

Extending SLICE. From top-down to bottom-up, from survey to social media corpora

Stefan Grondelaers

(with Dirk Speelman, Roeland van Hout, Stefania Marzo,
Chloé Lybaert & Paul van Gent)



Introduction

- Experimental SLICE-research has unveiled some of the main determinants of **fluidity and change** in European standard languages
- In communities which have preserved their standard language ideal
 - new **best languages** have appeared in the form of regional and **modern** standards (or neo-standards, Auer 2017 & 2018).
 - the emergence of neo-standards (“Low Copenhagen” speech in Denmark, Tussentaal in Flanders, Estuary English in the UK,...) is motivated by **dynamism** (modern prestige) which pertains to hipster coolness, media slickness & “street sophisticated toughness” (Trudgill 1986: 53)



Crucially

- Prestige is **not an intrinsic feature** of variants & varieties, but the outcome of competing language ideologies which hierarchize the different varieties spoken in a community in alternative ways:
 - conservative standard language ideology ranks varieties in terms of **traditional superiority** features (high birth, good education, fine manners, nice income)
 - most prestigious in that value system = **standard** variety
 - modern counter-ideology ranks varieties in terms of dynamism/modern prestige
 - most prestigious in that value systems = neo-standards, but also **multi-ethnolects** like Dutch Straattaal, Flemish Citétaal, German Kiezdeutsch



Concern

- **Almost no collaboration** within the SLICE-framework between researchers in the **experimental** and the **media strand**
- “It is widely accepted that broadcast mass media, meaning television and radio, have historically played important roles in the linguistic and ideological processes we refer to as 'standardisation' (SLICE-website)
- Logical extension to **social media**, which cancel sharp distinction between speaking and writing and “pluralise language norms” (Androutsopoulos 2011)



The methodological divide...

- Experimental SLICE-work is based on Kristiansen's (2009) speaker evaluation (matched-guise) **experiments**:
 - respondents evaluate audio clips representing different varieties on measures of speaker **personality** (the speaker is *nice, intelligent, cool,...*)
 - care is taken to **hide** the experimental purpose from the respondents in order to harvest **covert** evaluations they cannot actively or consciously manipulate
 - replicated in Flanders, The Netherlands, Lithuania, Ireland, Germany, Bosnia,...
- Work in the media strand is predominantly (though not exclusively) **qualitative**
 - social meanings are not stable, monolithic constructs which can be elicited as such
 - social meaning is eminently local, it emerges in interaction, and it has **no cross-contextual relevance** (so please stop **counting** it...)



Why we have to bridge the divide in the next SLICE-decade

- We need to **elaborate** our study of dynamism / modern prestige
 - the prestige of the standard is **easy to elicit**
 - but **dynamism** of neo-standards is **much more difficult** to extract – arguable more difficult to find **good scales**
 - do we – experimental SLICErs – treat dynamism too much as **stable monolithic social meaning**? Is there only one sort of dynamism?
- **Causal** relation between evaluation and production: does positive evaluation really cause / motivate / facilitate change?
 - scholars who explicitly argue for a causal relation **almost never have the evidence** to back up motivation, let alone causation
 - impossible in an experiment – you have to go back to production data
- In an experiment, it is difficult to test the **impact of mass media** on the acceleration of a change (still: *Eastenders* is a significant correlate of the acceleration of two Scottish innovations – Stuart Smith 2013)



Why go beyond the matched-guise technique? Recurrent criticism on matched guise technique

- Choice of measuring scales lacks “empirical motivation and construct validity” (Lee, 1971:413)
- Researchers typically copy these scales from preceding studies, at the risk of “circularity (. . .) and a deceptive semblance of exhaustiveness” (Garrett, 2005:1256)
- Not a good idea since an evaluation dimension like dynamism comes in many shapes: just copying Tore’s scales may not be the valid option...
- Typically include varieties which are known beforehand, or highly plausible to elicit specific evaluations, as a result of which new or emergent prestige varieties may stay under the radar



Study 1. Making perceptual maps with an enriched free response technique

- Can we uncover ideologies and their prestige reference points with a technique which does **not rely on scaled measures** and a **limited gamut** of testable varieties?
- Does this technique allow us to go “**as deep**” as the matched-guise technique?
 - some consensus that it takes **indirect techniques**, which keep participants ignorant of the linguistic purpose of an experiment, to access deep evaluations they cannot consciously manipulate
 - claim implied in Kristiansen (2009), but there is some counter-evidence (Rosseel 2017)
- Case study: Belgian Dutch

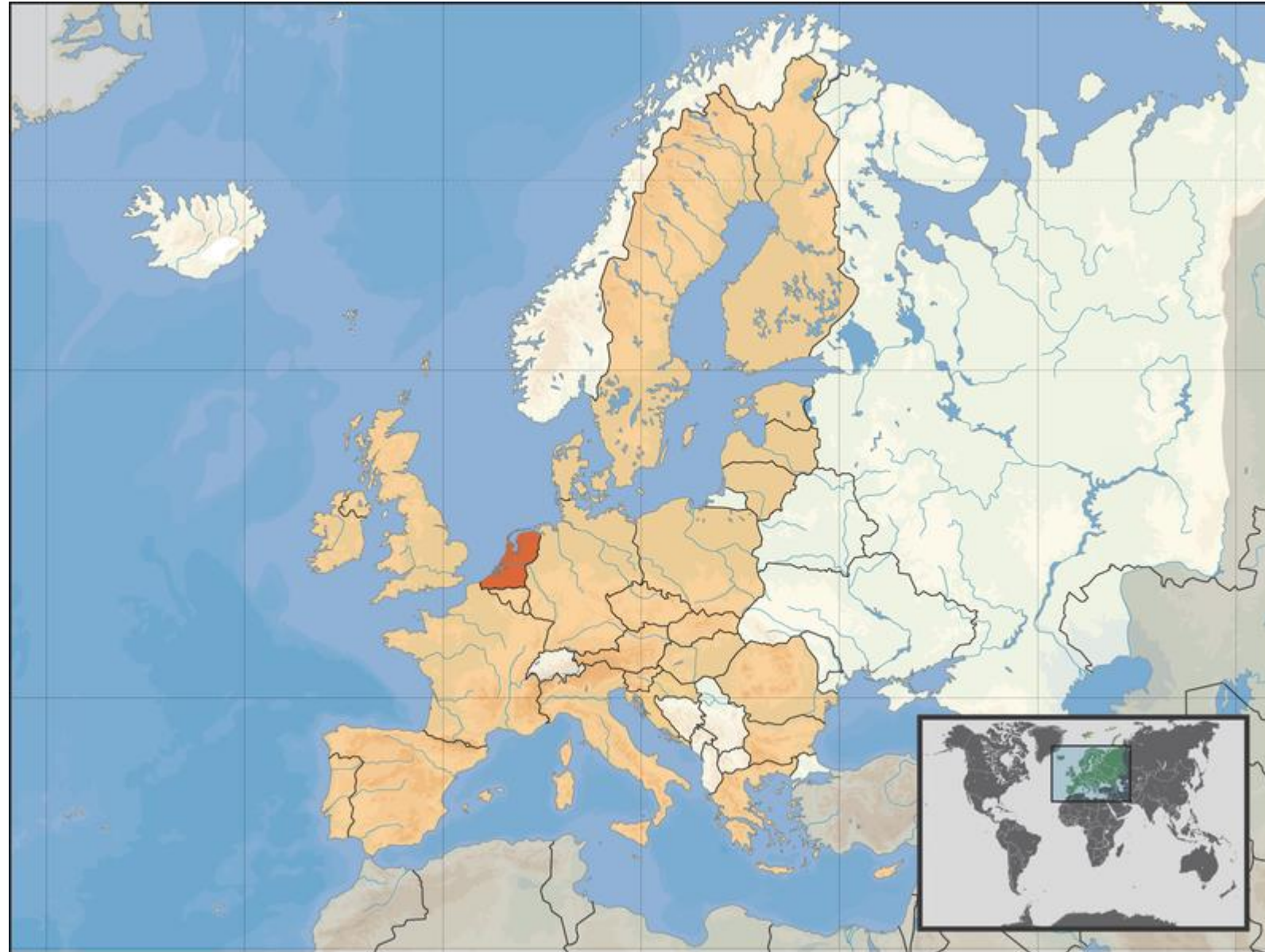


Dutch

- Dutch is a comparatively **small** language, spoken in a comparatively **small** area of Europe
- **Pluricentric** with 2 national varieties in Europe
 - Netherlandic Dutch (as spoken in The Netherlands)
 - Belgian Dutch (as spoken in Flanders, the northern half of Belgium)
 - (Dutch is also the official language of Suriname)
- Belgian and Netherlandic Dutch are **worlds apart**
 - ... although they started diverging only in the 16th century...



