Cultural models of the 'Standard Italian – dialect' continuum: a distributional semantic approach

Stefano De Pascale, Stefania Marzo, & Dirk Speelman

RU Quantitative Lexicology and Variational Linguistics

Overview

1. Cultural models in Cognitive Linguistics

2. A brief introduction to Distributional Semantics

3. Case study: 'Standard Italian – dialect' continuum
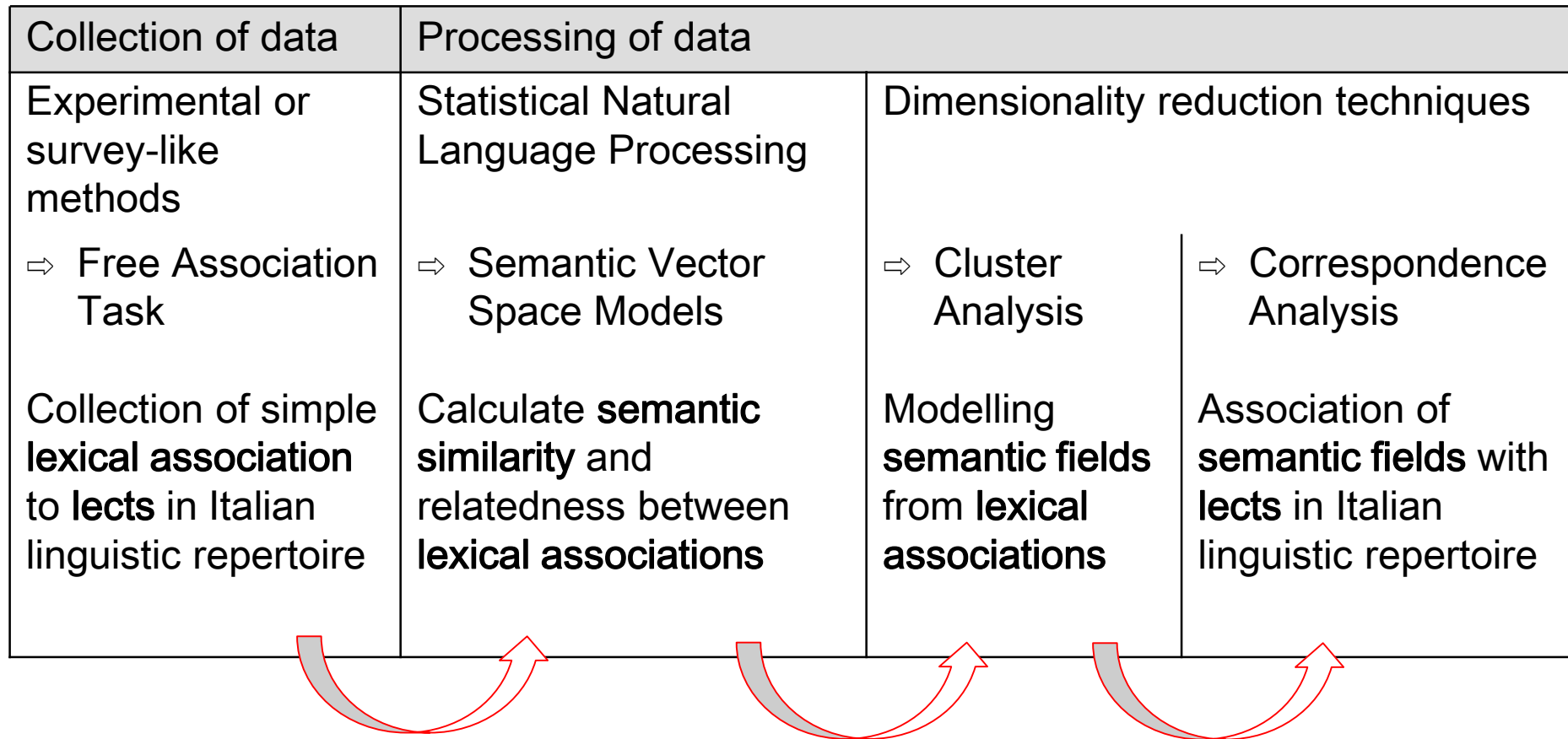
4. Discussion

5. Conclusion

1. Cultural models

Discipline	Research object	Methodology
Social Psychology	Attitudes	Quantitative
Critical Discourse Analysis	Ideologies	Qualitative
Cognitive Anthropology/Linguistics	Cultural models	Quantitative/qualitative

