



UC Leuven
Limburg

MOVING MINDS

Lexicale rijkdom in het professioneel hoger onderwijs: Aanzet tot sociolinguïstische staalkaart

Jose TUMMERS^{1,2} – Annelies DEVENEYNS¹

¹ UC Leuven-Limburg

² KU Leuven – Quantitative Lexicology and Variational Linguistics

Inhoud

1. Dure woorden in het (professioneel) hoger onderwijs
2. Lexicale rijkdom
3. Datagaring
4. Sociolinguïstische studentenkenmerken
5. Resultaten
6. Bespreking

Inhoud

- 1. Dure woorden in het (professioneel) hoger onderwijs**
2. Lexicale rijkdom
3. Datagaring
4. Sociolinguïstische studentenkenmerken
5. Resultaten
6. Bespreking

1. Dure woorden in het hoger onderwijs

- **Overgang secundair onderwijs – hoger onderwijs**
 - Nederlands: kloof tussen eindtermen secundair onderwijs en impliciete begintermen hoger onderwijs (Bogaert & Verheyden 2011; Van den Branden 2010)
 - Probleem: academisch en zakelijk taalgebruik
 - Probleem stelt zich des te meer in **professioneel** hoger onderwijs
 - Getrouwere afspiegeling diverse uitstroom secundair onderwijs
 - Studenten uit volledig spectrum van ASO over TSO en KSO tot BSO

1. Dure woorden in het hoger onderwijs

- **Overgang secundair onderwijs – hoger onderwijs**
- **Correlatie**
 - tussen **taalvaardigheid** en **studievoortgang** (De Wachter & Hee 2013)
 - tussen **score essay** en **lexicale rijkdom** (Lemmouh 2005)

1. Dure woorden in het hoger onderwijs

- **Taalbeheersingsonderzoek**

- Aandacht voor gevorderde moedertaalgebruikers: onderzoek naar taalvaardigheid studenten hoger onderwijs (*inter al.* Berckmoes & Rombouts 2009; Deveneyns & Tummers 2013; Peeters & Van Houtven 2010; Van Houtven et al. 2010)
- Hernieuwde aandacht voor tekst en kwantitatieve tekstanalyse (Crosley & McNamary 2011; Yannakoudakis et al. 2011), maar vooral in kader van vreemdetalenonderwijs

- **Lexicaal onderzoek**

Verband tussen lexicale rijkdom en sociolinguïstische determinanten
(Van Gijssel 2007)

1. Dure woorden in het hoger onderwijs

- Dure woorden in het (professioneel) hoger onderwijs: problematiek van **taalkloof / stijlkloof / registerkloof**
 - Studenten worden met een nieuw genre geconfronteerd dat ze zelf moeten eigen maken
 - Studenten – bij uitstek in het professioneel HO – moeten gesensibiliseerd worden voor de eigenschappen van het belang van dat genre (in bepaalde academische/professionele contexten)
- Onderzoeksvraag

In welke mate wordt de lexicale rijkdom in het professioneel hoger onderwijs (mede)bepaald door de sociolinguïstische kenmerken van de studenten?

Inhoud

1. Dure woorden in het hoger onderwijs
- 2. Lexicale rijkdom**
3. Datagaring
4. Sociolinguïstische studentenkenmerken
5. Resultaten
6. Bespreking

2. Lexicale rijkdom

- Onderzoeks aandacht

- **Taalbeheersingsonderzoek**

- Type/token ratio als maat van lexicale diversiteit (Breeze 2008) of lexicale variatie (Laufer & Nation 1995)
 - Lexicale originaliteit: $\frac{\text{auteurspecifieke tokens}}{\text{tot. tokens}}$ (Laufer & Nation 1995)
 - Lexical Frequency Profile (Lemmouh 2005; Li 1997)
 - % 1000 en % 2000 meest frequente woorden
 - % academische woorden
 - Lexicale dichtheid: $\frac{\text{lexicale tokens}}{\text{tot. tokens}}$ (Laufer & Nation 1995)

- **Lexicaal onderzoek**

2. Lexicale rijkdom

- Onderzoeks aandacht
 - **Taalbeheersingsonderzoek**
 - **Lexicaal onderzoek**
 - Modelleren distributie lexicale frequentie (Baayen 2001), met o.a. hapax legomena als indicator van productiviteit en creativiteit
 - Sociolinguïstische determinanten van lexicale rijkdom (Van Gijse 2007)

2. Lexicale rijkdom

- Kwantitatieve maten: complementaire invalshoeken
 - Lexicale **diversiteit**
 - Type/token ratio (TTR)
 - Type/token per woordsoort: N, V, A en FW
 - Lexicale **specificiteit**
 - Gemiddelde lexicale rang (CELEX, [Baayen et al. 1993](#)) en per woordsoort
 - Aantal hapax legomena en per woordsoort
 - Lexicale **densiteit**
 - Aandeel inhoudswoorden
 - Aandeel per woordsoort ([Van Gijssel 2007](#))

Inhoud

1. Dure woorden in het (professioneel) hoger onderwijs
2. Lexicale rijkdom
- 3. Datagaring**
4. Sociolinguïstische studentenkenmerken
5. Resultaten
6. Bespreking

