Register and lexicon in new media in Dutch

Promotors: Dirk Geeraerts & Dirk Speelman

Jocelyne Daems

KU Leuven
Quantitative Lexicology and Variational Linguistics

Background

- QLVL: Quantitative Lexicology and Variational Linguistics

- MA thesis: Register analysis in blogs

- Research apprenticeship: Sociolectometric studies
  on a pluricentric language

Overview

Overview

1. Introduction

Questions & Answers

- Answer the why's, how's, which's
- How?
    - Register and lexicon in new media in Dutch

1. Introduction

Questions & Answers

- Answer the why's, how's, which's
- How?
    - Register and lexicon in new media in Dutch

1. Introduction

Questions & Answers

- Answer the why's, how's, which's
- How?
    - Register and lexicon in new media in Dutch

1. Introduction

Questions & Answers

- Answer the why's, how's, which's
- How?
    - Register and lexicon in new media in Dutch

1. Introduction

Questions & Answers

- Answer the why's, how's, which's
- How?
    - Register and lexicon in new media in Dutch

1. Introduction

Research Questions

- What does language use in new media in Dutch show us about how and for which purposes the new media are used?

- How informal is the language use in new media registers in Dutch?

- How does language use in new media in Dutch (both in terms of functionality as in lexical richness) differs from more traditional registers, and how do the new media registers vary among themselves?

1. Introduction

Research Questions

- Descriptive
  - Creating a global overview of (the functions expressed by / the lexical diversity in) new media in Dutch

- Methodological
  - Bringing together profile-based lexical variation and multi-dimensional register research

1. Introduction

Concrete questions

- Is there a continuum from text messages over blogs to Tweets in terms of involvedness?

- How informal is language in new media, and precisely when is it not?

- Are text messages indeed close to spontaneous conversations?

- Do emails have their traditional counterpart in the hand-written letter?

- Do speakers of Belgian Dutch stand out lexically in text messages, but not so on Twitter?

1. Introduction

Concrete questions

- Is there a continuum from text messages over blogs to Tweets in terms of involvedness?

- How informal is language in new media, and precisely when is it not?

- Are text messages indeed close to spontaneous conversations?

- Do emails have their traditional counterpart in the hand-written letter?

- Do speakers of Belgian Dutch stand out lexically in text messages, but not so on Twitter?

1. Introduction

Plane leaves at 6pm...
Greece here we come!
-with @Lisa Parker

Last night as I was surfing around the Web, I came across this amazing cooking site which I desperately want to share with you. _link_ It has the most exquisite step-by-step fish recipes which I just gotta try! The tips and tricks allow you to look like a 5* cook among your friends. An example: it's called the firecracker technique: [...]

Dear Jeff, Concerning the optimalization meeting to be held on Wednesday, I was informed that there would be an opportunity to present last semester's sales results. Having worked on the new sales model, I would by honoured to give my views on next term's expectations. Would this be possible? Best regards, Jonathan

Just saw Quintana conquer the mountain, what a great little man #tourdefrance #greatachievements

Hi Stacy, how're u? what've u been up 2 lately? we should get together soon, i'm in town next weekend

Mom, ok if i stay at lisas? see you at dinner

1. Introduction

Continuum

Involvedness   **Detachment**

fpp   passive
spp   nominalization
hedge
question
it
imperative

1. Introduction

Continuum

Involvedness   **Detachment**

txt msg   FB wall post   blog   Tweet   office mail
                              FB status

1. Introduction

Continuum

Involvedness   **Detachment**

txt msg   FB wall post   blog   Tweet   office mail
                           FB status
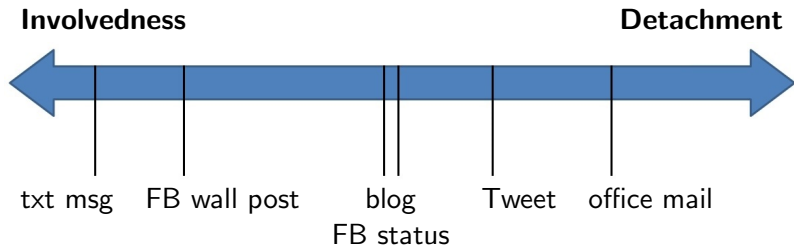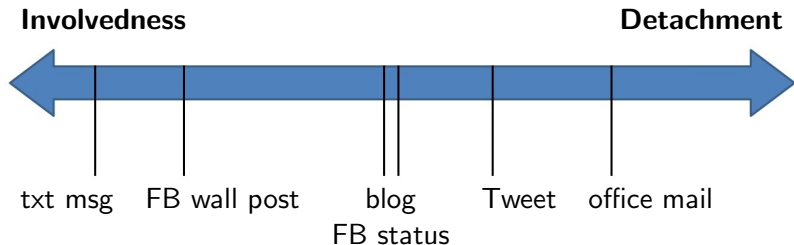
⇒ Continuum from text messages over blogs to Tweets

