months to 36 months: uses and misuses of possession markers. *French Language Studies* 22, 57-75.

Parisse, C., Maillart, C., & Tommerdahl, J. (2012). F-LARSP: A Computerized Tool for Measuring Morphosyntactic Abilities in French. In Martin J. Ball, David Crystal, Paul Fletcher (ed). *Assessing Grammar - The Languages of LARSP, Multilingual Matters*. (pp. 230-244). Bristol: UK.

Agrawal, Rakesh and Ramakrishnan Srikant. "Mining sequential patterns." *Proceedings of the Eleventh International Conference on Data Engineering* (1995): 3-14.