Dutch in Europe



Dutch in The Netherlands & Belgium



Synchronically

- In The Netherlands, the 16th century desire for a common best language materialized in a **flexible standard** variety which can stretch over registers and regions without tearing ...
- In Flanders, the desire for a common best language **did not materialize** in a consensus variety. What the Flemish got was
 - An unattainable spoken standard which is **widely cherished** but largely **virtual** because almost nobody speaks it (official norm Dutch of VRT, “Flemish Radio and Television”)
 - A **hysterical rejection** and **stigmatization** of the vital varieties, notably the colloquial informal variety called “**Tussentaal**”
 - **Hypersensitivity** to spelling and pronunciation errors (“language nazi’s”)
 - Extremely **conservative** standard language ideology

Historically

- Netherlands: textbook case of **normal, uninterrupted implementation** of standard language desire in terms of Haugen's 4 stages...
- Flanders:
 - **Problematic standardization**: political isolation from The Netherlands in 16th century and almost exclusive **reliance on French** until 1900;
 - **Forced, unspontaneous restandardization** from 1900
 - VRT-Dutch – a Flemish-flavored variety of Netherlandic Dutch – was **imposed** on the Flemish through **hyperstandardization** (Van Hoof and Jaspers 2012: 97), “a propagandistic, large-scale and highly mediatized linguistic standardization campaign”



Conservative standard language ideology, and how to “activate” it...

- In February 2019, I wrote an opinion piece in the leading Flemish newspaper in which I argued that the VRT might consider **relaxing its language policy** and officially embrace **some variation in some contexts**
- In the following days
 - suite of **fuming opinion pieces** by culturally prominent Flemings who denounced me for my “attack” on Standard Dutch...
 - hundreds of angry Facebook reactions ...
- Political reactions against my opinion piece
 - interview with Flemish prime minister: “standard language provides equal chances for everyone. That’s why we should continue to teach it and use it as a norm on television”
 - **parliamentary question** by the Flemish nationalist party NVA to the minister of culture: “what further actions is the VRT planning with respect to a revision of its language policy?”



Crucial

- “Who is going to determine [which variants will be admitted]? The faith healers of Dutch universities? Lord, deliver us from faith healers like Grondelaers. They know not what they do.”
- “In its opinion pages, De Standaard of 22 February 2019 gives room to one Stefan Grondelaers, who is connected to the Nijmegen Radboud University in an unnamed position.”
- “A pseudo-active guy who would otherwise end up in the unemployment statistics”
- Representative of a “humanistic field that should not be taught at university”.
- Gal & Irvine (1995: 975 ff.) on [erasure](#): “because a linguistic ideology is a totalizing vision, elements that do not fit its interpretive structure – that cannot be seen to fit – must either be ignored or be transformed”.



Crucial question

- Is this conservative ideology the **only value system** which impacts Flemish standard language dynamics?
- What happens if we **dig deeper**?
- Do we find ideologies which motor **change in progress**?
- Do these ideologies, if we find them, frame new types of prestige which motivate the emergence of **competitors** to VRT-Dutch?
- How do we pinpoint them & render them **visible**?



Free response experiment (2017)

- Task: name first 3 adjectives which spring to mind when confronted with the labels for a number of Belgian language varieties
 - “Dutch as spoken on the official broadcaster VRT”
 - “Dutch as spoken in soap series like *Thuis* or *Familie*” (for want of a better label for Tussentaal with a higher lay relevance)
 - “Dutch with a Ghent/Antwerp/Limburg/West-Flemish/Moroccan accent”
- 211 highly educated participants (119 f, 91 m; 132 younger, 79 older), sample controlled for province of origin/residence
- **Time pressure** implemented to guarantee prompt extraction
- 4,242 tokens for 1,013 adjective types
- For **83.3 %** of the non-hapaxes, **valence** information available (how positive/pleasant is x) from Moors et al. 2013 and Warriner et al. 2013).

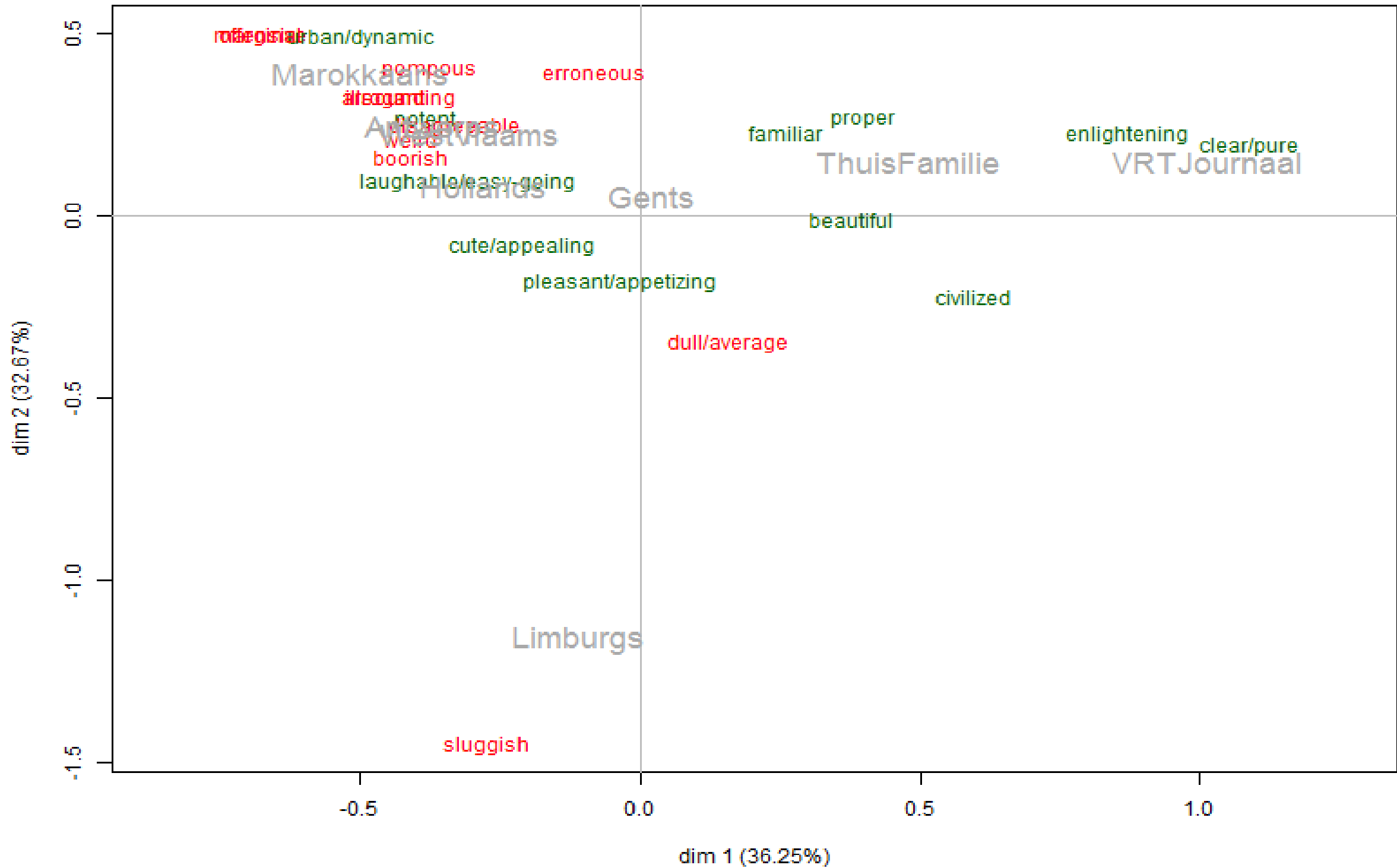


Computational processing

- From Leuven News Corpus, we extracted the 5000 most frequent content words
- For each occurrence of the adjectives returned in the experiment, we calculated the frequency of the top-5000 words in a window delimited by the four words preceding, and the four words following the adjective, resulting in a **row vector** of frequencies for 5000 alphabetically listed words
- Frequencies were transformed into pointwise mutual information scores, and the resulting row vector of 5000 pointwise mutual information scores subsequently represented the **semantic passport** of each adjective.
- *k*-medoids **cluster analysis** to extract clusters of semantically related adjectives.
- Correspondence Analysis in bi-plots (7 labels & 11 pos & 11 neg clusters)
- Hierarchical cluster analysis (pvclust) based on co-ordinates from CA



