# Analyser la variation syntaxique
# Un corpus annoté du suisse allemand

Eric Haeberli & Manuela Schönenberger

## Point de départ

- Background:
  - Le changement syntaxique (e.g. dans l'histoire de l'anglais (vieil anglais, moyen anglais, anglais moderne)).
  - Basée sur des corpus historiques électroniques syntaxiquement annotés.
  - Ces corpus sont une source riche de données pour l'étude du changement syntaxique. Mais ils ont des limites:
    - Le hasard détermine ce que nous pouvons étudier.
      - Aucun contrôle sur des variables importants concernant les locuteurs (âge, dialecte, sexe).
      - Quantités de données parfois très limitées.
    - Langue écrite vs. langue parlée.



## La variation

- Un phénomène central dans l'étude du changement linguistique: La variation (entre les locuteurs d'une même langue/d'un même dialecte et dans le langage d'un seul individu).
  - Lorsqu'un trait linguistique A est remplacé par un trait linguistique B,il y a nécessairement une phase de transition pendant laquelle et A et B sont utilisés par les locuteurs.
- Mais:
  "Not all variability and heterogeneity in language structure involves change, but all change involves variability and heterogeneity". (Weinreich, Labov & Herzog 1968:188)



## La variation

- La variation et son lien avec le changement ont été étudiés de façon détaillée depuis les années 1960, suite aux travaux de William Labov.
- Mais: Une grande partie des travaux dans cette tradition se concentrent sur la phonologie.
- La variation dans les domaines de la syntaxe et de la morphosyntaxe est relativement peu étudiée.



## Projet

- Projet "Studying Variation in Syntax: A Parsed Corpus of Swiss German" (SNF Projekt 146450; 2014–2016).
- Buts:
  - Création d'un corpus électronique d'un dialecte du suisse allemand qui permet l'analyse détaillée de la variation syntaxique:
    - Transcriptions d'entretiens avec des locuteurs
    - Environ 750'000 mots
    - Environ 40-50 locuteurs divisés en trois tranches d'âge (20-30, 40-55, 70+), équilibre hommes/femmes, différentes classes sociales.
    - Annotation: POS-tagging , parsing
  - Etudes de cas dans le domaine de la variation syntaxique et le changement syntaxique ("apparent time").
    - A long terme: "Real time" avec des corpus futurs.



## Pourquoi le suisse allemand?

- Dans les dialectes du suisse allemand, un bon nombre de propriétés (morpho-)syntaxiques sont variables.
  - E.g. Verb (Projection) Raising.
  (1) a. ... dass er am Òòbig es Lied *vòòrsinge wött* V2 V1
  b. ... dass er am Òòbig es Lied *wött vòòrsinge* V1 V2
  *... que il au soir une chanson {veut chanter/chanter veut}*
- Les dialectes du suisse allemand vivent dans un "habitat" favorable à l'étude de la variation:
  - Pas de stigmatisation du dialecte; utilisation dans le quotidien de tous les groupes sociaux.
  - Pas de pressions normatives (pas écrit, pas enseigné à l'école).
  - Résistants aux influences du standard.

## Le corpus

- Les données seront disponibles dans plusieurs formats:
  - Texte sans annotation.
  - Texte avec POS tagging.
  - Texte avec Parsing (cf. diapositif 8).
  - Texte avec lien au fichier sonore – alignement des données (EXMARaLDA, cf. diapositif 9).

7

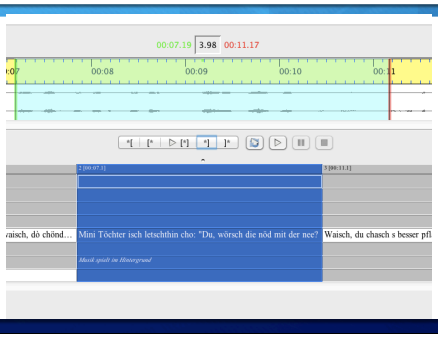
## Parsing

```
Mini Töchter isch letschthin cho: "Du, wörsch die nöd mit der nee?"

(IP-MAT (NP-SBJ (PRO$ Mini) (N Töchter))
        (BEP isch)
        (ADVP (ADV letschthin))
        (VBN cho)
        (, :)
        (" ")
        (CP-QUE-SPE
                (NP-VOC (PRO Du))
                (, ,)
                (IP-QUE-SPE (RDKII wörsch)
                            (NP-SBJ *pro*)
                            (NP-OB (D die))
                            (NEG nöd
                            (PP (P mit)
                               (NP (PRO der)))
                            (VB nee)
                            (, ?)))
                            (" "))
```

8

## Texte – fichier sonore




## Le rôle de l'informatique

- L'analyse quantitative et qualitative de la variation est facilitée énormément grâce à l'obention simple et rapide de données pertinentes.
- Le corpus peut être utilisé pour des questions de recherche dans n'importe quel domaine de la linguistique.
  - Et: Grâce a l'accès aux textes et aux fichiers sonores, des questions d'interfaces peuvent être traitées facilement (e.g. le rôle de la prosodie en syntaxe).

10