3. Datagaring

- Steekproef
 - Eerstejaarsstudenten KHLeuven (UC Leuven)
 - Alle opleidingen (n = 13)
 - Professionele bachelors
 - Domeinen: bedrijfsmanagement, sociaal werk, lerarenopleiding, gezondheidszorg en technologie
 - Clustersteekproef: 1 (klas)groep per opleiding ([Billiet 2001](#))
 - Tijdstip: einde 2^e semester

3. Datagaring

- Schrijfopdracht

- Ecologische validiteit

- Schrijven van tekst i.p.v. invuloefening, MKV, ... (Hyland 2001; Mukherjee 2006)
- Verbergen van taalkundige doelstelling
- In computerlokaal, alle nuttig geachte middelen toegelaten (interne spellingcontrole, ...)
- Geroosterd in uurrooster geselecteerde (klas)groepen

3. Datagaring

- Schrijfopdracht
 - **Argumentatieve/persuasieve** tekst
 - Genre indicatief voor EQF-6 (EP & EC 2008)
 - Opbouw rond uitwerking van thema's (Berman & Nir-Sagiv 2007)
 - Haalbaarheid
 - 60 minuten voor tekst van 500 woorden
 - Vertrouwd onderwerp: sociale media

3. Datagaring

- Kwaliteitscontrole
 - Originaliteitscontrole d.m.v. *TurnItIn*: verwijdering plagiaat (n=3)
 - Verwijdering onleesbare (n=3) en lege documenten (n=1)
 - Woordsoortannotatie d.m.v. *Tadpole* (van den Bosch et al. 2007)
- Resulterende **dataset**
 - 346 teksten
 - $n_{\text{woorden}} = 193,560$ ($\bar{x} = 559.4$, $SD = 86.8$)

Inhoud

1. Dure woorden in het (professioneel) hoger onderwijs
2. Lexicale rijkdom
3. Datagaring
- 4. Sociolinguïstische studentenkenmerken**
5. Resultaten
6. Bespreking

4. Sociolinguïstische kenmerken studenten

- **Geneste variabelen**

- Kenmerk van clustersteekproef
- Frequent in pedagogisch/didactisch onderzoek
- Groepsvorming:
 - student \in opleiding
 - student \in secundaire school

- **Model met twee niveaus**

- Niveau 2: opleiding, secundaire school
- Niveau 1: student

4. Sociolinguïstische kenmerken studenten

- **Socio-demografische kenmerken**

- Secundaire school [niveau-2]
- Opleiding [niveau-2]
- Student [niveau-1]
 - Geslacht: man, vrouw
 - Leeftijd diploma secundair onderwijs
 - Diploma secundair onderwijs: aso, bso, kso, tso
 - Voorgeschiedenis hoger onderwijs: nvt, ho, univ
 - Thuistaal: ned, meertalig, anderstalig
 - Opleidingsniveau moeder: lo, so, ho

4. Sociolinguïstische kenmerken studenten

- **Statistische modellering**

- Hiërarchisch / mixed-effects model ([Gelman & Hill 2007](#); [Baayen 2008](#))
 - Random effect terms \sim niveau-2 (groepen)
 - Fixed effect terms \sim niveau-1 (individuele kenmerken)

- Generalized linear mixed-effects model – family = gaussian

- lme4 library in R ([Bates et al. 2013](#))

- Voor iedere lexicale rijkdom-maat

$$Y \sim (1 | \text{opleiding}) + (1 | \text{sec.school})$$

$$+ \text{geslacht} + \text{opl.so} + \text{ho} + \text{lft.sec} + \text{thui.staal} +$$