all participants (two valence categories)



Some findings

- X-axis: **norm dimension** (Moroccan on the left, VRT-Dutch on the right)
 - Reproduces **conservative** ideology: association VRT-Dutch with **ideology-infused** dimensions ENLIGHTENING & CLEAR/PURE
 - No categorical dichotomy between 1 good (VRT-Dutch) & many bad varieties (experiment **does not reproduce most conservative ideology**)
 - Ghent accent & Tussentaal occupy middle position
 - Tussentaal **not rejected**, associated with standard qualities BEAUTIFUL & CIVILIZED
 - Ghent accent associated with positive **dynamic qualities** CUTE/APPEALING & PLEASANT/APPETIZING: **new finding** but converges with other evidence that Ghent accent is hip



Dynamism revisited

- At least 2 types of dynamism evaluations
- Y-axis **energy** dimension:
 - Limburgs = non-energetic (SLUGGISH)
 - Moroccan = over-energetic (OFFENSIVE & MARGINAL)
- Y-axis stratifies dynamism attributes in
 - **negative**: OFFENSIVE, MARGINAL, POMPOUS, ARROGANT
 - **positive**: LAUGHABLE/EASY-GOING, CUTE/APPEALING, PLEASANT/APPETIZING
- Both positive and negative dynamism found to **correlate with diffusion**
 - positive dynamism resembles Kristiansen's (2009) dynamism concept & original scales: typically correlates with neo-standards (“**hipster**” dynamism; “**yuppie** dynamism”)
 - negative dynamism has been found to correlate with multi-ethnolects (Moroccan accent; Citétaal) – “**badass** cool” & “**macho** prestige”

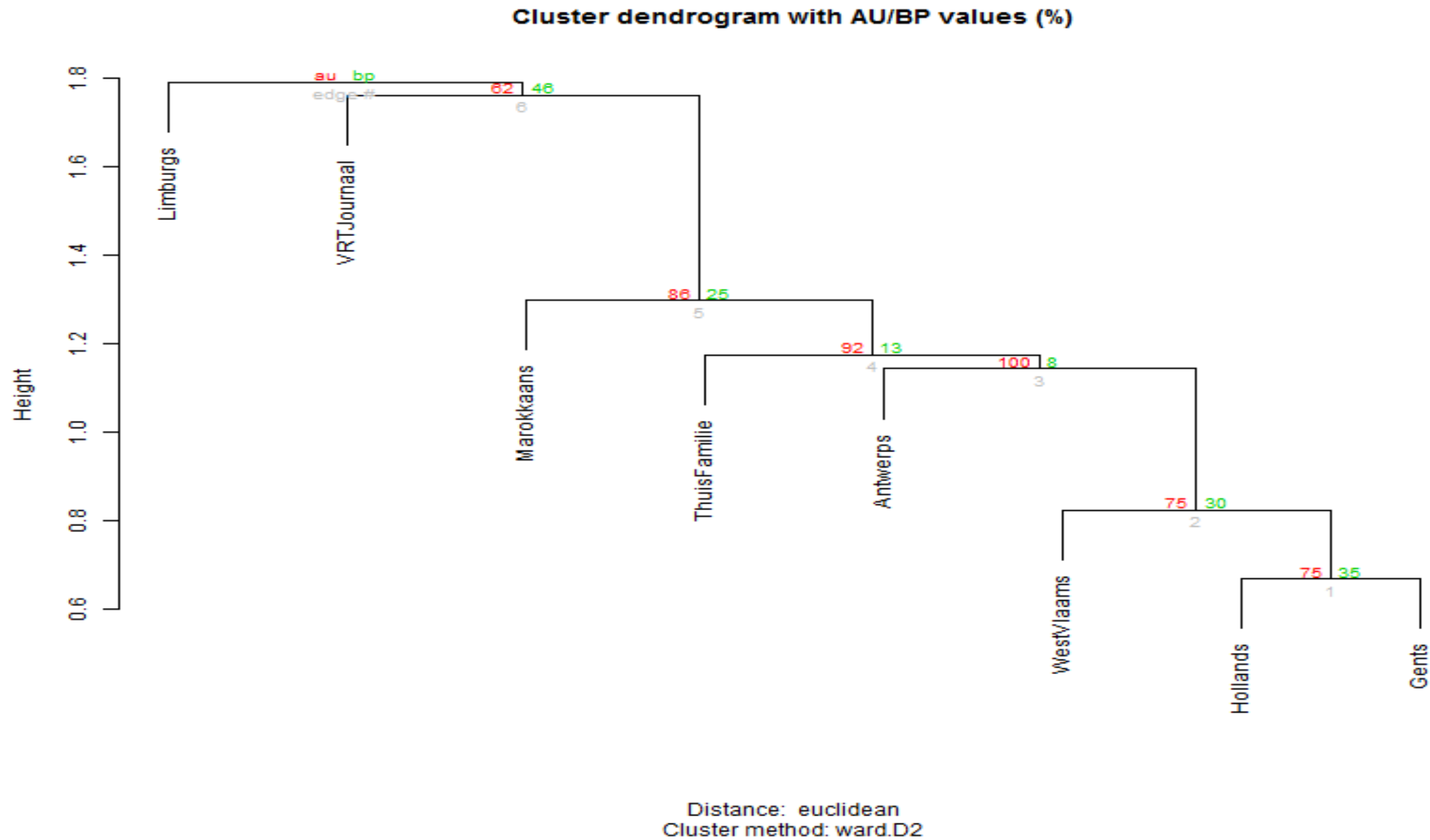


Most important...

