

Eliciting comparable spoken data in minor
languages: first observations from the
corpus *Kontatti*

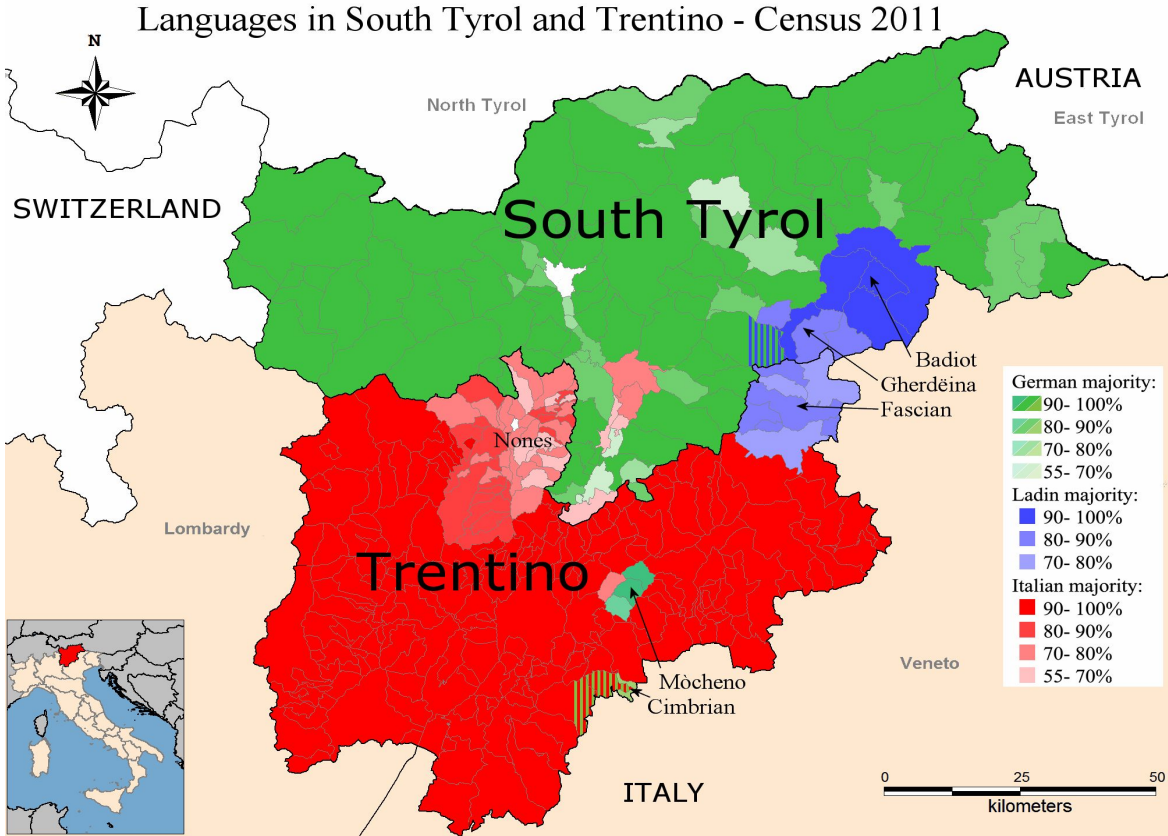
*Marta Ghilardi / Silvia Dal Negro,
Libera Università di Bolzano*

Corpus Linguistics: the Brown Corpus (1964) vs Kontatti (2018): what changes?

| Brown Corpus |
|-----------------------------|
| Written Data |
| American English |
| No sociolinguistic metadata |
| Million of words |

| Kontatti |
|--------------------------|
| Spoken Data |
| More than one language |
| Sociolinguistic metadata |
| Thousand of words |

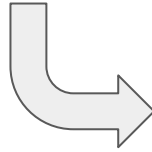
Trentino and South-Tyrol



Repertoires in the area

| | |
|--------------------------------|---|
| German Community | H>German/Italian L>Tyrolean/(Italian) |
| Bassa Atesina Community | H>Italian/(German) L>Tyrolean/Italian/Trentino |
| Italian Community | H>Italian/((German)) L>Italian/(Dialect traces) |
| Cimbrian Community | H>Italian/(German) L>Cimbrian |
| Ladin Community | H>Italian/German in Gardena (Ladin) L>Ladin/ (Tyrolean in Gardena) |

Kontatti 2.0: investigated areas



Bassa Atesina

Cimbrian

Ladin Valleys: Fassa, Fiemme, Gardena

Trentino

| Place | Interviews | Informants | Time |
|--------------------------|------------|------------|-------------|
| Bronzolo (Bassa Atesina) | 12 | 13 | 2:12:21 |
| Luserna (Cimbrian) | 4 | 9 | 35:55 |
| Fassa (Ladin) | 5 | 6 | 35:58 |
| Fiemme (Ladin) | 7 | 8 | 48:37 |
| Gardena (Ladin) | 3 | 6 | 1:01:43 |
| Val di Non (Italian) | 1 | 2 | 12:49 |
| Vallagarina (Italian) | 1 | 2 | 32:35 |
| Folgaria (Italian) | 1 | 4 | 48:49 |
| Tot. | 34 | 50 | 6:48 |

Corpus of spoken and conversational data

Digital recordings, transcribed and analysed
(Elan)

Monolingual corpora

- One language: different language varieties, registers.