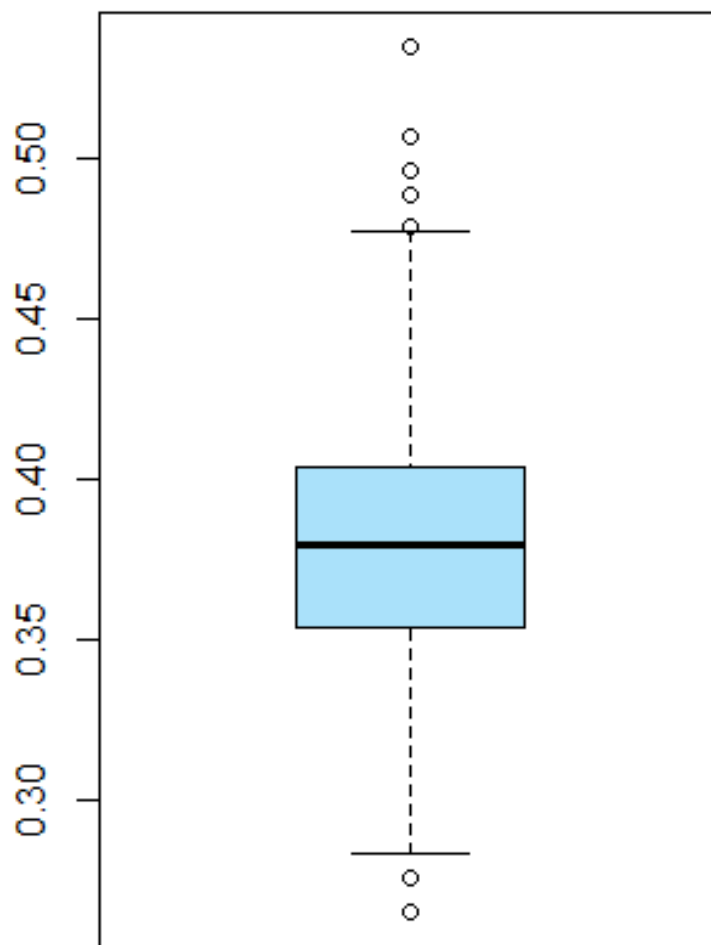
$$\text{opl.moeder}$$

Inhoud

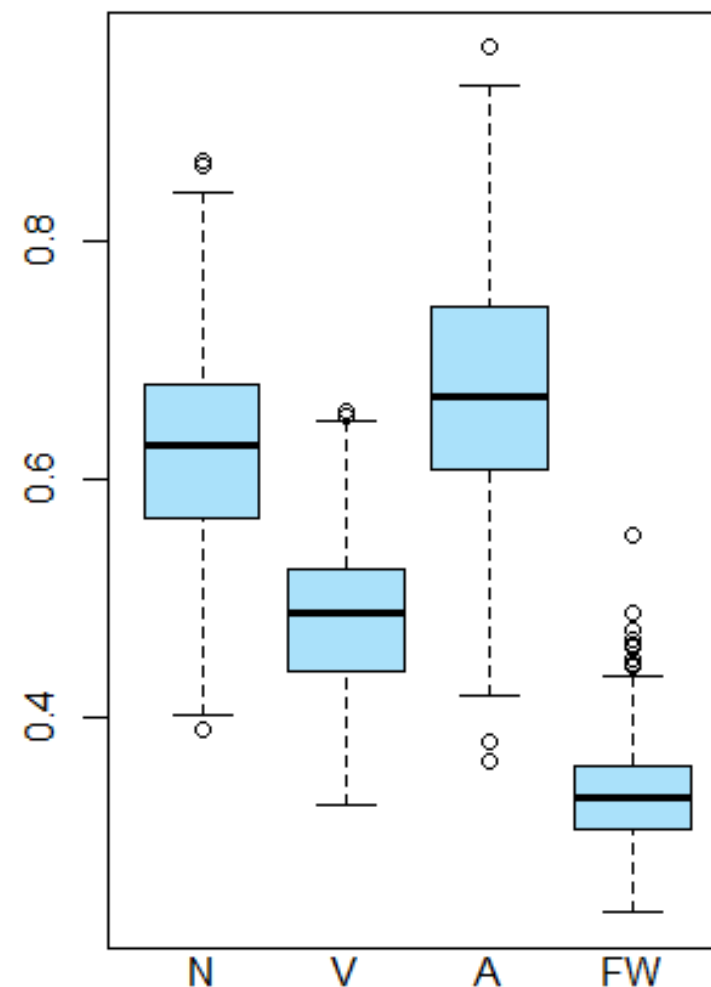
1. Dure woorden in het (professioneel) hoger onderwijs
2. Lexicale rijkdom
3. Datagaring
4. Sociolinguïstische studentenkenmerken
- 5. Resultaten**
 1. Lexicale diversiteit
 2. Lexicale specificiteit
 3. Lexicale densiteit
6. Bespreking

5. Resultaten – Lexicale diversiteit

TTR



TTR per woordsoort



5. Resultaten – Lexicale diversiteit

- Fixed effects

	geslacht	SO	leeftijd diploma SO	HO	thuis taal	opleid moeder
TTR	***	**	ns	ns	ns	ns
N.TTR	*	.	ns	ns	ns	ns
V.TTR	***	*	ns	ns	ns	ns
A.TTR	*	ns	ns	ns	ns	ns
FW.TTR	***	**	ns	ns	ns	ns

