

Student projects

Projet « Trump sucks »

Interested in: psycholinguistique computationnelle (i.e., IA « cognitive ») ; perception de parole

Want to learn: utiliser les outils standard de reconnaissance de ; construire des tests « cognitif » pour les systèmes de reconnaissance de parole automatique ; (possible selon temps et intérêt) passer des expériences comportementales de perception de parole avec les êtres humains, analyser les résultats (analyse de données expérimentales) ; (possible selon temps et intérêt) entraîner les réseaux neuronaux profonds en tant que modèle cognitif

Prior background: forte compétence en programmation (surtout en Python) ; des compétences en phonétique et/ou en machine learning seraient utiles

La psycholinguistique computationnelle consiste à entraîner des modèles computationnelles de (certains aspects de) l'être humain. Le défi est d'avoir des façon d'évaluer ces modèles qui nous permettent de comparer le comportement de l'être humain avec celui du modèle. Au cours de ces dernières années, nous avons vu (a) un fort intérêt en psycholinguistique computationnelle sous l'égide de « cognitive AI », les systèmes qui font des tâches quasi-humaines (reconnaissance de parole, reconnaissance des objets, traduction) non pas juste d'une façon utile mais aussi d'une façon similaire à celle de l'être humain ; et (b) le développement de plusieurs évaluations psycho-computationnelles pour le cas de la reconnaissance de parole : c'est à dire, des « expériences comportementales pour machines » qui correspondent à des véritables expériences de perception de parole chez les humains. L'idée de ce projet est d'appliquer cette idée d'évaluation psycholinguistique à des modèles de reconnaissance de paroles dans un cas intéressant de perception non-natif. Plusieurs dialectes de l'anglais distinguent les voyelles brèves des mots STRUT et FOOT, y compris les dialectes standards qu'on entend à la télévision et qu'on apprend à l'école. La voyelle de STRUT est distingué notamment par le fait qu'elle se produit sans arrondissement de lèvres. Les locuteurs de plusieurs langues qui manquent ce phonème (espagnol, bosnien/croate/serbe) ont tendance à le percevoir comme [a]. Mais il semble, de façon anecdotale, que les francophones ont tendance à entendre d'autres voyelles, que ça soit le [œ] de «sœur» ou le [ɔ] de «potable» dans les mots comme SUCKS ou peut-être le [œ̃] ou [ɛ̃] de «un» dans les mots comme TRUMP. Mais on dirait que le [a] de «chat» reste quand même assez proche, et il n'est donc pas clair pourquoi on observe cette différence (semblable) entre les francophones et les locuteurs d'autres langues. En préparation pour une vrai expérience de perception pour mieux comprendre pourquoi on voit cette différence entre les langues, on cherche une ou des lignes de bases computationnelles, donc, des évaluations du comportement d'un modèle de perception de parole basé sur l'acoustique (e.g., un modèle de reconnaissance de parole standard).

Dispersion

Interested in: typologie ; phonétique ou phonologie

Want to learn: typologie quantitative

Prior background: programmation R et Python ; des compétences en statistique seraient utiles

Avec des grandes bases de données prises de beaucoup de langues, on peut découvrir plusieurs tendances intéressantes. Par exemple, un de nos papiers récents a montré qu'il existe des tendances vers certaines symmétries abstraites dans les inventaires phonémiques des langues du monde. Ça a été montré en utilisant beaucoup de langues—une sorte de « big data » typologique—et des techniques statistiques spécialisées. Par contre, une des grandes affirmations typologiques par rapport aux inventaires phonémiques, l'affirmation qu'il existe une tendance vers la « dispersion »—que les phonèmes cherchent les réalisations plus distinctes l'un de l'autre—n'a jamais été montrée empiriquement. Dans ce projet, on va essayer de valider cette affirmation.

Expériences de perception de parole et/ou de langue artificielle

Interested in: phonétique, phonologie ou morphologie

Want to learn: mise en oeuvre et analyse d'expériences psycholinguistiques

Prior background: compréhension de phonétique et phonologie, surtout de l'anglais et du français ; un peu d'expérience antécédente avec le montage et l'analyse des expériences serait utile

Les inventaires phonologiques, les règles phonologiques, et les paradigmes morphologiques ont tous tendance à être systématiques. Par exemple, en phonologie, on trouve que les inventaires ont souvent [p t k b d g] mais jamais [p q' c b Ø g] et, en morphologie, on trouve souvent que, par exemple, il y a une seule terminaison pour l'accord verbal au pluriel, partagée entre 1ère, 2è, et 3è personne au pluriel, mais non pas qu'il y a, par exemple, une terminaison qui est partagée entre la première personne au pluriel et la deuxième et la troisième personne au singulier. En préparation pour un projet de recherche qui débute ici l'année prochaine où on va faire des expériences psycholinguistiques pour voir si on peut observer ces tendances vers la systématisme pendant (i) la perception des sons non-natifs et (ii) l'apprentissage des patrons non-linguistiques (pour essayer, en fin de compte, d'expliquer ces phénomènes linguistiques), on doit reposer sur des expériences pilotes qui montre que la méthodologie expérimentale (y compris certaines nouvelles tâches expérimentales) et la méthodologie d'analyse statistique est valide. Dans ce projet exploratoire, alors, on construit des stimuli, on passe les manips (surtout en ligne), et on fait l'analyse statistique. Si tout se passe bien, une suite au sein de projet de recherche, une fois débuté (pendant le M2) serait également prévisible.