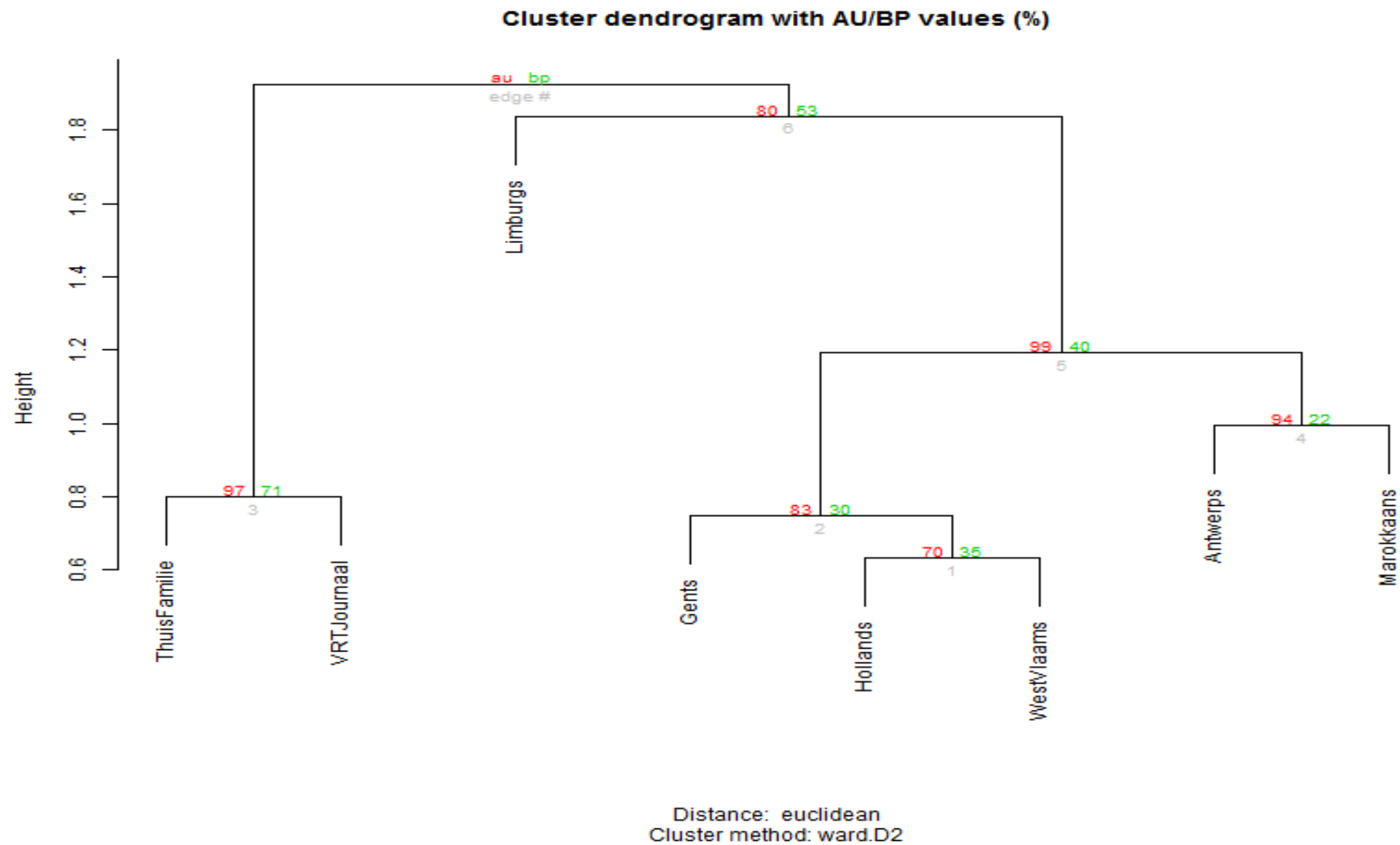
- Very much a **system in motion**:
 - Change inherent in different prestige types from left to right on norm dimension:
 - On left, **unprestigious** varieties
 - Ghent has dynamic prestige which is typical correlate of **emergent standards** (such as Modern Copenhagen speech)
 - Tussentaal has the **traditional** prestige correlates of a **practical** standard
 - VRT-Dutch has the **ideological** correlates of the **ideal/virtual** standard
 - Inter-generational change: obvious **change in conceptualizations** of the younger respondents



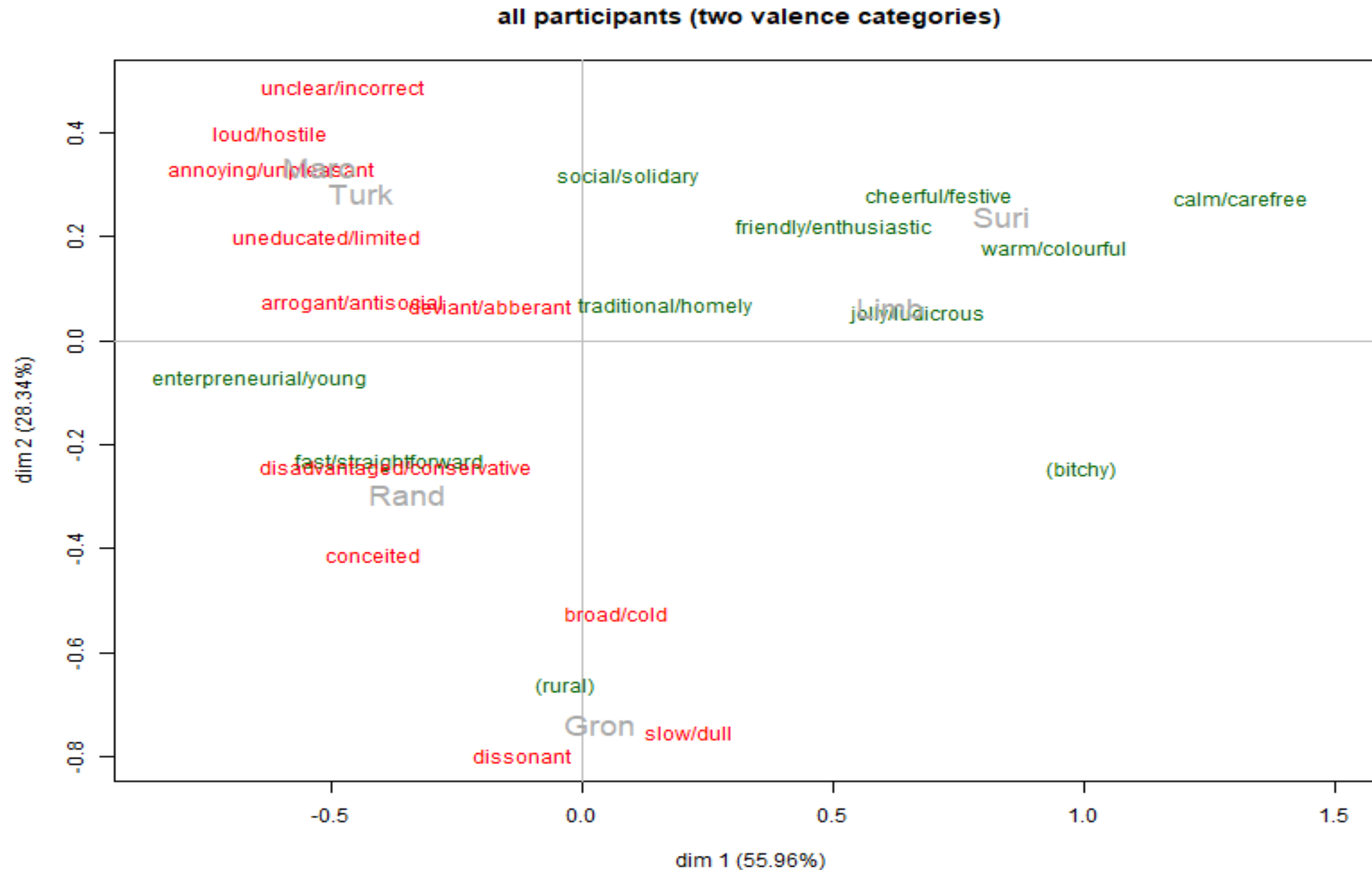
Cluster dendrogram for older participants



Cluster dendrogram for younger participants



(Situation in The Netherlands)



Study 2. Social meaning as a driving force on Twitter

- What is the **exact role** of language evaluation in language change processes?
- Is **positive evaluation**, or rather, a specific type of positive evaluation, a reason why rapid linguistic diffusion takes place
 - *be like* across world Englishes (D'Arcy; Tagliamonte)
 - Cockney phonetic innovations in Glasgow English (Stuart-Smith 2013)
 - diffusion of multi-ethnolectal features in Belgian Dutch
- What is the role of the **mass media** in this process?
 - watching *Eastenders* is a **significant co-determinant** of the diffusion of Cockney features in Scotland