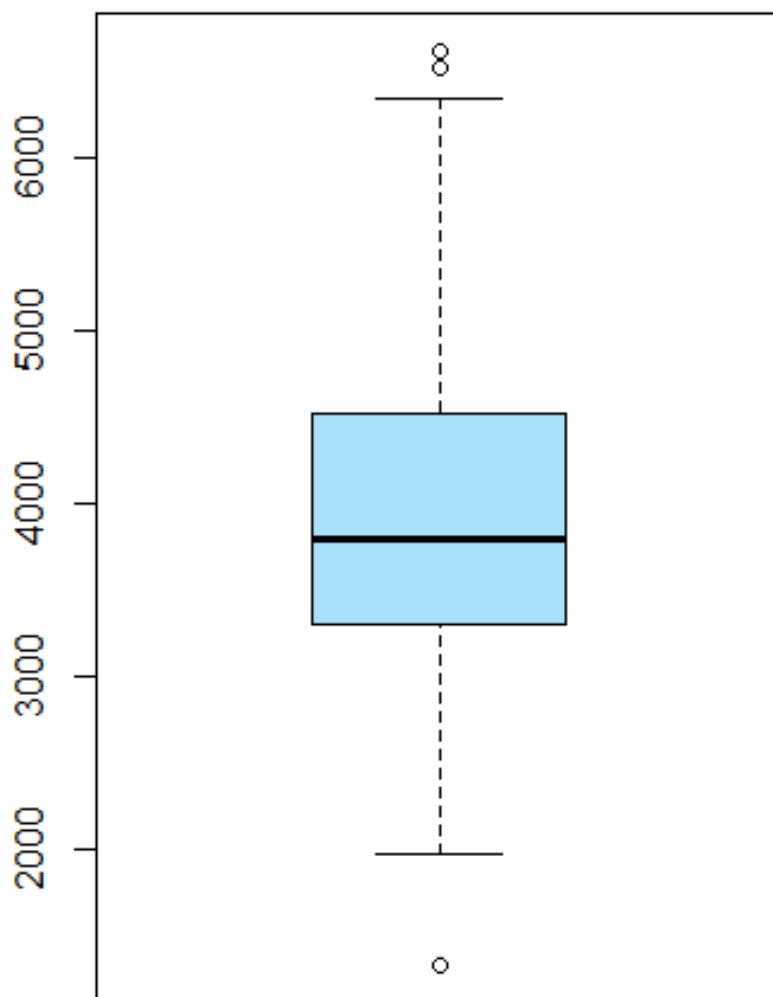
5. Resultaten – Lexicale diversiteit

- Random effects

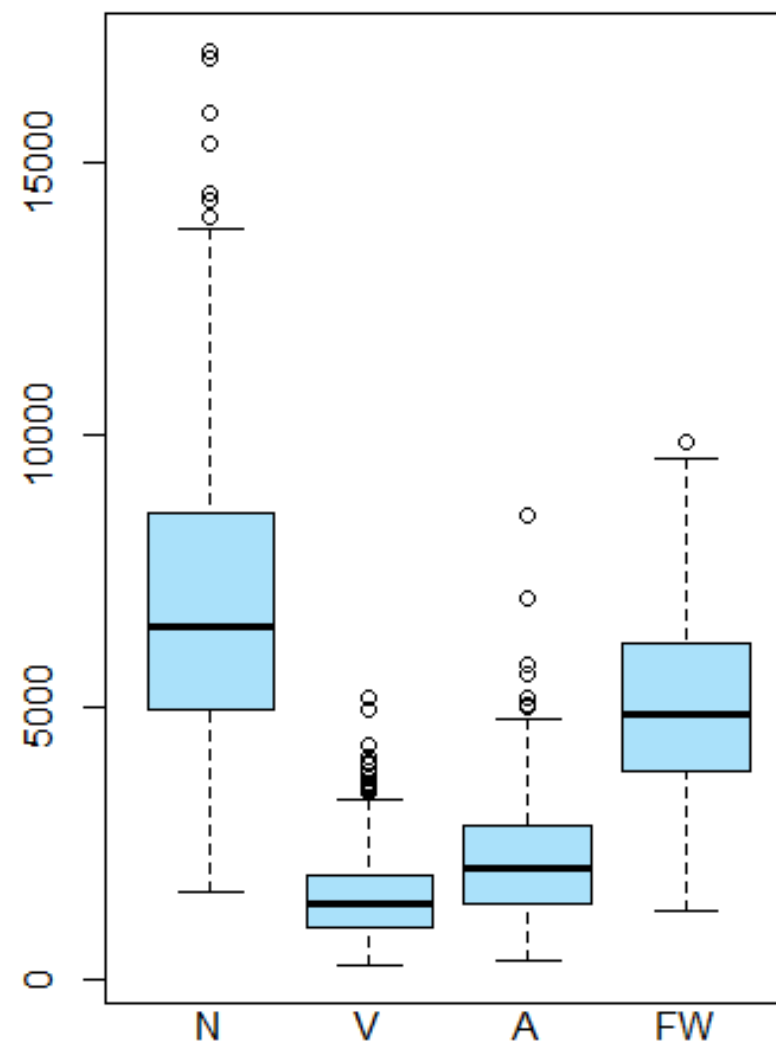
	ICC(opleiding)	ICC(secundaire.school)
TTR	0.0957	0.0426
N.TTR	/	0.0506
V.TTR	0.0573	0.0255
A.TTR	/	/
FW.TTR	0.0692	/

5. Resultaten – Lexicale specificiteit – rang

Gemiddelde rang



Gemiddelde rang
per woordsoort



5. Resultaten – Lexicale specificiteit – rang

- Fixed effects

	geslacht	SO	leeftijd diploma SO	HO	thuis taal	opleid. moeder
rang	***	ns	ns	**	*	ns
rang.N	*	ns	ns	.	ns	ns
rang.V	***	ns	ns	ns	ns	ns
rang.A	**	ns	ns	ns	ns	ns
rang.FW	***	ns	ns	*	.	ns

