

Longitudinal UC Corpus of English Accents (LUCEA)

Hugo Quené & Rosemary Orr

Utrecht Inst voor Linguïstiek OTS & University College Utrecht



Universiteit Utrecht



university college utrecht

Vragen

convergentie: gesprekspartners zijn geneigd om zich aan elkaar aan te passen in taal en uitspraak (accommodatie) [1], verkleint sociale afstand [2]

- ook bij “lingua franca” (voertaal)?
- wat is eindpunt van convergentie, welk accent resulteert? [3]
- longitudinale spraakopnames
- fonetische analyses
- verstaanbaarheidsproeven

Uitdagingen

- consistentie gedurende 6 jaar (~12 interviewers)
- logistiek (alleen ‘s avonds, op campuslocatie, ~900u)
- validatie

werd inderdaad X gezegd?
- transcriptie

wat werd gezegd?
- bescherming van privacy...

“I was so mad at my friend Lisa for sleeping with Thomas...”

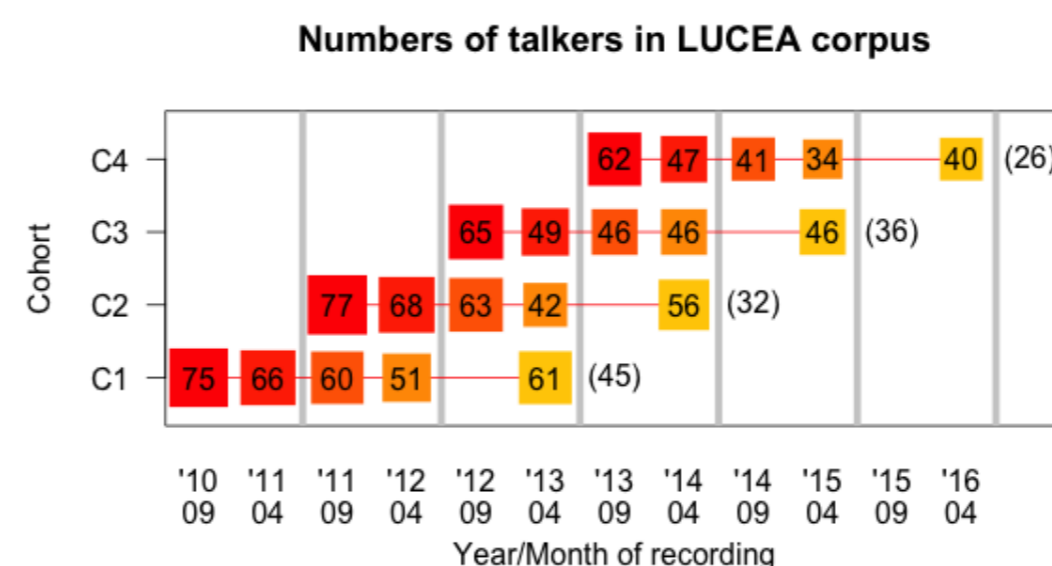
...met toekomstige sprekerherkenning
- koppeling aan juiste metadata (statisch, dynamisch) over 5 opnames
- data management (elk interview ~20 GB, tijdrovende backups), totaal ~2.5 TB