Stuart-Smith (2013): eye-openers

- Specifically, diffusing features are often linked with **very locally defined social meanings** that in turn relate to **broader supralocal language ideologies** (Eckert 2000 & 2008) (p. 502)
 - Local: “fronting and L-vocalization have become closely linked with the development of **local Glaswegian street style**” (p. 527)
 - Supralocal ideologies: “**not posh**” - ... “Trudgill (1986,1988) also suggested social meanings linked with **urban toughness**, and Williams and Kerswill's (1999) discussion of '**youth norms**' adds a further dimension. Wherever these features are proliferating in the UK, they **are linked with a bundle of overlapping social meanings of being 'us', 'not posh', '(cool) youth', and 'urban tough'.**”
- **No effect of positive evaluations** of the Cockney accent, but failure to find those may have been an artefact of the methodology (direct elicitation; no dynamism scales, p. 528)

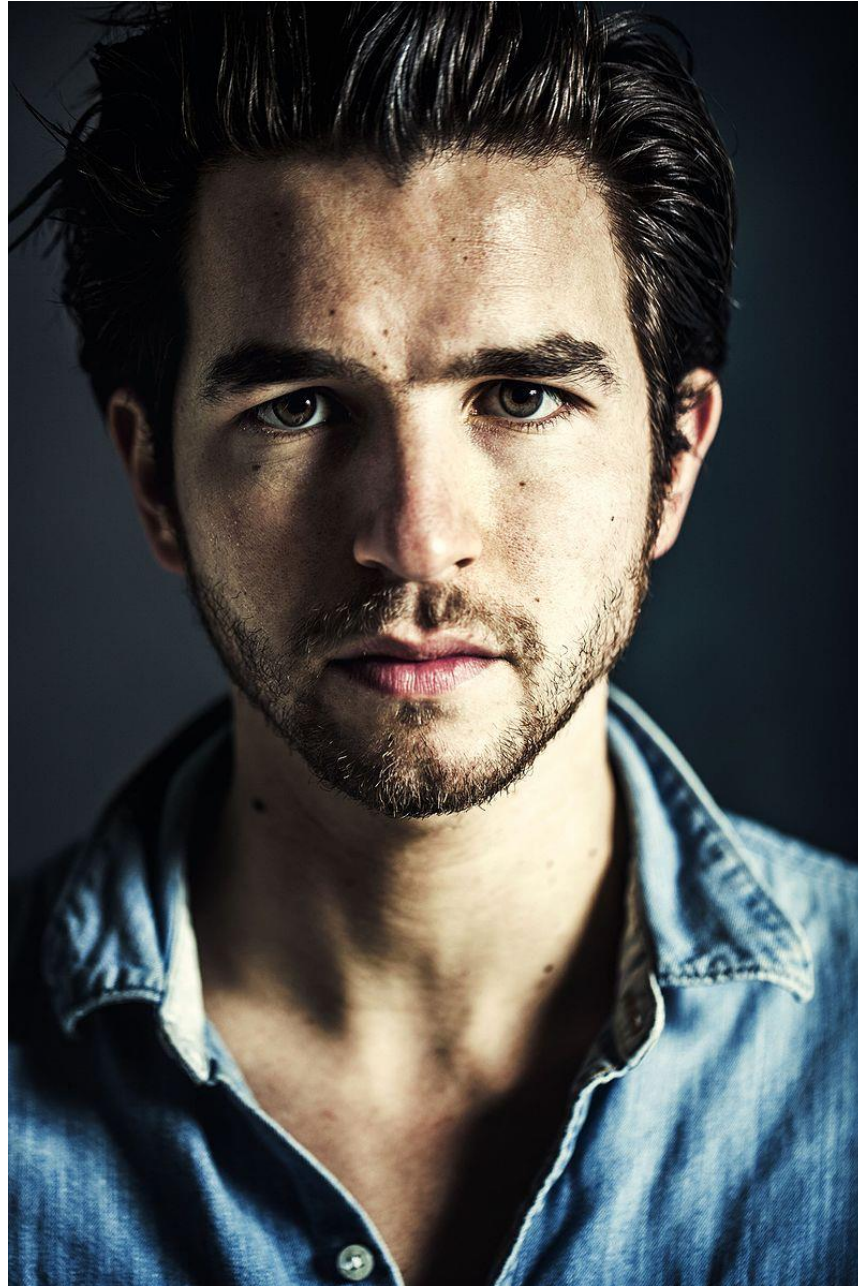


The case of *shtijl*

- Most remarkable rapid linguistic diffusion in Flanders
- Why do a growing number of Flemish adolescents (especially girls) **palatalize** the /s/ of *stijl* “style” to [ʃ]?
- Pronunciation variant that was born in the **Cité**, the multi-ethnic housing estates of miners in Genk (Belgian Limburg)
- Why do pronunciation features travel so **far beyond their original habitat**? Which factors determine their vitality?
 - Positive evaluation? Macho prestige (cool youth & urban tough)?
 - Mass media impact?



Most plausible explanation



- Youthful prestige of a **macho**, **cool** and **streetwise** personality
- **Signature** feature of Cité-speakers (especially of Italian descent)
- Boosted by **television success** of Matteo Simoni, Belgian actor of Italian descent
- **Attractive to adopt** by adolescents
- By adopting the *sjtjil*-variant, the borrower hopes to **inherit** some of the original user's cool.

Media impact on acceleration of Citétaal

- Popular television series and movies featuring Matteo Simone with Citétaal features:
 - *Safety First*: 2013 & 2014; movie in 2015
 - *Marina*: movie on iconic singer-with-Italian-roots Rocco Granata in the movie *Marina*.
- Local authorities and advertisers exploited the growing popularity of Citétaal in political and commercial campaigns to evoke a **new local and urban identity**
- All this boosted the **enregisterment** of Citétaal –no longer used to express ethnicity and migration but an urban streetwise identity (Cornips et al. 2019).
 - converges with the dynamic social meaning (macho prestige) we believe to be responsible for the diffusion of Citétaal...

Does the *shtyle* spread on account of modern prestige or mediatization?

- Ideally, **correlational** study with
 - spontaneous production data by Flemish adolescents
 - evaluations of Citétaal & other varieties by same adolescents
 - data on their television viewing frequency, interaction with other media...
- Not feasible on a sufficiently large scale to study diffusion across Flanders
- Instead, we rely on Twitter
 - open access micro-blogging platform for messages up to 140 characters
 - between 1.5 and 3 % of tweets have **geo-positioning** data (small proportion but global availability is enormous)
 - hashtags as **content cues**



Advantages of Twitter

- All computer-mediated communication (Androutsopoulos 2011; Hilte et al. 2018) shares 2 essential features:
- **Conceptual orality**: netspeak's similarity to casual speech, and approximation of casual speech features
 - spoken syntax constructions (Ruppenhoffer 2018)
 - non-standard orthography which is the result of error, or **indexical resourcefulness**: Twitter shares with authentic speech the presence of phonetic, lexical, and morphosyntactic cues which **systematically reveal the regional descentance** of tweeters
- **Compensation**: strategies to compensate for the absence of suprasegmental facial and other expressive cues in normal speech
 - emoticons
 - Intensifiers, capitalization, expressive lengthening (*coollllllllll*) – **affective discourse strategies** (Bamann, Eisenstein and Schnoebelen 2014)



Disadvantages of Twitter

- Spoken discourse properties do not render Twitter purely “vernacular” in the Labovian definition of a style of “speaking” to which the speaker pays the least attention (Bohman 2016: 163)
- Space limitations (140 characters), the asynchronous design, and the eminently public character of most Twitter communication probably make tweeters **more conscious** of what they type than chatters in public or private forums
- Twitter ≠ proxy for casual speech