- "knowledge structures/cognitive schemas that are shared intersubjectively by a community"
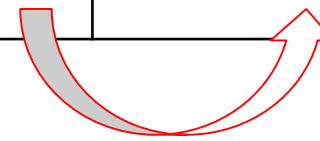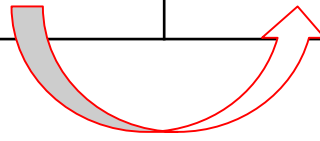 (Dirven, Wolf & Polzenhagen 2007: 1207)

1. Cultural models
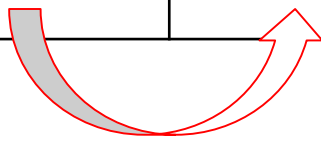
- ## Focus on cognitive organization of cultural meaning
  - Linguistic enterprise: centrality of semantic theory and description (prototype theory, conceptual metaphor theory)

- ## Focus on the community rather than on the individual
  - Large and diversified data collection

- ## How do we tap into and represent the cultural knowledge of large groups in a semantically responsible way?

1. Cultural models: methodological workflow

Collection of data	Processing of data		
Experimental or survey-like methods	Statistical Natural Language Processing	Dimensionality reduction techniques	
⇨ Free Association Task	⇨ Semantic Vector Space Models	⇨ Cluster Analysis	⇨ Correspondence Analysis
Collection of simple **lexical association** to **lects** in Italian linguistic repertoire	Calculate **semantic similarity** and relatedness between **lexical associations**	Modelling **semantic fields** from **lexical associations**	Association of **semantic fields** with **lects** in Italian linguistic repertoire

1. Cultural models: methodological workflow

Collection of data	Processing of data		
Experimental or survey-like methods	Statistical Natural Language Processing	Dimensionality reduction techniques	
⇨ Free Association Task	⇨ Semantic Vector Space Models	⇨ Cluster Analysis	⇨ Correspondence Analysis
Collection of simple **lexical association** to **lects** in Italian linguistic repertoire	Calculate **semantic similarity** and relatedness between **lexical associations**	Modelling **semantic fields** from **lexical associations**	Association of **semantic fields** with **lects** in Italian linguistic repertoire

2. Distributional Semantics

- Linguistic origin: Distributional Hypothesis
  - "You shall know a word by the company it keeps" (Firth)
  - A word's meaning can be induced from its co-occurring words

- DS in Computational Linguistics
  - Semantic Vector Spaces, Word Space Models …
  - Standard technique in statistical NLP for large-scale automatic modelling of (lexical) semantics
  - Words occurring in the same context have related meanings
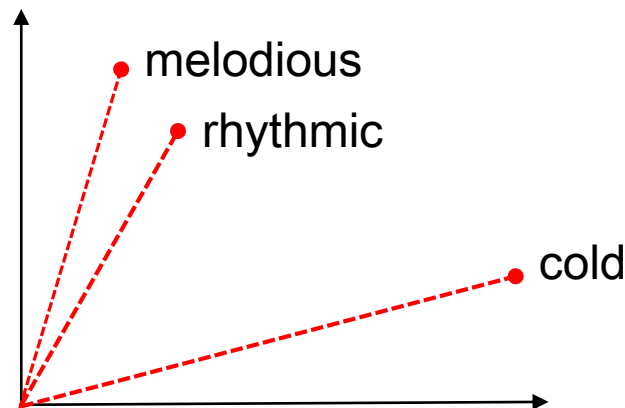
2. Distributional Semantics

- Weighted corpus frequencies by collocational strength (positive pointwise mutual information)

	musical	listen	piano	beat	tone	thermometer	winter
melodious	1.2	3.2	2.9	0.7	2.2	-0.3	-0.2
rhythmic	1.4	3.3	2.7	0.9	2.5	0.1	-0.6
cold	0.2	-0.2	0.1	-0.4	-0.2	3.7	2.2