Overview

2. Dutch

Why Dutch?

- Native speaker
- QLVL tradition
- New

2. Dutch

Why Dutch?

- Native speaker
- QLVL tradition
- New

2. Dutch

Which Dutch?

- Pluricentric language (Clyne 1992)
  - ⇒ >1 national variety

  - Belgian Dutch
  - Netherlandic Dutch

How?

- Large corpora
  - QLVL: blog, Usenet, IRC chat, newspaper, Twitter, ...
  - SoNaR (interuniversitary STEVIN project): blog, chat, text message, Twitter, newsletter, e-magazine, ...

Overview

3. New media

Why new media?

- Newness
    - Dutch
    - QLVL
- In fashion

3. New media

Why new media?

- Newness
    - Dutch
    - QLVL
- In fashion

! Also problematic !

3. New media - Terminology & Definitions

3. New media - Terminology & Definitions

register

cyberspeak

new media

digitally mediated communication (DMC)

electronically mediated communication (EMC) style mass media netlish

multimedia weblish   text type

computer-mediated communication (CMC)

netspeak   medium social media   output electronic media

genre e-language

electronic discourse

3. New media - Terminology & Definitions

Lievrouw & Livingstone (2006: 2) on **new media**

"[We defined them as:] the *artefacts or devices* used to communicate or convey information; the *activities and practices* in which people engage to communicate or share information; and the *social arrangements or organizational forms* that develop around those devices and practices."

Baron (2011: 119) on **CMC**

"Some urged incorporating the study of mobile messaging under the rubric of computer-mediated communication, while others proposed alternative terminology. I favor the **umbrella term 'electronically mediated communication' (EMC) to encompass language used with any online or mobile device, though the term CMC is more broadly used**, at least in referring to language actually composed on a computer."

3. New media - Terminology & Definitions

Crystal (2011: 9-10) on **outputs**

"A rather different **terminological question** is what to call the various entities which form Internet discourse, such as email, blogs, chats, and tweets. A main aim of Internet linguistics is to establish their linguistic character. They are often described as *genres*, but that suggests a homogeneity which has not yet been established. The same question-begging would arise if they were called *varieties or dialects or registers* or any of the other terms for situationally related uses of language provided by sociolinguistics and stylistics. Linguists have to demonstrate linguistic coherence, not assume it. We need a term that is **theoretically neutral, from the linguistic point of view, and for the present book I propose to use outputs**. [...] The term implies nothing about its linguistic character, or how it relates to other outputs."

Overview

1. Introduction

2. Dutch

3. New media

4. Lexicon: A profile-based method

5. Register: A multi-dimensional approach

6. Conclusions

4. Lexicon: A profile-based method

Why the lexicon?

- QLVL tradition
- Scope
- Written text corpora

4. Lexicon: A profile-based method

How?

- Profile-based method introduced and elaborated in Geeraerts, Grondelaers & Speelman (1999) and Speelman, Grondelaers & Geeraerts (2003)
- ⇒ Some terminology
    - Lexical variation
    - Onomasiology
    - Profile
    - City-block distance

4. Lexicon: A profile-based method

MOBIELE TELEFOON	Bel.Dutch	%	Neth.Duth	%
cellulaire telefoon	2	0.000	1	0.000
draagbare telefoon	174	0.011	66	0.004
gsm	11443	0.754	1108	0.063
mobiel	19	0.001	10859	0.619
mobiele telefoon	2678	0.176	3930	0.224
mobieltje	619	0.041	1209	0.069
zaktelefoon	249	0.016	375	0.021

4. Lexicon: A profile-based method

How?