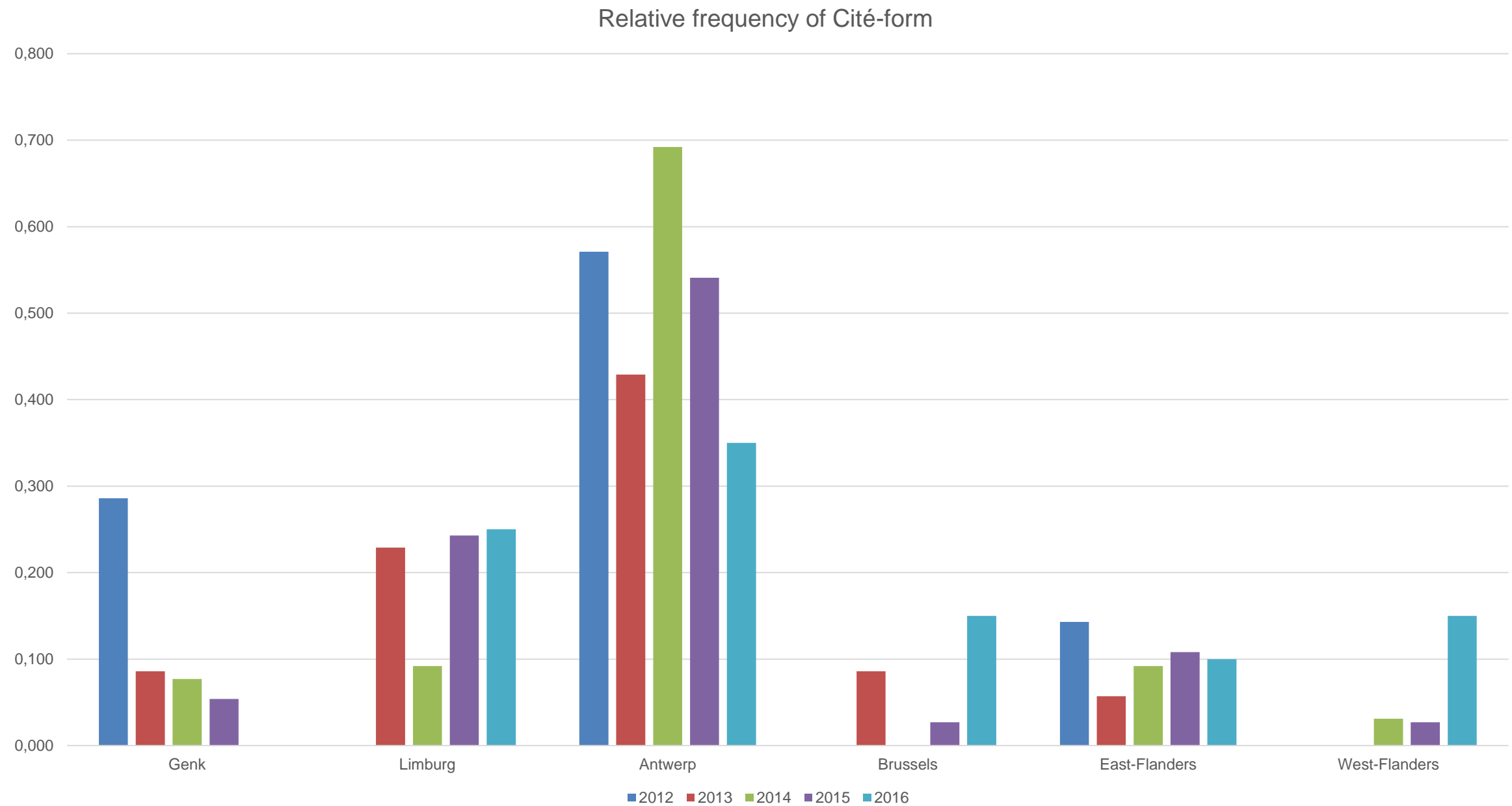


Design

- All tweets featuring the standard spelling *stijl* and its non-standard allographs (including *stijl*, *sjtijl*, *shtijl*, *shhhtijl*)
- 2012 to 2016
- Search restricted to tweets with [geo-coding](#) (latitude and longitude of the normalized location) and [user profile](#)
- Initial dataset n = 1411
- After removal of spurious hits (misspelled references to the homophonous *steil* “steep”, or references to *steil haar* “straight hair”), resulting dataset contains 1050 tweets with standard spelling *stijl*, and [180 \(14.83 %\) with non-standard palatalized spellings](#).

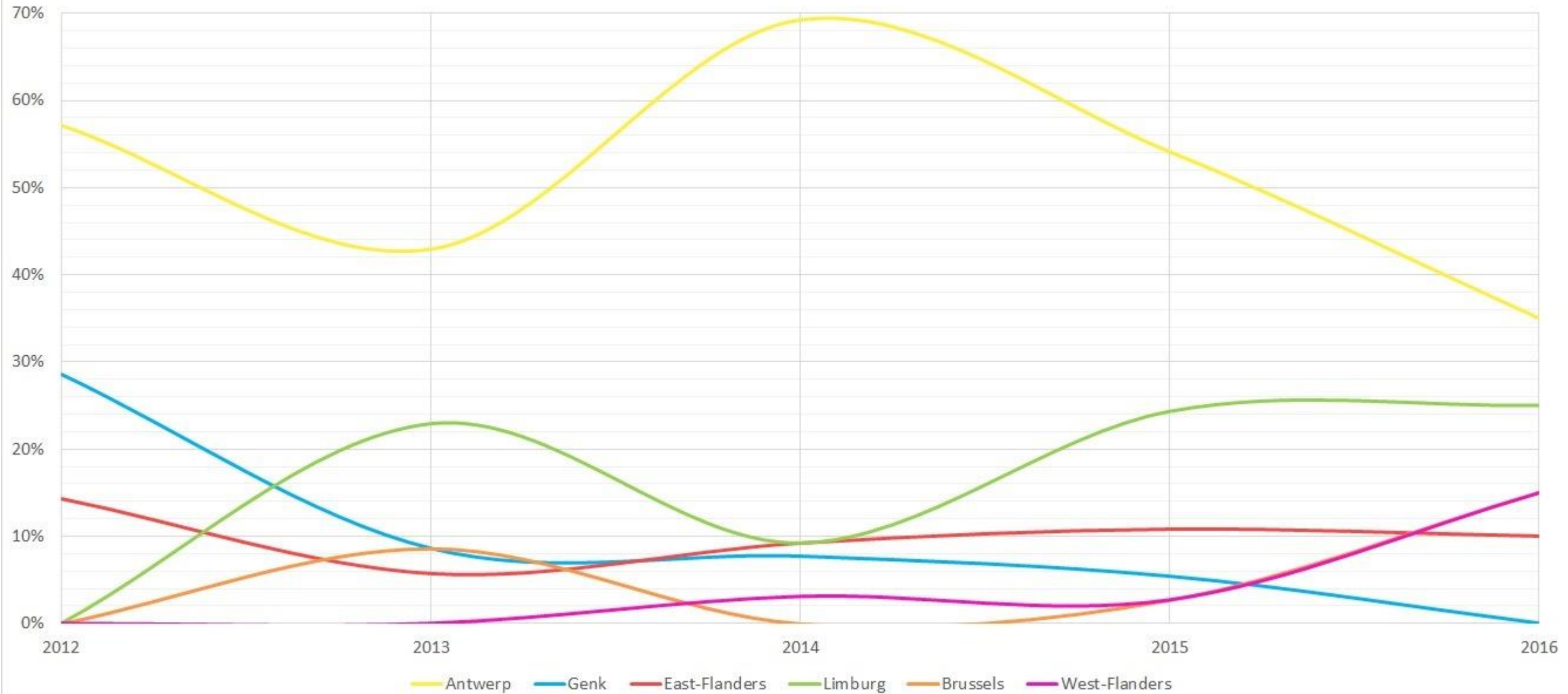


Diffusion through space and time



Diffusion through space and time

Relative frequency of Cité-form



Interpretation

- Mediatization does not seem to lead to nation-wide appropriation: no synchronic mediatization boost through television (as predicted in Sayers' 2014 *Mediated Innovation Model*).
- Data appear to be more compatible with a **gradual, estafette-like** change through social interaction.
- **Short-lived change**: expanding stain of the spread is already drying up in its core during the spread
- Little evidence for television-induced peak
 - popularity of *Safety First* peaked end 2013, Twitter-proportions of *sjtijl* peak in 2014, with a steady decline
 - only 1 explicit reference to *Safety First* in the hashtags, none to *#sjtijl*



