Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan

unibz

Code-mixing patterns

Investigating variation in a German-Italian bilingual corpus

Silvia Dal Negro

(Free University of Bolzano-Bozen)

This presentation

o Bi- plurilingual corpora and the documentation of bilingual speech.

o The sociolinguistic context: German and Italian in South Tyrol.

o Modelling bilingual speech: a corpus-based research on insertions.

Documenting bilingual speech

A corpus-based perspective

The priority of language use in contact linguistics

Studies of contact-related language change, while usually acknowledging the importance of sociolinguistic norms and constellations in motivating change [...], tend largely to disregard aspects of actual bilingual linguistic performance at the discourse level.

(Matras / Sackel, 2007: 848)

The priority of systematic empirical research in contact linguistics

From the point of view of grammar and sociolinguistics, the priority is to set up comparative studies which will allow us to gauge the effect of the different types of variable which have been found to affect CS, both
linguistic and sociolinguistic. (Gardner-Choloros, 2009: 177)

Some authors treat the corpora collected primarily as the potential source for individual codeswitching examples, meant to confirm or argue against a particular 'model' or 'constraint' for codeswitching in general. In this case the overall features of the corpus play much less of a role, if any role at all. This is not the approach we want to take here. On the contrary, we want to link the study of codeswitching to the insight from the sociolinguistic research of the last 40 years or so that bilingual speech, just as monolingual speech, shows variation, but that this variation is patterned or structured, not random. (Deuchar, Muysken & Sung-Lan Wang, 2007: 298-299)

The situation is exacerbated by the fact that there is so little pertinent data, as will be detailed herein, the (possibly related) methodological predilection for butterfly collecting, and a concomitant devaluing of accountable analyses of large corpora of spontaneous bilingual speech. (Poplack & Dion, 2012: 280)

«Plurilingual» vs. «multilingual» corpora

To differentiate our materials from these multilingual corpora, here we use the term "plurilingual corpora" to designate corpora that illustrate not only instances of codeswitching and codemixing but also *languaging* through the use of various linguistic resources […].

(Léglise / Alby, 2016)

The issue of defining a corpus of bilingual speech

GIVEN:

➢ a potentially bilingual community and

➢ potentially bilingual speakers

WE LOOK FOR:

• contact phenomena being or not being present and

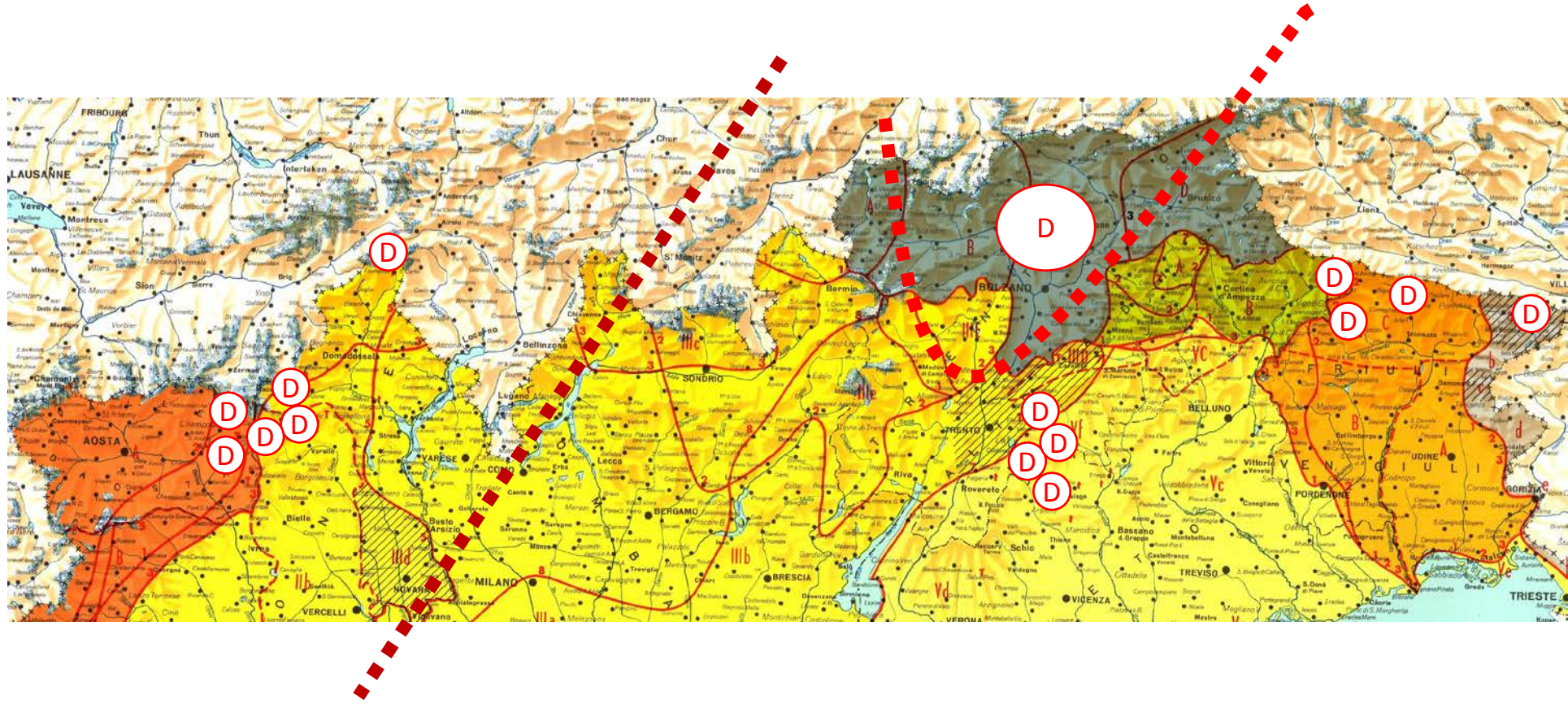• the type of contact phenomena actually been observed