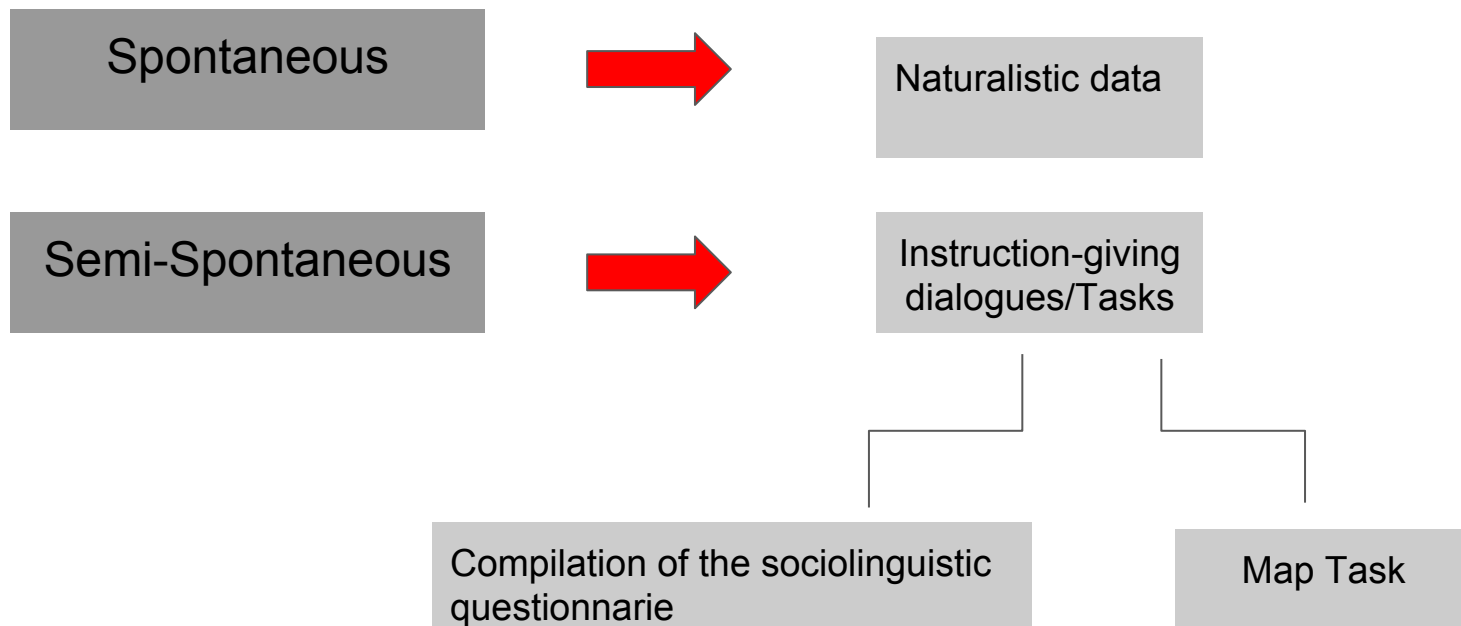
Multilingual corpora

- Different linguistic codes.

Parallel corpora

Bilingual people

Methodology (semi)spontaneous spoken data



Semi-spontaneous speech: Reformulations and spontaneous translations (between mother and daughter)

K_Fol01_S3 pensionata occupazione metti lineetta pensionata metti lì pensionata
lì sóra ok pensionata

K_Fol01_S3 scólta chi bisòn che té scrivi occupazione pensionata qua té dévi
ndè che te hai fàt quéla ics lì té devi scrìver pensionata, Elisa

*Retired employment (you) write underscore retired (you) write there retired up ok
retired> (0) Listen to me it is necessary that you write employment retired here
where you marked that “x” there you have to write retired, Elisa*

K_Fol01_S3 e lingua dei genitori

K_Fol01_S3 la tó màma el tó papà che lingua parlàveli

Language of your parents> Your mother and your father, what language did they speak?

K_Fol01_S3 lingue conosciute in ordine di apprendimento e imparata per prima

K_Fol01_S3 la lingua cos'è che tài imparà prima

Languages known in order of acquisition and learnt as first> The language, what is the language that you learned first?