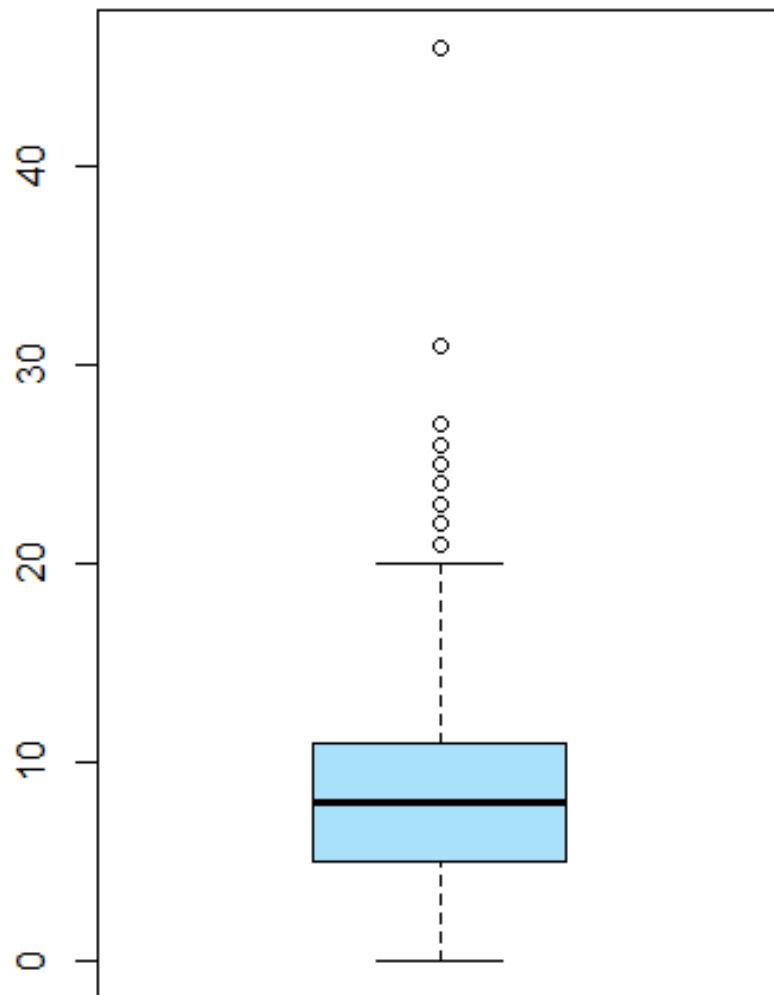
5. Resultaten – Lexicale specificiteit – rang

- Random effects

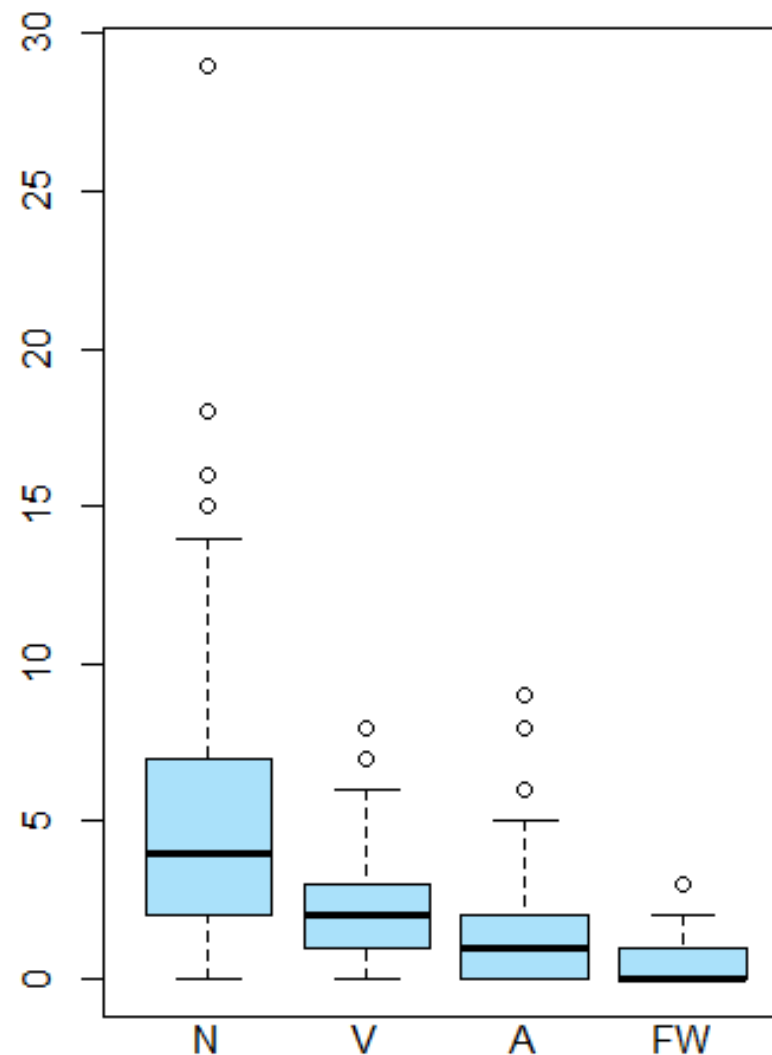
	ICC(opleiding)	ICC(secundaire.school)
rang	0.0317	/
rang.N	0.0432	0.0224
rang.V	0.0146	0.0293
rang.A	/	/
rang.FW	0.0514	/