Fieldworkers «force» speakers into a «bilingual mode» (Grosjean, 2008)
by letting members of the same bilingual community interact in natural or
naturalistic situations.

The sociolinguistic context

German as a minority language in Italy

German minorities in Italy

Excerpt from the *Carta dei dialetti d'Italia* (G.B. Pellegrini)

South Tyrol

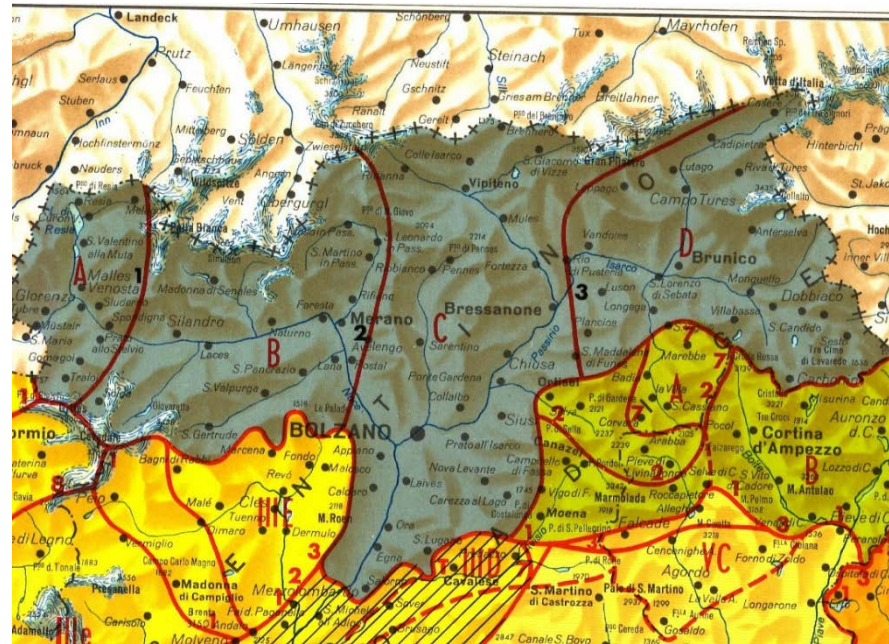ICL20 - The Dynamics of Language

Focus on South Tyrol

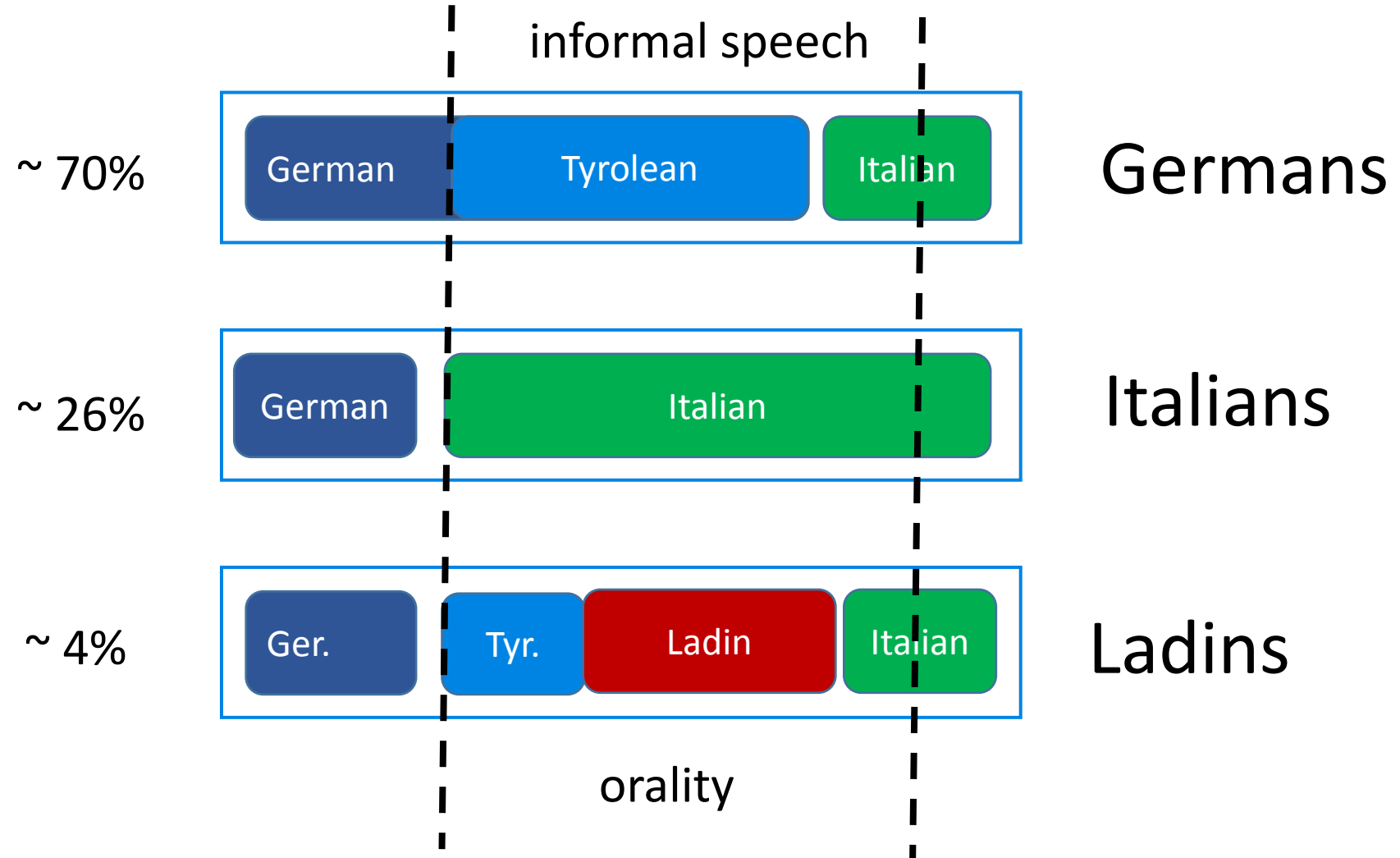*De jure* bilingualism (Standard German / Standard Italian)

South Tyrol has been part of Italy for «only» 100 years.
Forced italianization policies between 1922-1945. Full recognition of German since 1972.

De jure trilingualism (Standard German / Standard Italian / Ladin) in two valleys