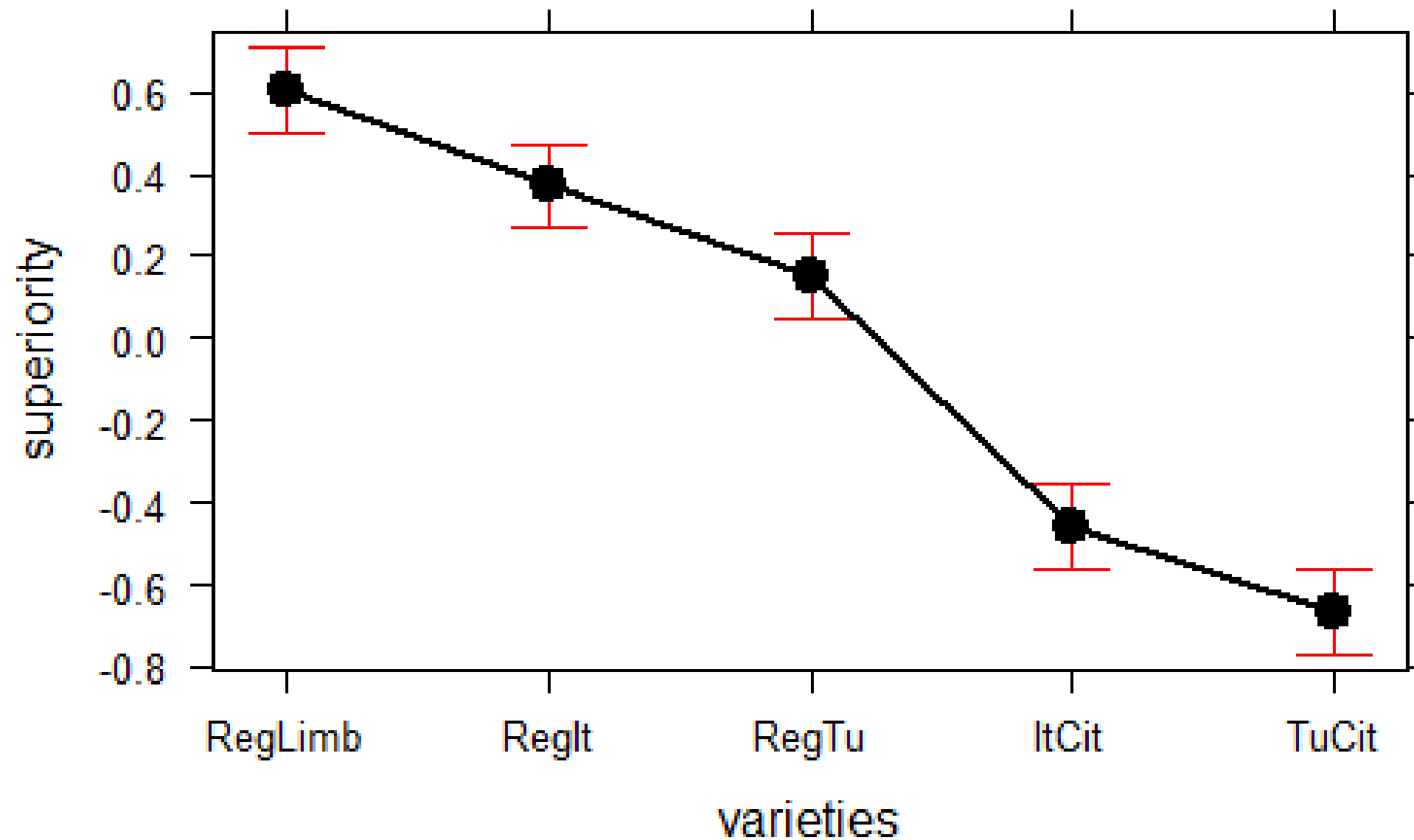
Does social meaning drive *shtyle*? Experimental data

- Stimuli
 - Regional Limburg: near-standard Dutch with an audible Limburg accent
 - Regional Italian: previous but produced by speaker of Italian descent
 - Regional Turkish: previous but produced by speaker of Turkish descent
 - Italian Cité: regional Italian guise with 4 iconic Citétaal features
 - Turkish Cité: regional Turkish guise with 4 iconic Citétaal features
- Scales. 9 dynamism scales included in function of
 - **hipster/yuppie interpretation** (Kristiansen 2009): this person is *confident, hip, cool, easy-going*
 - **macho interpretation**: *loud, macho, sturdy, sarcastic, provocative*
 - factor analysis only retains *cool, loud, sturdy, provocative* (**macho**)
- 161 young respondents (average age = 17.47), panel was regionally stratified