5. Resultaten – Lexicale specificiteit – hapax

Aantal hapax legomena



Aantal hapax legomena per woordsoort



5. Resultaten – Lexicale specificiteit – hapax

- Fixed effects

	geslacht	SO	leeftijd diploma SO	HO	thuis taal	opleid moeder
hapax	***	ns	ns	.	ns	ns
hapax.N	***	.	ns	ns	ns	ns
hapax.V	*	ns	ns	.	ns	ns
hapax.A	***	ns	ns	ns	ns	ns
hapax.FW	ns	ns	ns	*	ns	*

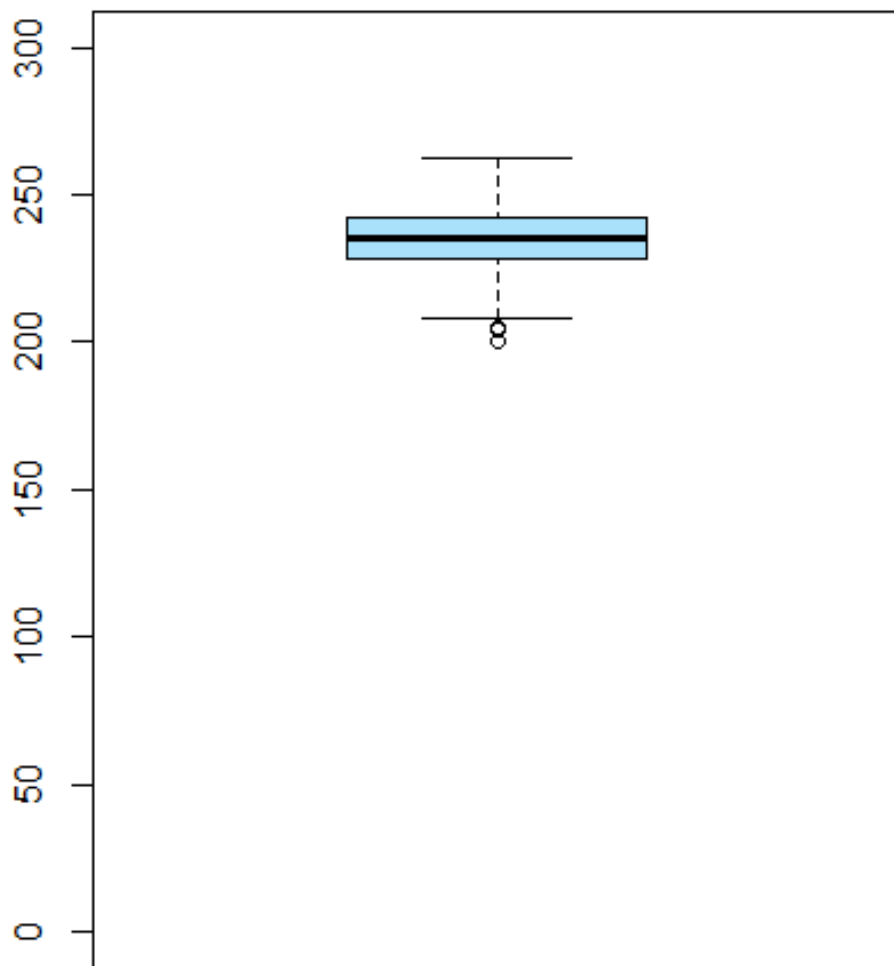
5. Resultaten – Lexicale specificiteit – hapax

- Random effects

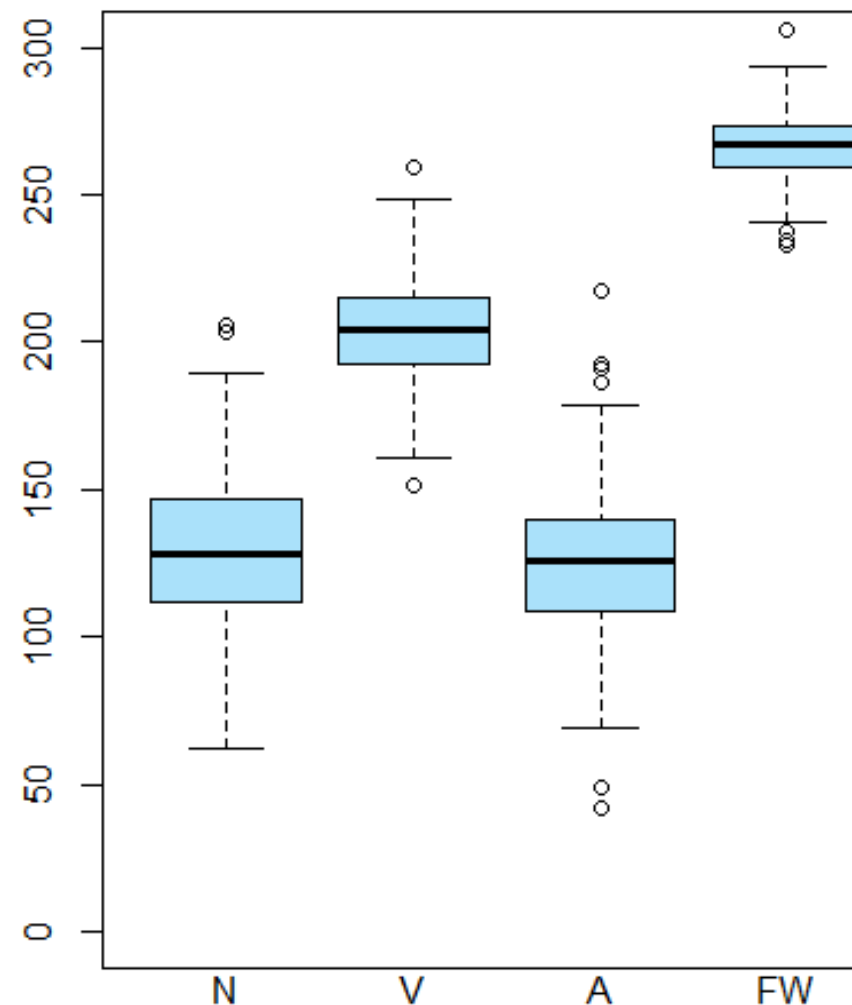
	ICC(opleiding)	ICC(secundaire.school)
hapax	/	0.0095
hapax.N	/	0.0117
hapax.V	/	0.0306
hapax.A	/	0.0206
hapax.FW	/	0.1000