Three main speech communities

«Kontatto»: a corpus-based project
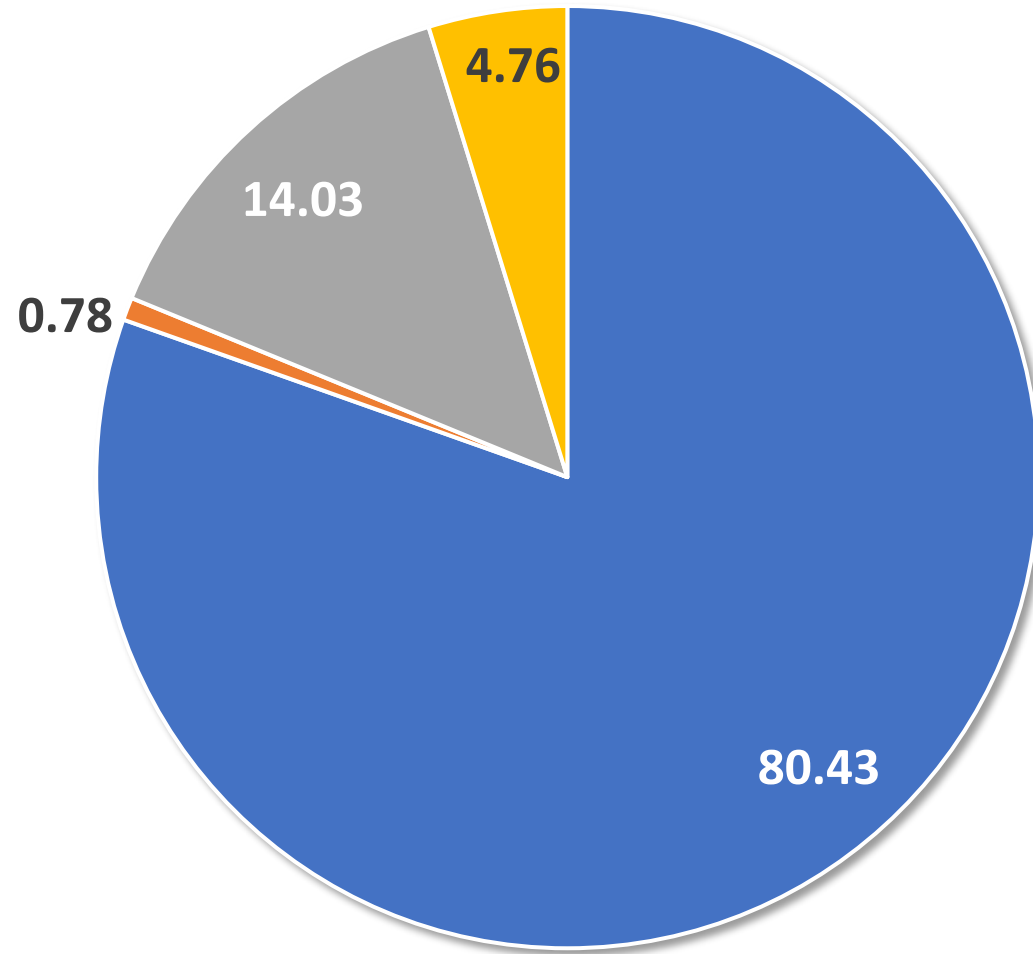
📍	📄	⏳	🔗
South Tyrol (special focus on **bordering areas**)	**55 (61)***	**17:57 h**	**147.596 words**