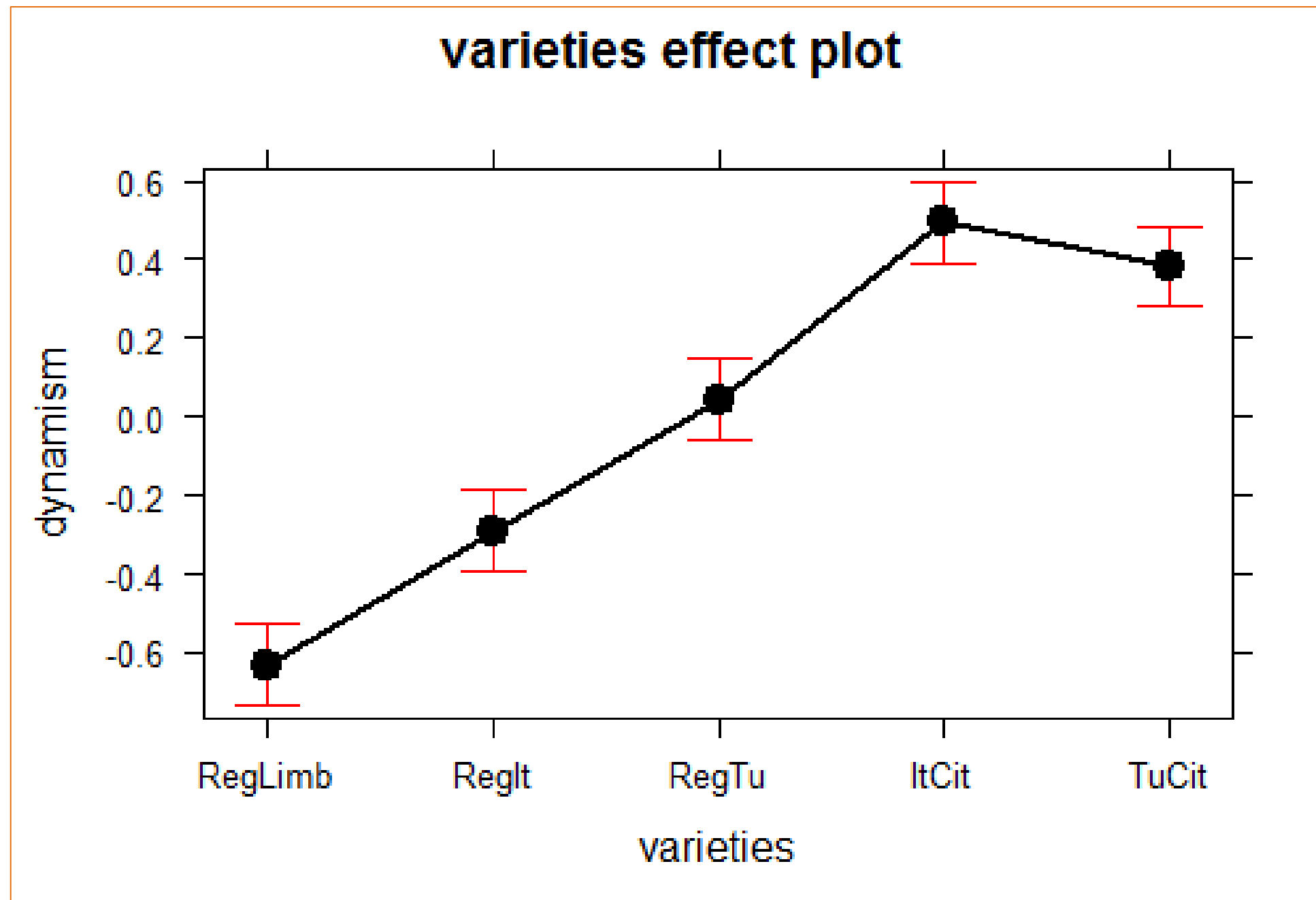


Findings for superiority

varieties effect plot

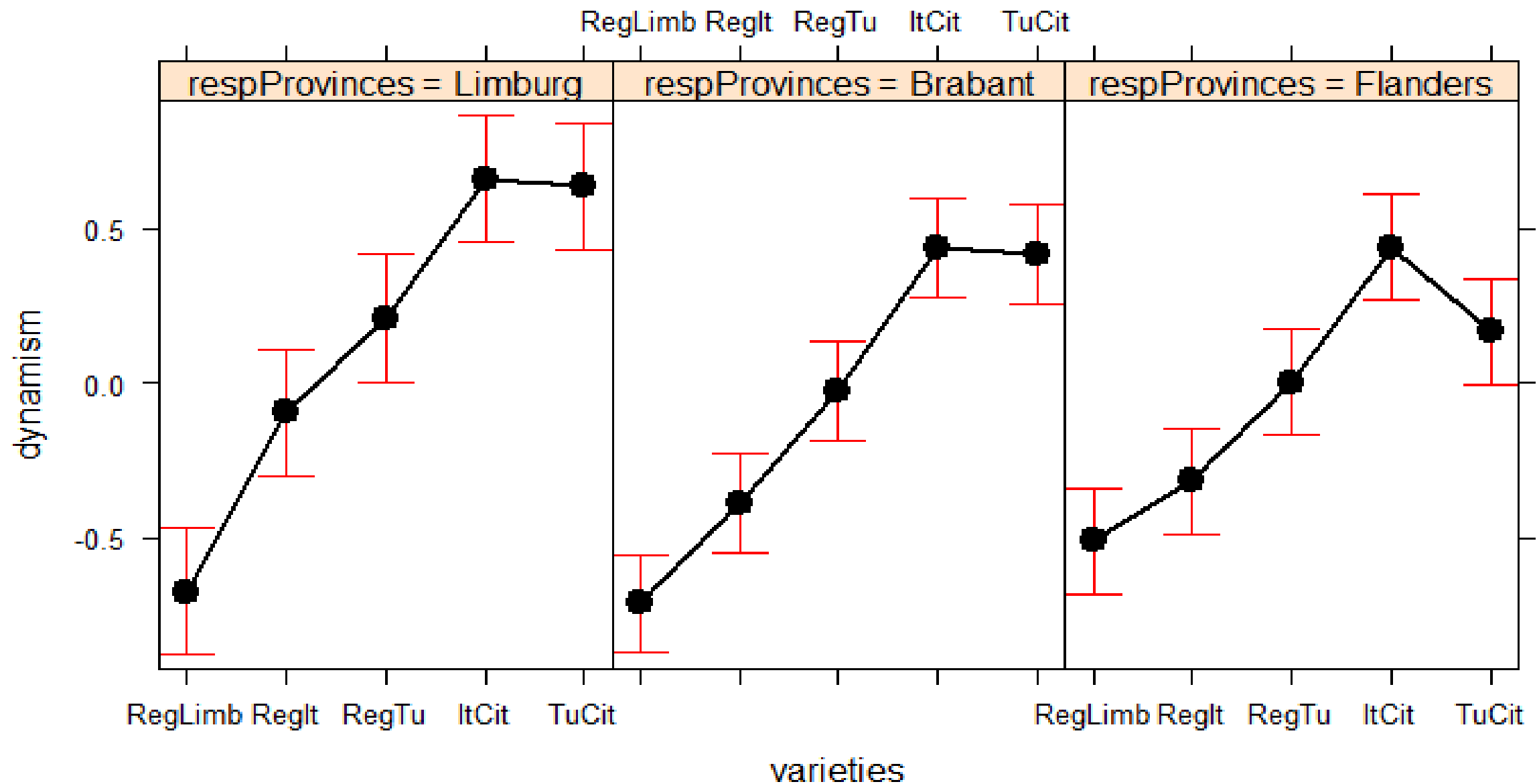


Findings for dynamism



Dynamism*Respondent Province

varieties*respProvinces effect plot



So

- On the superiority dimension, there is **extreme downgrading** for the Citétaal guises (no traditional prestige)
- Adolescents all across Flanders find the Cité guises – especially the Italian one – **much more dynamic** than the other guises
- Not hipster but **macho prestige**
- Stuart-Smith (2013): “specifically, diffusing features are often linked with **very locally** defined social meanings that in turn relate to broader **supralocal language ideologies**
 - very local social meaning: association with multi-ethnic community in Genk
 - association between Citétaal features and this community available to Flenish adolescents through **mediatization**
 - supralocal ideology: modern prestige of urban, young, cool, streetwise

Causality issue

- No better explanation for diffusion than media-boosted social meaning
- Still, “our what-comes-first problem [evaluation or change] will remain open, since **correlation is not explanation**” (Kristiansen & Jørgensen 2005: 295).
- What would constitute **valid evidence** that a language change is facilitated by a change in evaluation?
- Modern prestige is a **hearer** category, but speakers can try to **enact a modern prestige personality** (a process known as “stylisation”, cf. Rampton 2011).
- Do speakers who stylise themselves as somewhat provocative, streetwise, and cool dudes and gals manifest a **higher preference for the Cité-forms** in the enactment of this style?



Problem

- Bohmann (2016) twitter-based study of innovative *because*
- “Garber (...) describes the new construction as “**exceptionally bloggy and aggressively casual and implicitly ironic**” (2013). (p. 157)
- Carey (2013): “**snappy, jocular feel**”
- A possible explanation could be that what is meant here is not the casualness of unmonitored quotidian talk, but a studied, consciously constructed one that is exploited as a poetic device. **In other words, *because X* is perhaps not so much an indicator of a generally casual style as a resource that is exploited in the *stylization* of casualness (Coupland 2001). (p. 175)**
- But: “Since it is near impossible to model irony or humor quantitatively, the analysis presented here cannot address these aspects systematically“



Problem & possible solution

- Bohmann (2016) codes “stylized casualness” in terms of “colloquialization” – formality and variant spellings of *because*
 - doesn’t work well
- We code “stylized dynamism” in terms of implementations of the compensation principle
 1. Bent really has **THE** style that I love 😊 hmMMM #tvvv
 2. I know that, but it is really **shhhhtyle** eh?
 3. From style to swag. **wtf**
 4. We want to make style, we want to make style
 5. She has that drama shtyle ... but does she also have that ... ???
 6. A **dirty** (“awful”) lot of style, dude...



Some findings

- Random forest analysis to estimate impact of different predictors on *sjtjil*-use:
 - Most important factor is the **semantics** of *stijl*: when “this is style” means “this is cool”, it is palatalized in 92.11 % of all cases, vs. only 9.53 % of the non-cool/ meanings ($\chi^2 = 389.17$; 1 Df; $p < .0001$)
 - Stylised dynamism is the **second most important** linguistic factor: only 19.84 % of all tweets was coded as enacting dynamism, but 27.46 % of these were palatalised vs. only 11.46 % of the others ($\chi^2 = 40.08$; 1 Df; $p < .0001$)



Conclusions

1. The enriched free response technique is an **informative, explorative** tool which allows you to get a comprehensive grip on standard language dynamics (even on dynamism & modern prestige factors)
2. Dynamism is a **pivotal driving-force**, but it is not a stable, monolithic social meaning – it comes in different guises which play very different roles in standard language dynamics (neo-standards vs. multi-ethnolects)
3. Mass media boost and accelerate rapid linguistic diffusion, and it is **useless** to study the latter without including the former as a factor (but how?)
4. Social media (computer-mediated communication) offer us language data which allow us to study both the **actual diffusion** of an innovation and some **social meaning factors** which co-determine the change
5. The experimental and the media strand of SLICE should collaborate more