5. Resultaten – Lexicale specificiteit – originaliteit

Lexicale originaliteit



Lexicale originaliteit per woordsoort



5. Resultaten – Lexicale specificiteit – originaliteit

- Fixed effects

	geslacht	SO	leeftijd diploma SO	HO	Ned in opl HO	thuis taal	ople moed
orig	***	**	.	ns	ns	ns	ns
orig.N	**	ns	ns	ns	ns	.	ns
orig.V	***	*	ns	.	ns	ns	ns
orig.A	*	ns	ns	ns	ns	ns	ns
orig.FW	*	**	ns	ns	ns	ns	ns

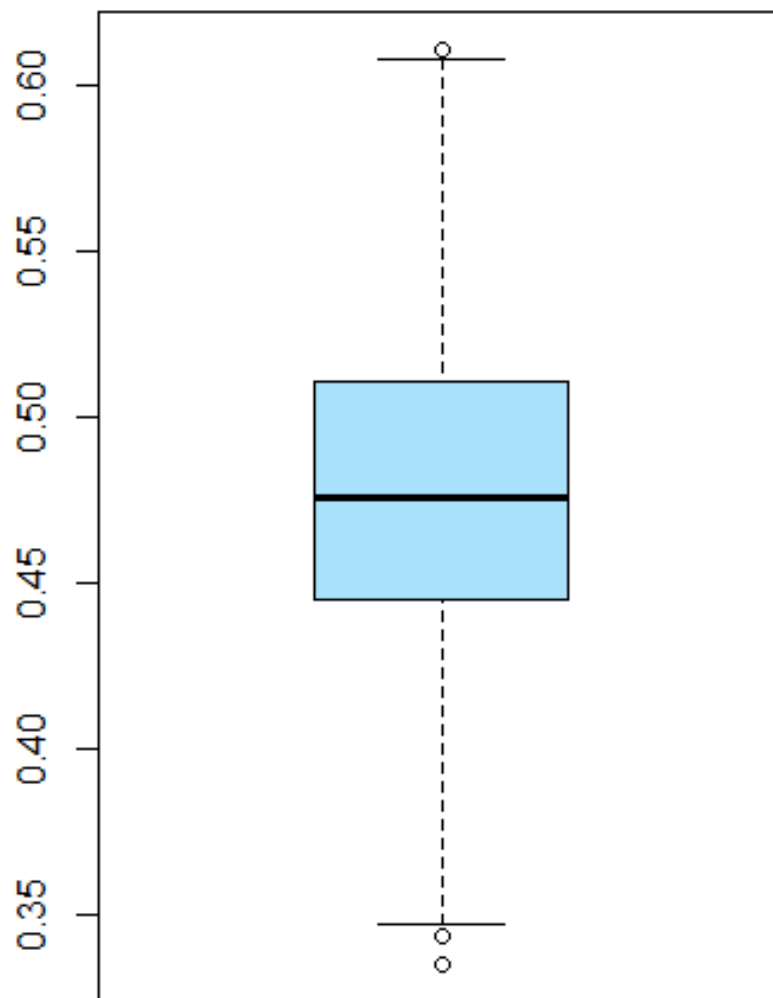
5. Resultaten – Lexicale specificiteit – originaliteit