2. Distributional Semantics

- Word by word similarity matrix (cosine similarity)
- Geometrical metaphor: semantic distances/spaces

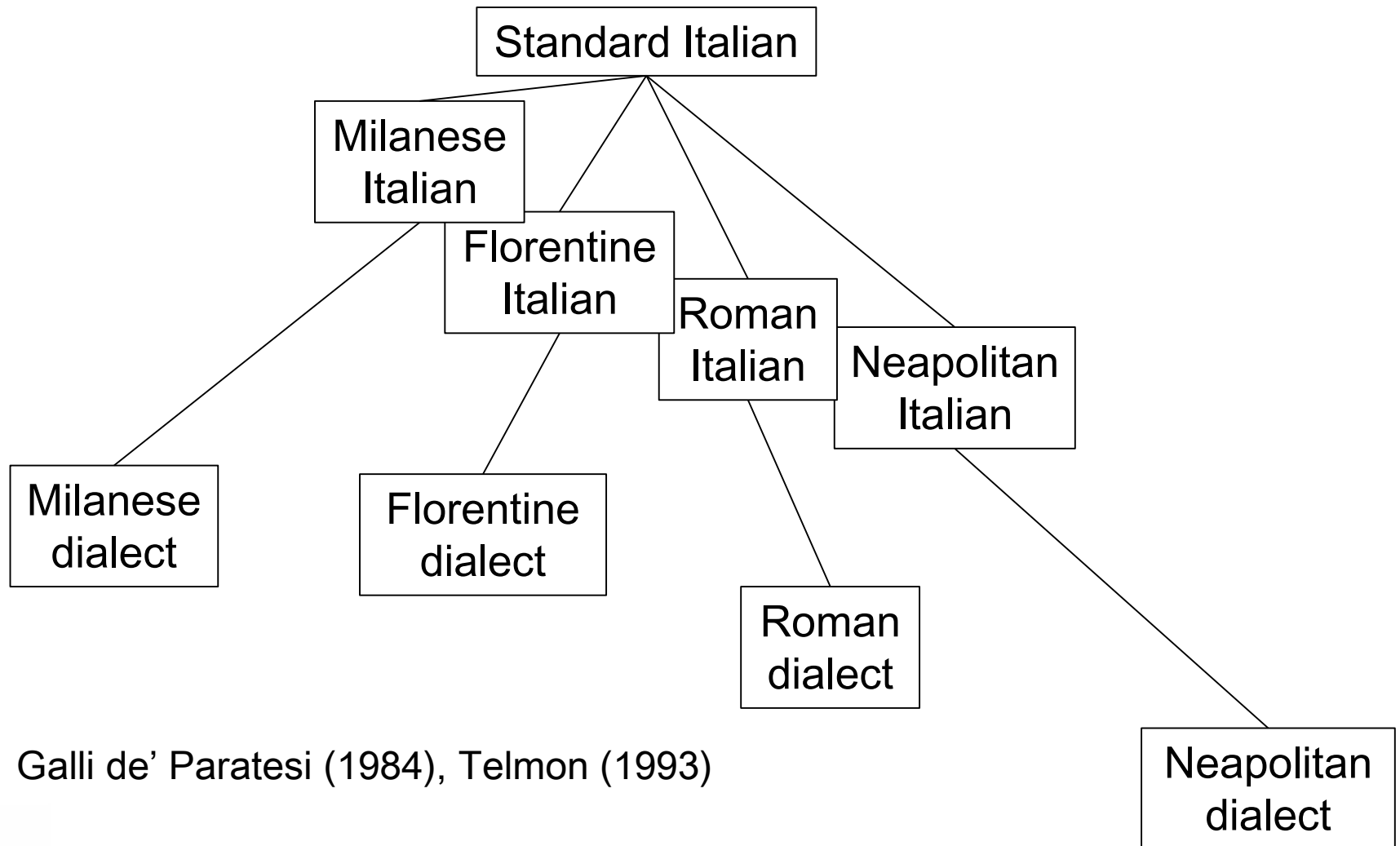|	melodious	rhythmic	cold
melodious	1	0.91	0.08
rhythmic	0.91	1	0.17
cold	0.08	0.17	1