1. Calculate the relative frequencies for each alternative
2. Subtract the relative frequencies two by two
   - 0.000 − 0.000 = 0.000; 0.011 − 0.004 = 0.008; 0.754 − 0.063 = 0.690; 0.001 − 0.619 = -0.618; 0.176 − 0.224 = -0.048; 0.041 − 0.069 = -0.028; 0.016 − 0.021 = -0.005
3. Add the absolute values of the subtractions
   - |0.000|+ |0.008|+ |0.690|+ |-0.618|+ |-0.048|+ |-0.028|+ |-0.005|= 1.397
4. Divide by two
   - 1.397 ÷ 2 = 0.698

4. Lexicon: A profile-based method

How?

1. Calculate the relative frequencies for each alternative
2. Subtract the relative frequencies two by two
   - $0.000 - 0.000 = 0.000$; $0.011 - 0.004 = 0.008$; $0.754 - 0.063 = 0.690$; $0.001 - 0.619 = -0.618$; $0.176 - 0.224 = -0.048$; $0.041 - 0.069 = -0.028$; $0.016 - 0.021 = -0.005$
3. Add the absolute values of the subtractions
   - $|0.000| + |0.008| + |0.690| + |-0.618| + |-0.048| + |-0.028| + |-0.005| = 1.397$
4. Divide by two
   - $1.397 \div 2 = \boxed{0.698} \Leftarrow$ dissimilarity measure

Overview

5. Register: A multi-dimensional approach

What is register?

- $\neq$ genre (Biber 1994: 51)
- $\neq$ text type (Biber 1994: 52)
- "a cover term for any language variety defined in situational terms, including the speaker's purpose in communication, the topic, the relationship between speaker and hearer, spoken or written mode, and the production circumstances" (Biber et al. 2002: 10)

5. Register: A multi-dimensional approach

What is register?

- ≠ genre (Biber 1994: 51)
- ≠ text type (Biber 1994: 52)
- "a cover term for any language variety defined in situational terms, including the speaker's purpose in communication, the topic, the relationship between speaker and hearer, spoken or written mode, and the production circumstances" (Biber et al. 2002: 10)

Who is Biber?

- *Variation across speech and writing* (1988)

5. Register: A multi-dimensional approach

What is multi-dimensional analysis (MD)?

- MD groups co-occurring patterns among linguistic features into factors, which are then interpreted as dimensions of functional variation

How to apply multi-dimensional analysis?

1. Relevant linguistic features ⇒ matrix
2. Factor analysis ⇒ expose correlations
3. Shared situational, social and cognitive functions
   ⇒ underlying dimensions

5. Register: A multi-dimensional approach

	nomz	**PrepP**	**fpp**	**spp**	**absN**	$\cdots$	**netsp**
text$_1$	0.743	0.541	0.825	0.287	0.196	$\cdots$	0.987
text$_2$	0.831	0.697	0.032	0.221	0.868	$\cdots$	0.471
text$_3$	0.977	0.751	0.194	0.212	0.980	$\cdots$	0.002
text$_4$	0.321	0.554	0.725	0.836	0.642	$\cdots$	0.312
text$_5$	0.759	0.709	0.209	0.126	0.794	$\cdots$	0.082
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
text$_n$	0.897	0.795	0.329	0.304	0.921	$\cdots$	0.381

5. Register: A multi-dimensional approach

	nomz	PrepP	fpp	spp	absN	$\cdots$	netsp
text$_1$	0.743	0.541	0.825	0.287	0.196	$\cdots$	0.987
text$_2$	0.831	0.697	0.032	0.221	0.868	$\cdots$	0.471
text$_3$	0.977	0.751	0.194	0.212	0.980	$\cdots$	0.002
text$_4$	0.321	0.554	0.725	0.836	0.642	$\cdots$	0.312
text$_5$	0.759	0.709	0.209	0.126	0.794	$\cdots$	0.082
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
text$_n$	0.897	0.795	0.329	0.304	0.921	$\cdots$	0.381

5. Register: A multi-dimensional approach

	nomz	PrepP	fpp	spp	absN	$\cdots$	netsp
$\textbf{text}_1$	0.743	0.541	0.825	0.287	0.196	$\cdots$	0.987
$\textbf{text}_2$	0.831	0.697	0.032	0.221	0.868	$\cdots$	0.471
$\textbf{text}_3$	0.977	0.751	0.194	0.212	0.980	$\cdots$	0.002
$\textbf{text}_4$	0.321	0.554	0.725	0.836	0.642	$\cdots$	0.312
$\textbf{text}_5$	0.759	0.709	0.209	0.126	0.794	$\cdots$	0.082
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\textbf{text}_n$	0.897	0.795	0.329	0.304	0.921	$\cdots$	0.381

$\Rightarrow$ indication of an information-focused dimension

5. Register: A multi-dimensional approach

Why use MD?