Systematic annotation and the making of a corpus

ELAN

MPI - TLA

Modelling bilingual speech

Focus on **insertions** and their role in triggering bilingual speech

[Muysken 2000; Clyne 2003; Matras 2009]
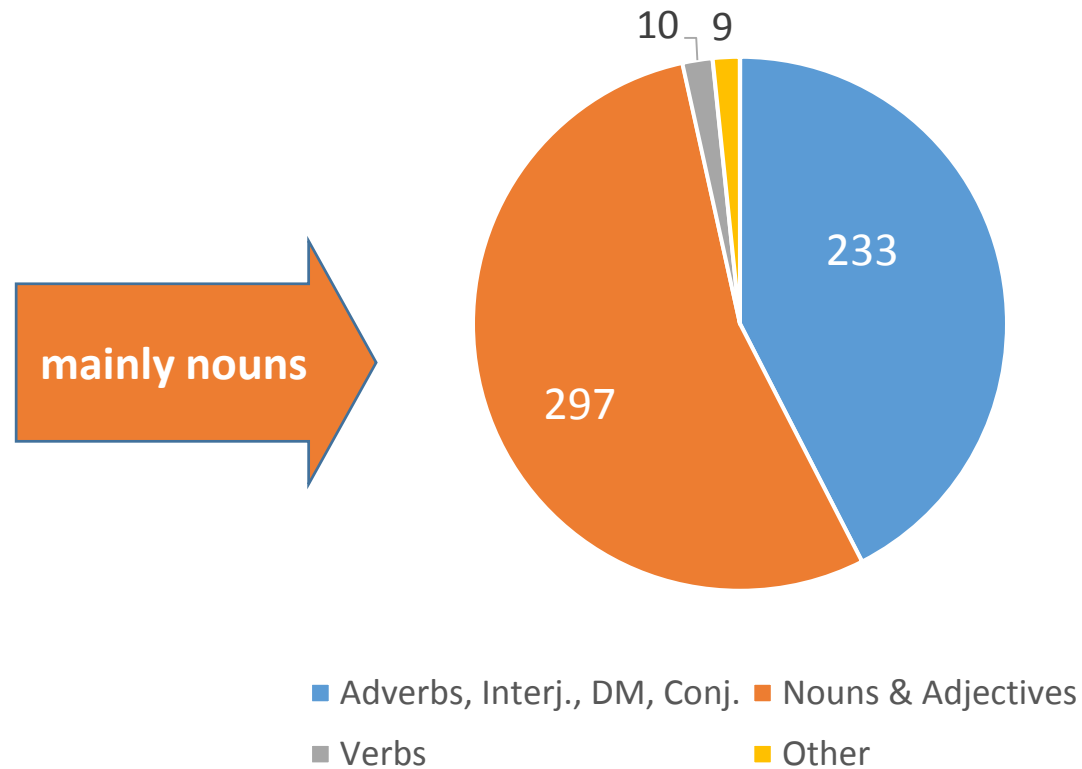
Investigating patterns: single word insertions

	Towards Italian	Towards German	**Total**
Peripheral	735 [1]	127 [4]	862
Nested	495 [2]	54 [3]	549
Total	1230	181	**1411**

Italian/Trentino > German