Overview

1. Cultural models in Cognitive Linguistics

2. A brief introduction to Distributional Semantics

3. **Case study: 'Standard Italian – dialect' continuum**

4. Discussion

5. Conclusion

3. The Italian linguistic repertoire

- **Standard Italian (H-variety)**
  - Amended literary Florentine of the 14th century
  - Mainly a written language, learned and spoken by an elite

- **Italo-Romance 'rural' dialects (L-varieties)**
  - 5 systems scattered across Italy, very often mutually unintelligible
  - The language of everyday communication

- **Critical juncture: economic miracle in '50s and '60s**
  - Standard Italian gains access to domains formerly reserved to the dialects => Italian regional varieties
  - Abandoning dialects as sole means of communication

3. Regional variation in Italian

Standard Italian

Milanese Italian

Florentine Italian

Roman Italian

Neapolitan Italian

Milanese dialect

Florentine dialect

Roman dialect

Neapolitan dialect

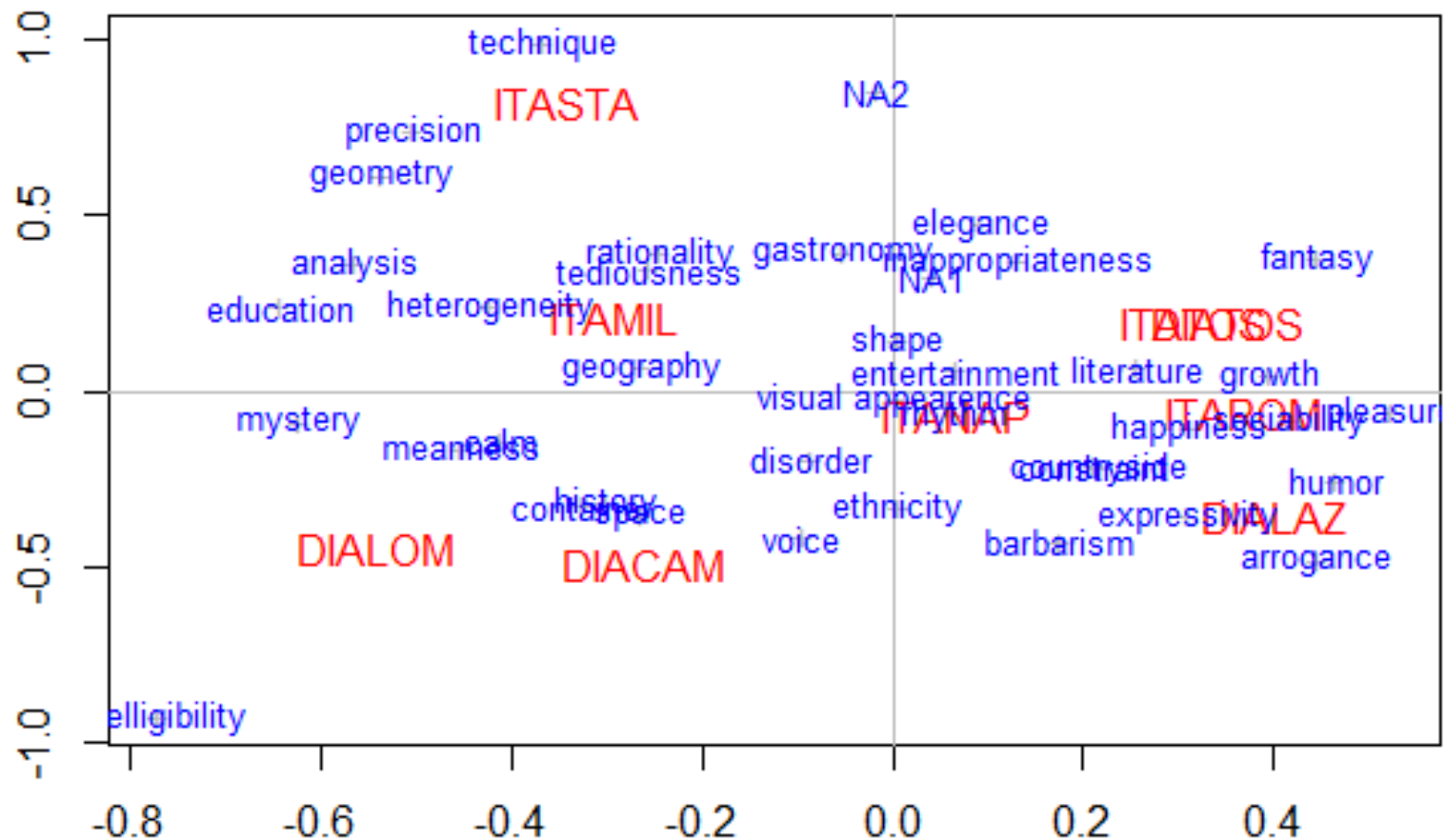Galli de' Paratesi (1984), Telmon (1993)