Comparison with Vivaldi

Methodology



Translation tasks

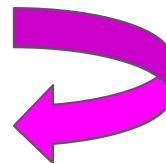


Recording of a translation of a list of words/sentences;
translation of the *Parable of the Prodigal Son*



No spontaneous dialogical
communication/translation

Loyalty to the written version
1 speaker=1place



Map Tasks as a way to elicit comparable data

Code-switching as a turn-taking device

| Main language of interaction | switch | n. | Example |
|------------------------------|--------------|----|---|
| German | no | 17 | also iã ebm mir sain (1.8) afn parkplätz |
| German | Ital./Trent. | 5 | alóra du bisch schtartäsch fo dän hotäl dâ |
| Trentino | no | 3 | alóra da li té vai én Zó dé dói |
| Trentino | Italian | 3 | allora té vègni av- té vègni avanti dé dó quadri |
| Italian | no | 4 | allora # partiamo dalla zona numero uno |
| Ladin | no | 5 | bën po l A chël ie ehm a man ciancia dla plata gonz su insom |
| Ladin | German | 2 | also ## tu es pu la Piazz |
| Ladin | Ital./Trent. | 5 | alóra parte dal prum # riquadro |
| Cimbrian | Italian | 6 | alóra du geast # atta recht |

Map Tasks as a way to elicit comparable data

Code-switching as a turn-taking device: some results

| | | |
|----|--|---|
| 01 | Italian is the most switched to language | <ul style="list-style-type: none">• Italian, or Italian with Trentino, is the majority language |
| 02 | Ladin and Cimbrian are never switched to | <ul style="list-style-type: none">• Cimbrian and Ladin are typically minority languages |
| 03 | Cimbrian always implies some switching | <ul style="list-style-type: none">• Cimbrian is the weakest language in the area |

Map Tasks as a way to elicit comparable data

Dealing with target words

Sigsch T fän där **tabaktrafik**?

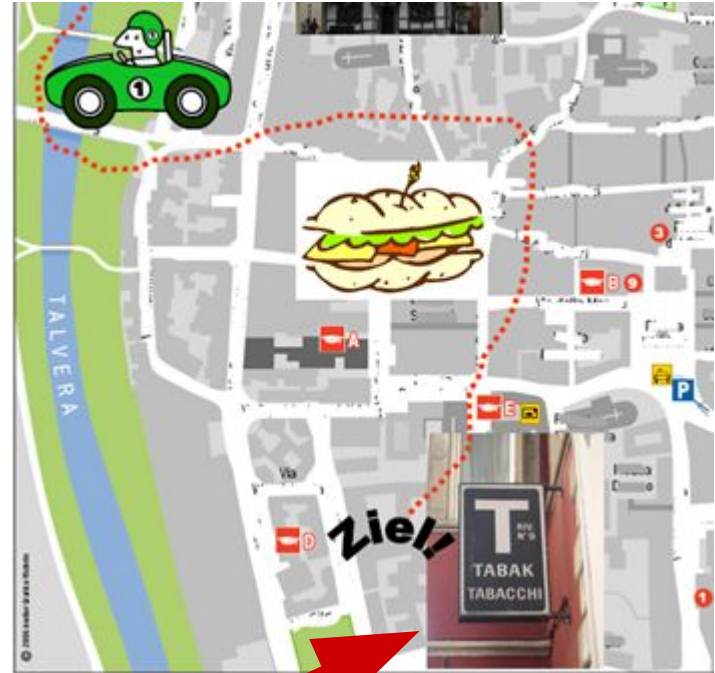
unt därnoch foorsch äh ban schilt **tabak** net

ja, **tabacchi**, genau

häm foorsch när # or bis wo 's foto fän **tabacchino** unheb

in **tabacchin** do schun hån i gsegn

rächts # fän **tabakilä**



Conclusion: corpora of minority languages, which applications?

Corpus-Driven

- No prior linguistic theories to confirm. Theories come from the analysis.



Reconstruction of the community repertoires

Currently: Working principle of subject expression system in Trentino dialect

and

Corpus-Based

- Expound, exemplify prior linguistic theories and description



Exploring the real degree of multilingualism in the area

Analysing the CS/CM data in order to verify the theories of multilingual speech

Comparing our subject system with prior dialect data collected with different methodology and different purposes

Methodology and minority languages

Map Task

Eliciting comparable data in multilingual context .

Sociolinguistic questionnaire

Reformulation and cultural translation.

Sociolinguistic Information

Social parameters can play the role of sociolinguistic variables, and cast light on the diachronic study of the language. Language attitudes arisen from the questionnaires can better explain the community repertoires.

Thank you for your attention!