Peripheral insertions > Nested insertions

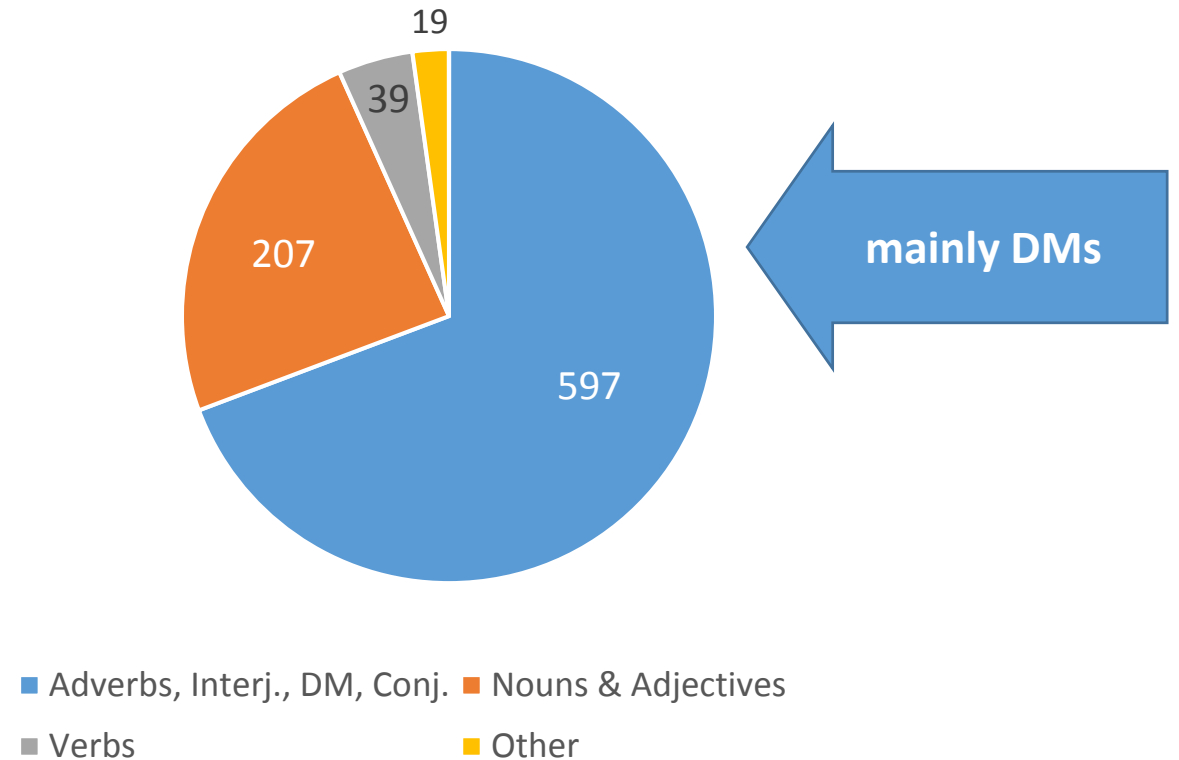1) **ecco # si årbetet a ba schportlår**
   'there you are, she also works for Sportler'
2) **zäm wo där calcetto isch**
   'there where the table-football is'
3) **in dialetto tedesco unt italiano standard**
   'in German dialect and Standard Italian'
4) **sono favorevole alle scuole bilingui xxx jå quello molto d'accordo**
   'I am in favour of bilingual schools, yes, that I approve of very much'