3. Free Association Task

- "Give the first 3 adjectives that come to mind for":
  - Standard Italian
  - A base dialect from Lombardy, Tuscany, Latium and Campania
  - Italian spoken in Milan, Florence, Rome and Naples

- Database of 5628 adjectives (935 types)
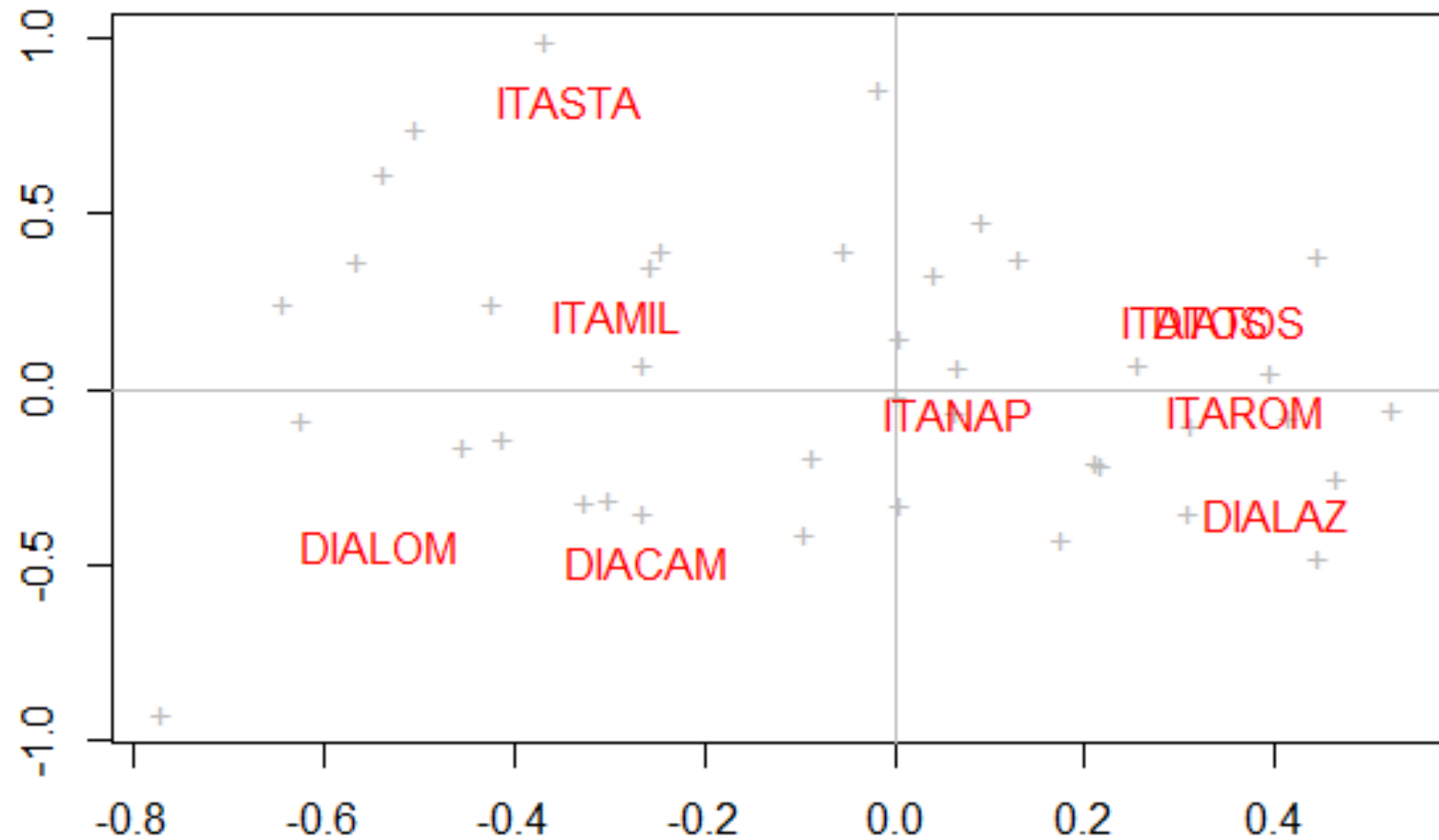- ± 200 university students all over Italy

