

BOOTPHON - A computational approach to early language bootstrapping

- **Responsable : Emmanuel Dupoux (LSCP)**
- **Participants : 2 doctorants, 3 post-docs, 1 ingénieur de recherche, 1 assistant de recherche**
- **Recherche financée par l'ERC, septembre 2012-aout 2017**

Au cours de leur première année de vie, les nourrissons se sensibilisent à leur langue maternelle à une vitesse phénoménale, avec peu ou pas de supervision de la part de leur environnement linguistique. Si les recherches en psychologie expérimentale ont, au cours des trente dernières années, établies les grandes lignes de ce processus d'acquisition, on comprend encore très peu les mécanismes cognitifs qui rendent cette acquisition possible. En particulier, aucune théorie n'arrive à expliquer comment les deux composantes élémentaires de la langues que sont la structure sonore et la structure lexicale semblent apprises, non pas l'une après l'autre, mais simultanément par l'enfant.

Le projet BOOTPHON a pour objectif de construire une théorie quantitative de cet apprentissage précoce en s'appuyant des outils issus des sciences de l'ingénieur (traitement du signal, modèles probabiliste), et s'attachera à élaborer des algorithmes d'apprentissage statistiques non supervisés visant à découvrir la structure phonologique et lexicale dans plusieurs langues typologiquement distinctes dont on dispose de corpus de parole spontanée suffisamment étendus (français, japonais, hollandais).

L'objectif de ce travail théorique est de reproduire, et par là même éclairer les grandes acquisitions de l'enfant, pendant sa première année de vie. Cette démarche 'synthétique' sera complétée par une validation (ou réfutation) empirique des hypothèses proposées, en utilisant les méthodes traditionnelles de la psychologie du développement (tests comportementaux, imagerie cérébrale, collecte et analyse de corpus d'interaction parent-enfant).

Ce programme se déroulera sur 5 ans, et impliquera de nombreuses collaborations avec des collègues issus de différentes branches des sciences formelles (traitement automatique des langues, apprentissage machine, mathématiques appliquées), des sciences du langage (psycholinguistique, phonologie et phonétique, socio-linguistique, linguistique computationnelle) et des sciences expérimentales (psychologie cognitive, psychologie du développement, neurosciences cognitives), en France, en Europe et aux Etats-Unis.