Corpus

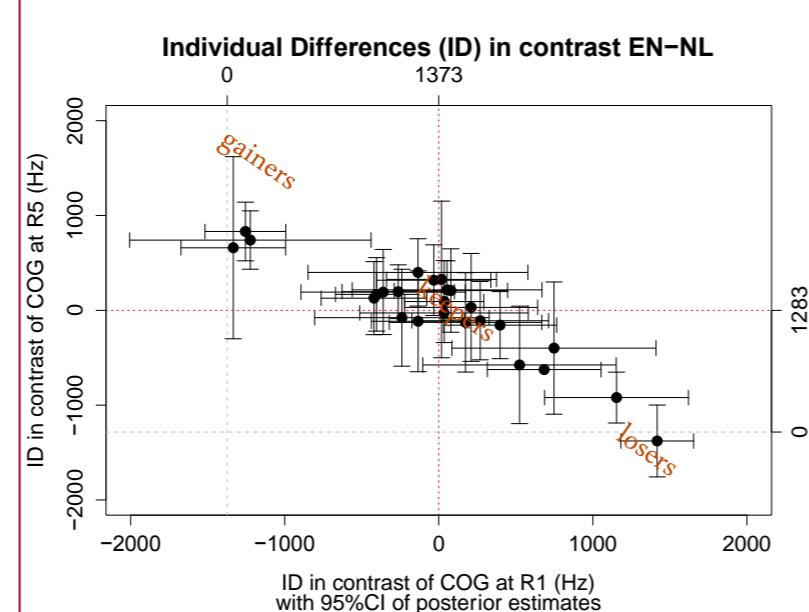
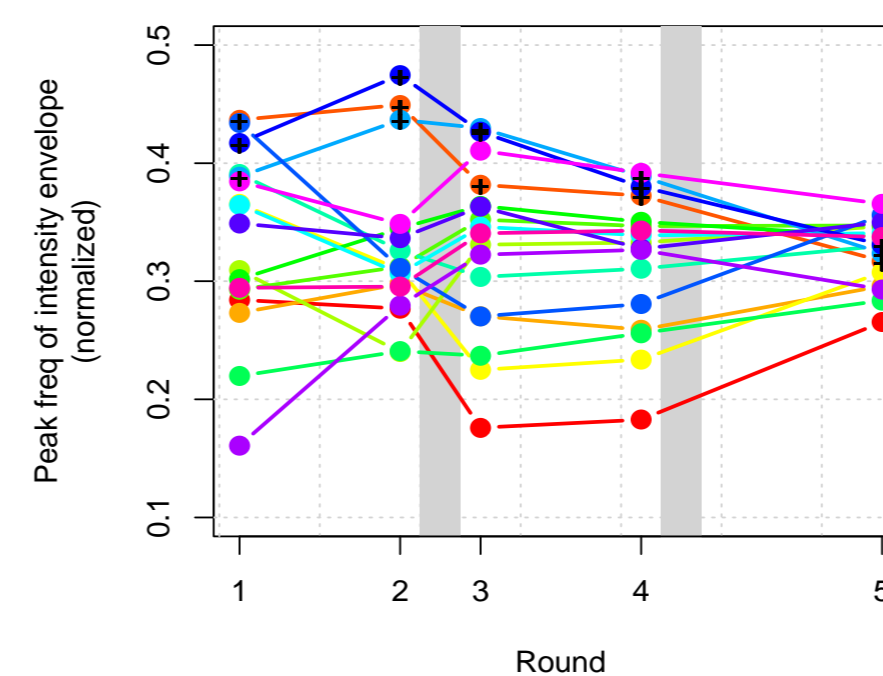


- n=285 UCU studenten uit cohorten 2010-2013 (32%)
- moedertaal ca 60% Nederlands, 5% Engels, 35% rest
- 5 interviews, **longitudinaal**, gedurende 3 jaar
- reeks is compleet voor 139 studenten (49%)
- totaal 1095 interviews
- plus interviews van interviewers, docenten, e.a.
- **Engels:** voorgelezen zinnen en teksten, delen *UN Decl Human Rights*, informele en formele monologen, dialoog.
- **L1:** vert. *UN Decl Human Rights*, informele monoloog
- metadata: vragenlijsten, audiometrie, administratie