3. Processing data

- Parameter selection for Vector Space Model; building a similarity matrix with all adjectives

- Cluster analysis on similarity matrix yields 40 clusters of related/similar adjectives (= semantic fields)

- Correspondence analysis on the frequencies of the 40 semantic fields for each of the 9 varieties

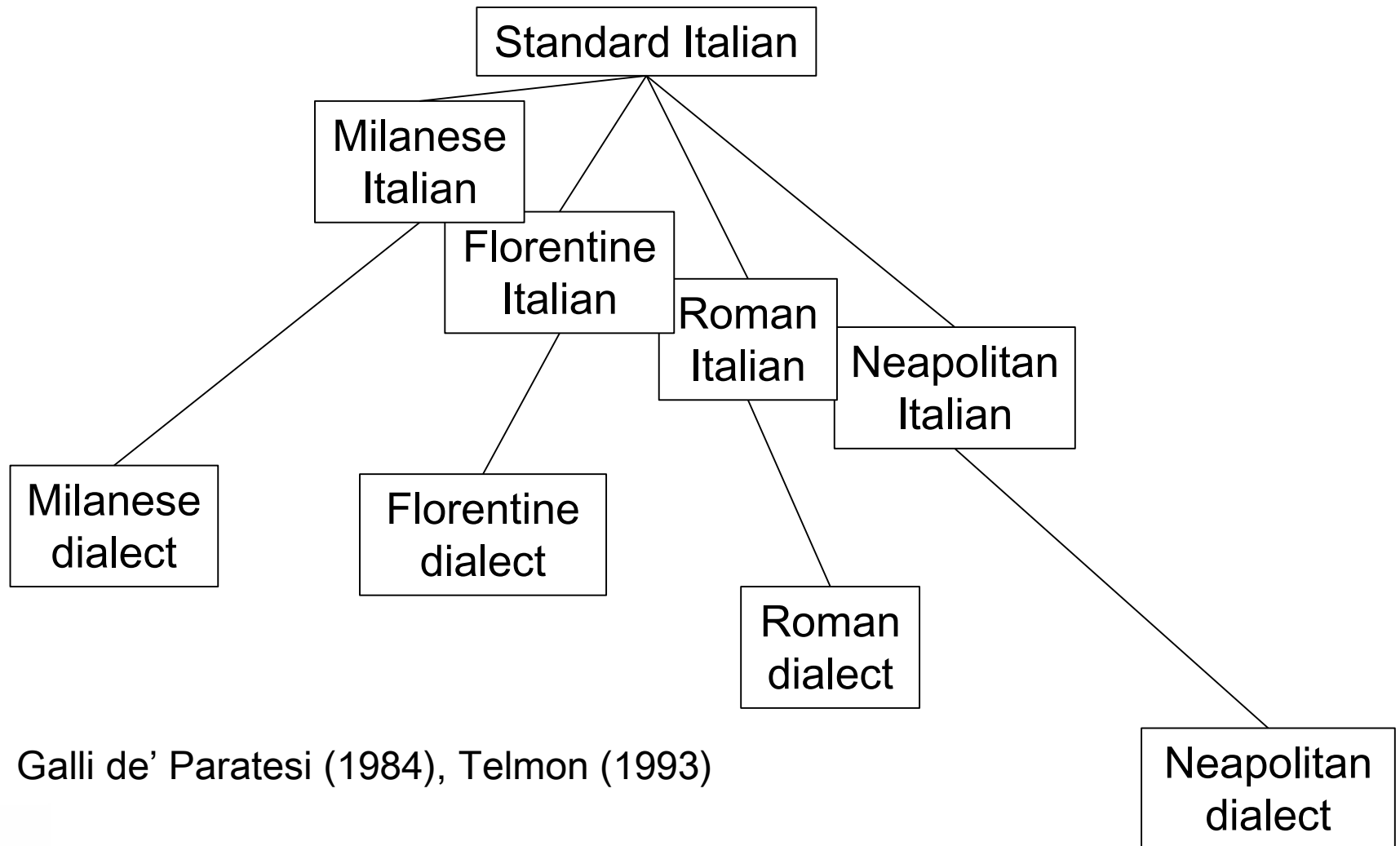⇨ Use a two-dimensional plot for an eight-dimensional semantic space!

