

Discrimination Automatique des Sens d'Emploi des Mots par l'Intonation

François NEMO

Mélanie PETIT

DIASEMIE

CaSciModOt

26 novembre 2013

Relation entre son et sens

- Tout signe associe une forme sonore et un sens
- Tout énoncé associe une forme sonore (une intonation) et une phrase
- On distingue les unités sonores de base (phonèmes) de l'intonation qui elle est dite suprasegmentale (prosodie) car elle porte sur un ensemble de phonèmes

Qu'est-ce qu'une onde sonore ?

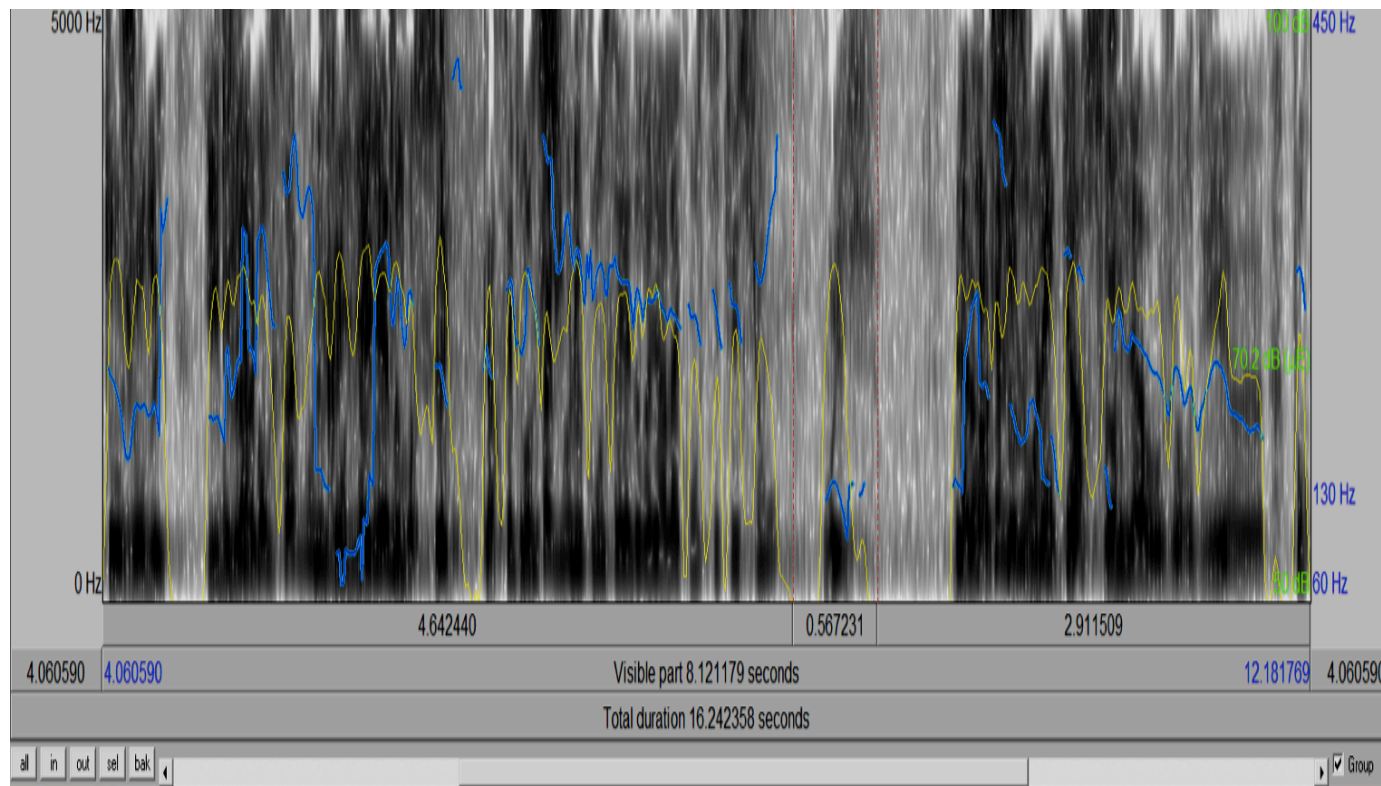
« La phonétique s'intéresse à la production des sons chez l'être humain, sans se préoccuper a priori de la notion de sens. Elle est liée à la physiologie, puisque la connaissance des organes vocaux permet de comprendre la production des sons. Elle est liée également à l'acoustique qui permet l'analyse des sons produits. »

(<http://m.antoniotti.free.fr/phonetique.htm>)

Qu'est-ce qu'une onde sonore ?