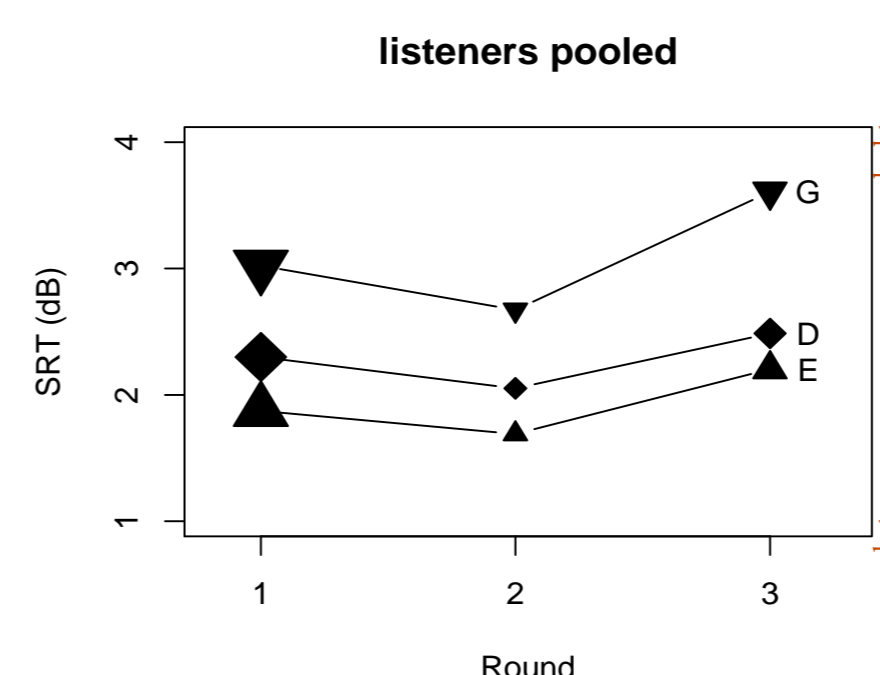
Enkele resultaten

- 1. ritme:** reductie van onbeklemtoonde lettergr Dutch L2ers: **geen verandering** English L1ers: convergentie naar (Ned) **L2** waarden! afname van var tussen sprk



- 2. is /s/ scherper** in Engels dan in Nederlands? R1 en R5: ja, contrast is aanwezig, **geen verandering** maar interessante **individuele verschillen**

- 3. is geconvergeerde spraak** (uit R2 of R3) beter **verstaanbaar** dan uit R1? R1: zonder convergentie R2: lagere SRT, en minder variantie in SRT tussen sprekers en tussen luisteraars, **beter verstaanbaar** R3 na zomerreces: **als R1** geen “interlanguage” voordeel!



Verspreiding

- **corpus is beschikbaar**, inclusief metadata
 - eerste helft reeds online; tweede helft binnenkort
 - nieuwe ISOcat labels <http://www.isocat.org/rest/dcs/649> regardSelfAsNativeSpeaker(S, L)
 - opslag conform richtlijnen van CLARIN-NL
1. MPI: The Language Archive
 2. Meertens Institute
 - beschikbaar “for linguistic research and development” na instemming met gebruiks-voorwaarden
 - anonimiteit van sprekers moet beschermd blijven

Toekomst

- onderzoek naar **integratie** van productie en perceptie
 - zelfde individuen als sprekers en als luisteraars, en correleer bevindingen
- **netwerk-analyse**
 - is convergentie gerelateerd aan sociaal prestige? is het symmetrisch? wie convergeert naar wie?
- **fonetische transcriptie** van Engelse monologen, m.b.v. spraakherkenning

Bibliografie

- [1] H. Giles, N. Coupland, J. Coupland, in *Contexts of Accommodation: Developments in applied sociolinguistics* (CUP, Cambridge, 1991), pp.1-68.
- [2] J. Pardo, *J. Acoust. Soc. Am.* 119(4): 2382-2393 (2006).
- [3] J. E. Flege, in *Speech Perception and Linguistic Experience: Issues in Cross-language research* (York Press, Timonium, MD, 1995), pp. 229-273.

Informatie

lucea.wp.hum.uu.nl



Met dank aan alle interviewers en alle deelnemers, en aan David van Leeuwen voor technische ondersteuning.

Hugo Quené www.hugoquene.nl