Semantic fields associated with Italian lects

Semantic fields associated with Italian lects
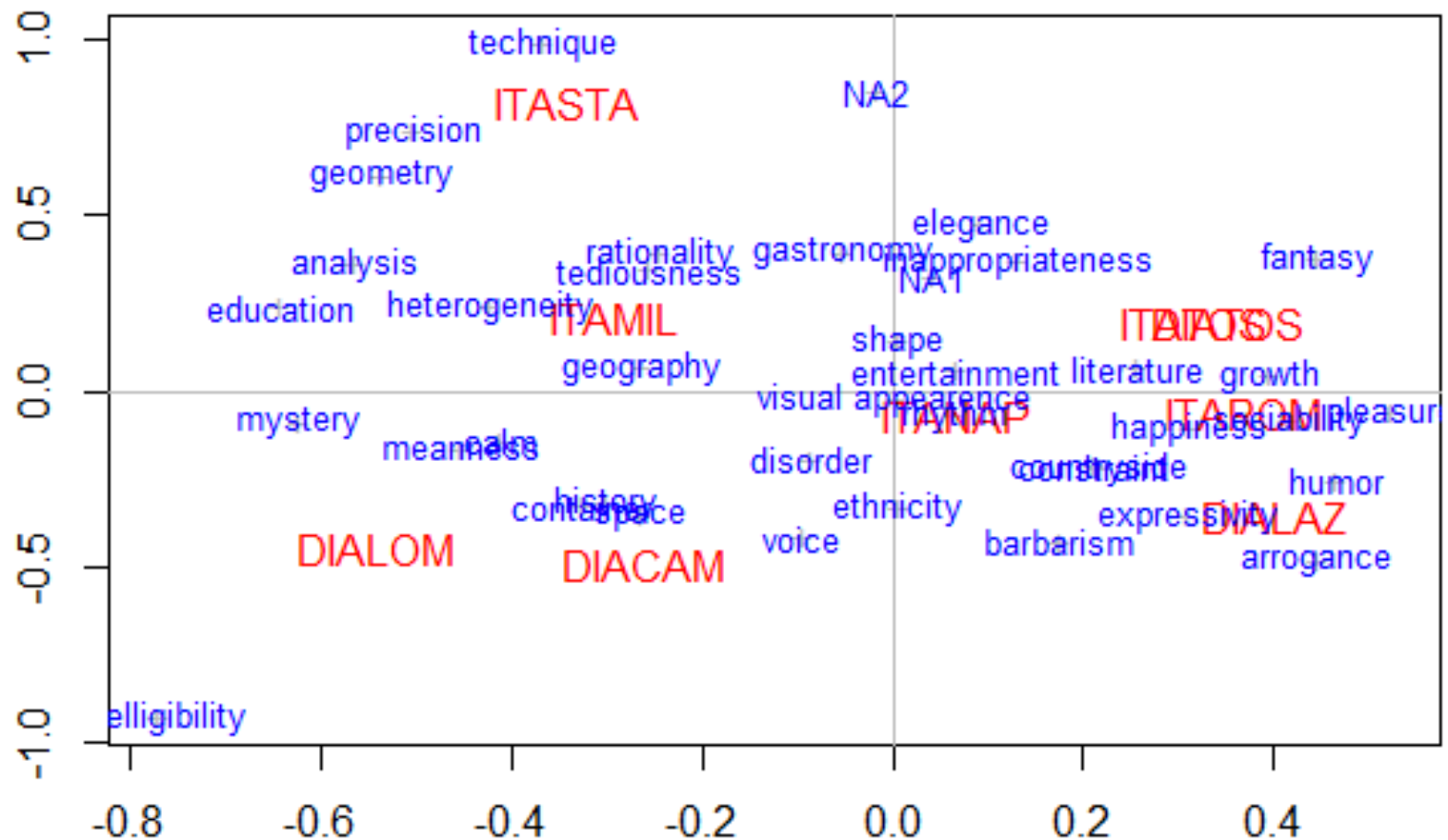
3. Regional variation in Italian