« Un son est le résultat de la mise en vibration de l'air. Celle-ci peut être obtenue par la mise en vibration d'un corps, qui transmet sa vibration à l'air ambiant, comme sur un instrument de percussion ou un instrument à cordes. Le son peut être obtenu en agissant sur de l'air en mouvement, en modifiant son écoulement, comme l'air sous une porte, ou dans un instrument à vent. On peut classer les sons en deux catégories : périodiques et apériodiques. Ces deux types de sons donnent une sensation de hauteur, mais les seconds sont qualifiés de bruits. L'appareil vocal humain permet de produire les deux types de sons, voire de les combiner. »

Spectrogramme



Principaux paramètres prosodiques

- Fréquence
- Intensité
- Longueur
- Rupture ou intégration prosodique
- Pauses

Sens d'emploi des mots

- La plupart des mots ont plusieurs sens (polysémie). Expliquer la polysémie est une des deux tâches de base de la sémantique
- Depuis une quinzaine d'années les théories de la polysémie ont commencé à se poser la question du rôle de la prosodie dans la construction et la discrimination du sens

Travaux

- Les premiers travaux consacrés exclusivement à cette question ont moins de 10 ans (e.g. Petit, 2009)
- Ils ont permis de montrer que la prosodie joue au niveau lexical un rôle majeur dans la polysémie et que la prise en compte de la prosodie est indispensable en sémantique

Exemple

- Polysémie de « enfin » :
 - Irritation
 - Soulagement
 - Reformulation
- Polysémie et prosodie : découverte de sens d'emplois de « enfin » : l'interprétation-type de « soulagement » s'avère avoir deux emplois-types (sens d'emplois) 📢 📢

La question de l'automatisation

- Travailler sur la polysémie suppose d'étudier des centaines d'emplois d'un mot
- Travailler sur la prosodie lexicale suppose d'étudier des milliers d'emplois d'un mot
- Le linguiste est amené sur de tels corpus à faire des navettes constantes entre forme prosodique et interprétation
- Il est confronté à des problèmes de discrimination, qui implique une quantification fine des phénomènes

Le projet DIASEMIE

- Il associe linguistes, informaticiens, mathématiciens et spécialistes du traitement du signal dans un quadruple objectif :
 - Mettre au point une chaîne de traitement automatisée des corpus
 - Utiliser des outils mathématiques de discrimination (classification)
 - Cartographier tous les sens d'emplois de « oui » et « yes »
 - Mettre au point un outil de discrimination automatique brevetable



Etapes de la démarche

- Recueil des données
- Création de banques d'emplois
- Caractérisation sémantico-pragmatique
- Caractérisation prosodique
- Classification
- Navette
- Définition des clefs de discrimination
- Implémentation des clefs de discrimination pour permettre un traitement automatique du sens

Recueil des données

- Corpus :
 - Diversité des sources et des situations (exhaustivité)
 - Diversité des locuteurs (variabilité)
- Extraction :
 - Découpage
 - Transcription
- Obtention d' une banque d' emplois (population)

Caractérisation sémantico-pragmatique

- Etape cruciale et complexe mobilisant une dizaine de linguistes
- Obtention d'un ensemble de caractérisations testables
- Plurisémie (diversité des strates interprétatives associées à un emploi)
- Métalangage à construire

Caractérisation prosodique

- Extraction de données numériques

Time_s	F0_Hz
8.927429	126.774419
8.939929	132.139510
8.952429	136.922802
8.964929	139.203823
8.977429	138.826809
8.989929	137.389701
9.002429	135.069914
9.014929	130.154839
9.027429	124.717814
9.039929	118.543486
9.052429	109.161091
9.064929	101.382775
9.077429	135.366470
9.089929	136.907615
9.102429	137.128086

Classification

- Sur la base de l'intégration à la banque d'emplois des descriptions sémantico-pragmatiques d'une part et prosodiques d'autre part, lancement de la phase de classification (automatique)
- Hétérogénéité des conditions d'enregistrement + variabilité contraignent la classification (alignement)

Navette

- En fonction des résultats obtenus, tests d'homogénéité sur des sous-ensembles
- Explication de l'éventuelle hétérogénéité

Définition des clefs de discrimination

- Identification de critères de discrimination
- Identification des zones de discrimination insuffisantes
- Croisement des tests de perception et des caractérisations prosodiques (granulométrie)

Modélisation mathématique

- La question fondamentale qui se pose est de savoir si les outils mathématiques standards sont suffisants pour traiter cet ensemble de problèmes et les étapes successives du projet
- La question annexe s'il est apporté une réponse négative à la première est de savoir quels outils il est possible de créer