- Univocal ground: 1 method $\rightarrow$ various registers
- Compare among new media registers
- Compare to other more traditional registers
- Allows to incorporate profile-based lexical variation

5. When register meets lexical variation...

How?

- Combine register & lexicon methodology
  - Incorporate MD (frequency $\rightarrow$ distance) in lexical variation
  - Incorporate lexical variation (distance $\rightarrow$ proportion) in MD
  - $\Rightarrow$ Evaluation without & with

Why?

- Tool to investigate functions expressed by new media registers & deal with lexical variation

5. When register meets lexical variation...
or when one plus one equals three

Contributing to...

- Lexical variation
    - Tradition: new media as representations of substandard language use
    - Extension: grasp to what extent functional dimensions influence lexical variation (and is there a difference for Belgian Dutch and Netherlandic Dutch)

- Register/multi-dimensional analysis
    - Tradition: limited set of lexical variables in research by Biber
    - Extension: grasp to what extent systematic (profile-based) lexical variation research influences the determination of functional dimensions

Overview

6. Conclusions

Return to RQs

- What does language use in new media in Dutch show us about how and for which purposes the new media are used?
  - ⇒ Register
- How informal is the language use in new media registers in Dutch?
  - ⇒ Lexicon
- How does language use in new media in Dutch (both in terms of functionality as in lexical richness) differs from more traditional registers, and how do the new media registers vary among themselves?
  - ⇒ Register & lexicon

6. Conclusions

Return to RQs

- Descriptive
  - Creating a global overview of (the functions expressed by / the lexical diversity in) new media in Dutch
    - ⇒ Register & lexicon
- Methodological
  - Bringing together profile-based lexical variation and multi-dimensional register research
    - ⇒ Register & lexicon

6. Conclusions

What should you retain?

- Register
    - Role: multi-dimensional analysis
    - New: influence of functional dimensions on lexical variation

- Lexicon
    - Role: profile-based study of lexical variation
    - New: influence of lexical variation on functional dimensions

- New media
    - Role: data
    - New: status for Dutch

- Dutch
    - Role: tradition
    - New: systematic global comparison (and extrapolation?)

Overview

Thank you! Suggestions? Questions?

More information?
http://wwwling.arts.kuleuven.be/qlvl
jocelyne.daems@arts.kuleuven.be

References

- Baron, Naomi S. 2011. "Assessing the Internets Impact on Language." In: Mia Consalvo & Charles Ess (eds). *The Handbook of Internet Studies.* Malden: Wiley-Blackwell, p. 83-115.

- Biber, Douglas. 1988. *Variation across Speech and Writing.* Cambridge: Cambridge University Press.

- Biber, Douglas. 1994. "An Analytical Framework for Register Studies." In: Douglas Biber & Edward Finegan (eds). *Sociolinguistic perspectives on register.* New York: Oxford University Press, p.31-56.

- Biber, Douglas, et al. 2002. "Speaking and writing in the university: A multi-dimensional comparison." *TESOL Quarterly* 36 (1): 9-48.

- Clyne, Michael. 1992. *Pluricentric languages: differing norms in different nations.* Berlin/New York: Mouton de Gruyter.

- Crystal, David. 2011. *Internet linguistics: A student guide.* Abingdon: Routledge.

- Geeraerts, Dirk, Stefan Grondelaers & Dirk Speelman. 1999. *Convergentie en divergentie in de Nederlandse woordenschat: een onderzoek naar kleding- en voetbaltermen.* Amsterdam: P.J. Meertens-Instituut.

- Lievrouw, Leah A. & Sonia Livingstone. 2006. *The Handbook of New Media: Updated Student Edition.* London: Sage Publications Ltd.

- Speelman, Dirk, Stefan Grondelaers & Dirk Geeraerts. 2003. "Profile-based linguistic uniformity as a generic method for comparing language varieties." *Computers and the Humanities* 37 (3): 317-337.