- Random effects

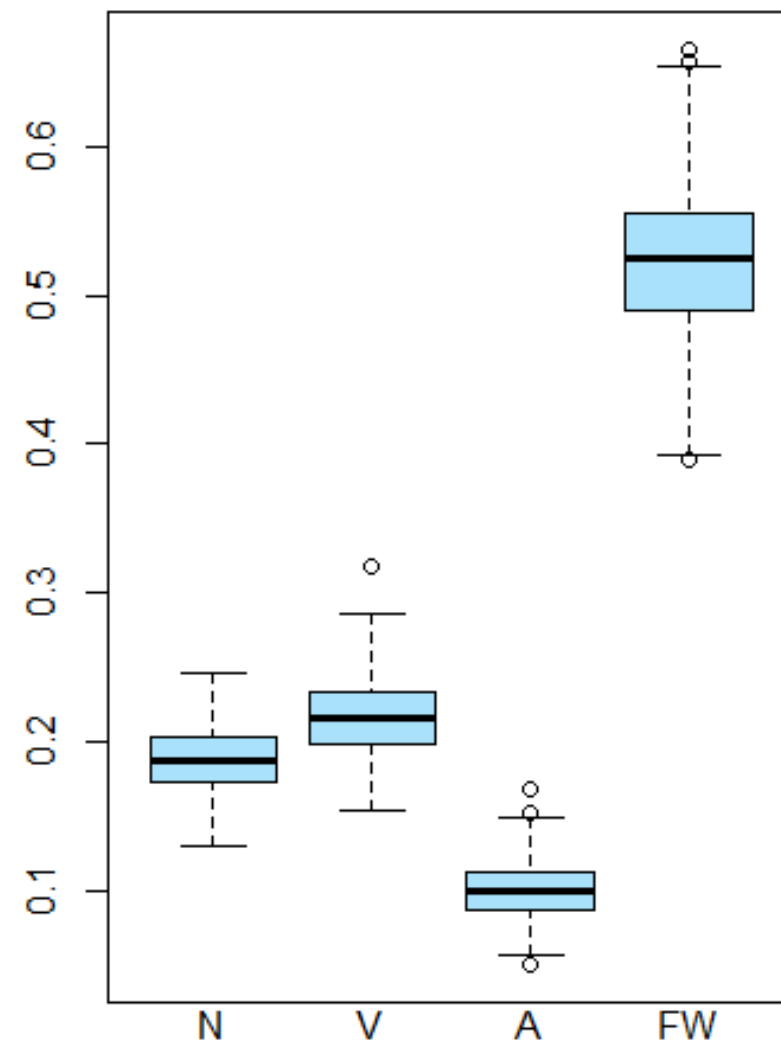
	ICC(opleiding)	ICC(secundaire.school)
orig	0.0057	0.0432
orig.N	/	/
orig.V	0.0371	0.0067
orig.A	/	/
orig.FW	0.0140	0.0478

5. Resultaten – Lexicale densiteit

Lexicale densiteit



Lexicale densiteit per woordsoort



5. Resultaten – Lexicale densiteit

- Fixed effects

	geslacht	SO	leeftijd diploma SO	HO	thuis taal	opleiding moeder
algemeen	***	.	ns	ns	ns	ns
N	***	ns	ns	ns	ns	ns
V	**	ns	ns	ns	ns	ns
A	ns	ns	ns	ns	ns	ns
FW	***	.	ns	ns	ns	ns

5. Resultaten – Lexicale densiteit

- Random effects

	ICC(opleiding)	ICC(secundaire.school)
algemeen	0.0385	/
N	/	/
V	/	/
A	/	/
FW	0.0385	/

Inhoud

1. Dure woorden in het (professioneel) hoger onderwijs
2. Lexicale rijkdom
3. Datagaring
4. Sociolinguïstische studentenkenmerken
5. Resultaten
6. **Bespreking**

6. Bespreking

- Resultaten – algemeen
 - Grote variatie over studenten in steekproef
 - Socio-demografische determinanten van lexicale variatie
 - Diversiteit: geslacht, SO
 - Specificiteit: geslacht > HO >> thuistaal, opl.moeder >> S
 - Densiteit: geslacht >>> SO
 - Bemerkingen
 - Beperkt aso-effect
 - Mannelijke prestaties ~ opbouw van tekst (Van Gijssel 2007)

6. Bespreking

- Resultaten – groepskenmerken
 - Opleiding
 - Effect
 - TTR, lexicale specificiteit_{rang}, lexicale densiteit
 - Secundaire school
 - Beperkt effect
 - TTR

6. Bespreking

- Resultaten – woordsoort
 - N: hoge TTR, hoge specificiteit, matige densiteit
 - V: matige TTR, lage specificiteit, matige densiteit
 - A: hoge TTR, lage specificiteit, lage densiteit
 - FW: lage TTR, lage specificiteit, hoge densiteit
- Beleidsmatige vertaalslag
 - Aandachtsgroepen op basis van geslacht, diploma SO, voorgeschiedenis HO
 - Belang en haalbaarheid van sensibilisering in HO
 - Vooral werken aan V, A en FW

6. Bespreking

- Beperkingen inherent deze *aanzet*
 - Beperkte steekproef
 - 1 hogeschool
 - 1 tekstgenre
- Toekomstig onderzoek
 - PCA om onderlinge verbanden te ontrafelen tussen determinanten
 - Samen met syntactische/grammaticale maten verbinden aan sco

Meer weten ?

- E-mail
 - jose.tummers@ucll.be
 - annelies.deveneyns@ucll.be
- Internet
 - <http://marco.khleuven.be>
 - <http://wwwling.arts.kuleuven.be/qlvl/>