Standard Italian

Milanese Italian

Florentine Italian

Roman Italian

Neapolitan Italian

Milanese dialect

Florentine dialect

Roman dialect

Neapolitan dialect

Galli de' Paratesi (1984), Telmon (1993)

4. Discussion

Conceptual-semantic distances between lects correlate with **linguistic distances** between lects

1. Clear on vertical dimension of 'Italianization':
   - ITAMIL closest to ITASTA, but far from DIALOM
   - ITANAP far from ITASTA, and far from DIANAP

2. No horizontal dimension of 'geographic contiguity':
   - DIACAM and DIALOM are close together, just like ITANAP and ITAMIL
   - No horizontal differentiation between DIALAZ, ITAROM, DIATOS, ITAFIR

Semantic fields associated with Italian lects

4. Discussion

- Standard Italian and Milanese Italian (topleft quadrant): **rationality, tediousness, education** and **precision**

- Lombard and Campania dialects (bottomleft quadrant): **intelligibility, obscurity, space** and **history**

4. Discussion

- Florentine Italian and Tuscan dialects (topright quadrant): **literature, fantasy, elegance** and **inappropriateness (=negative use**)

- Neapolitan and Roman Italian, as well as Latian dialects (bottomright quadrant): **ethnicity, barbarism, expressivity, countryside**

5. Conclusion

Cultural conceptualizations seem to organize/partition the Italian linguistic repertoire in three/four groups

	Interpretation 1	Interpretation 2
H-varieties	Standard Italian and Milanese Italian	
L-varieties	Southern dialects and northern dialects	
More 'Italianized' regiolects	Florentine Italian and Tuscan dialects	Central dialects and 'central' Italian
Less 'Italianized' regiolects	Roman and Neapolitan Italian, Latian dialects	Neapolitan Italian

5. Conclusion

Future work:

- Elaborate semantic analysis of semantic fields (prototype theory, conceptual metaphor theory)

- Investigate variability in cultural conceptualizations in different regions of Italy (North, Centre, South)

- Explore other ways to visualize semantic space

Thank you very much!

for further information:
stefano.depascale@kuleuven.be

http://wwwling.arts.kuleuven.be/qlvl

Bibliography

- Dirven, R., Wolf, H.-G., & Polzenhagen, F. (2007). Cognitive Linguistics and Cultural Studies. In D. Geeraerts & H. Cuyckens (Eds.), *The Oxford Handbook of Cognitive Linguistics* (pp. 1203–1221). Oxford: Oxford University Press.

- Galli de' Paratesi, N. (1984). *Lingua toscana in bocca ambrosiana. Tendenze verso l'italiano standard: un'inchiesta sociolinguistica*. Bologna: Il Mulino.

- Telmon, T. (1993). Varietà regionali. In A. A. Sobrero (Ed.), *Introduzione all'italiano contemporaneo* (Vol. 2, pp. 93–149). Roma-Bari: Laterza.