Insertional patterns and POS frequency

Nested insertions

10 9

297

233

mainly nouns →

Peripheral insertions

19

39

207

597

← **mainly DMs**

■ Adverbs, Interj., DM, Conj. ■ Nouns & Adjectives
■ Verbs ■ Other

■ Adverbs, Interj., DM, Conj. ■ Nouns & Adjectives
■ Verbs ■ Other

Insertional pattern and type frequency
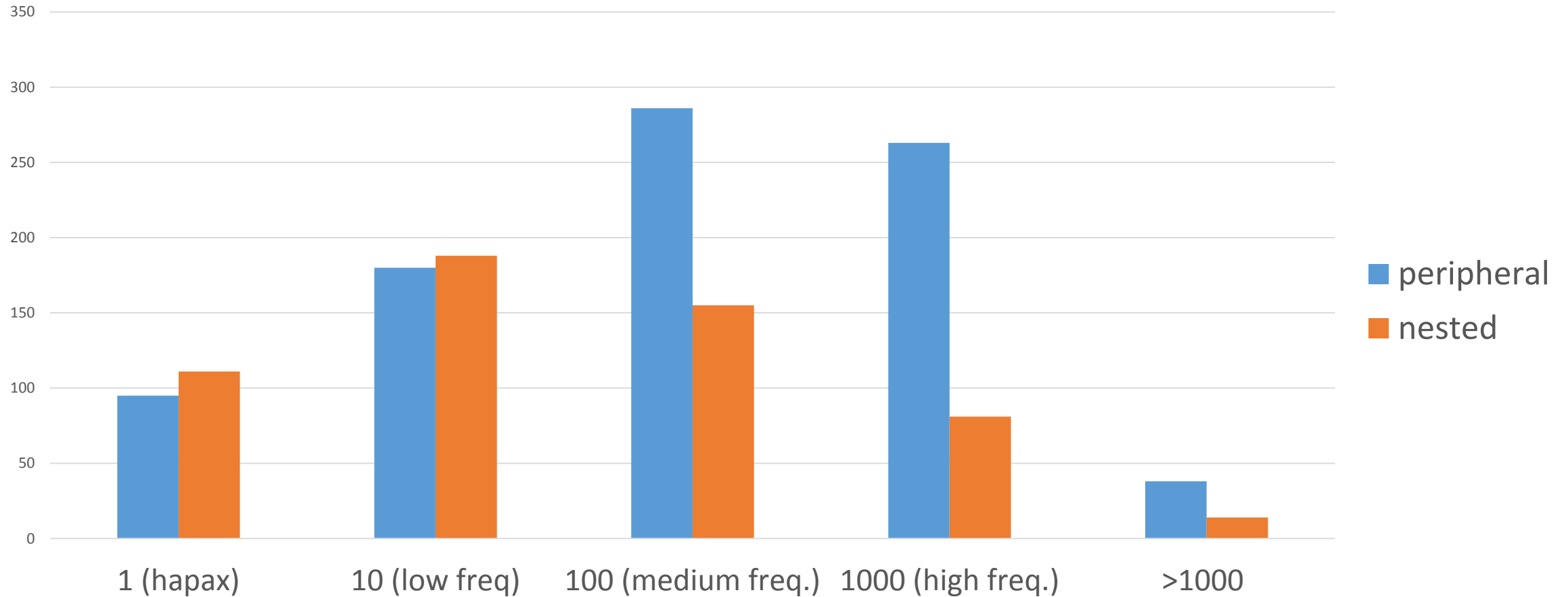
dår acquisto des jaarhundårts wor des °wail i glab°

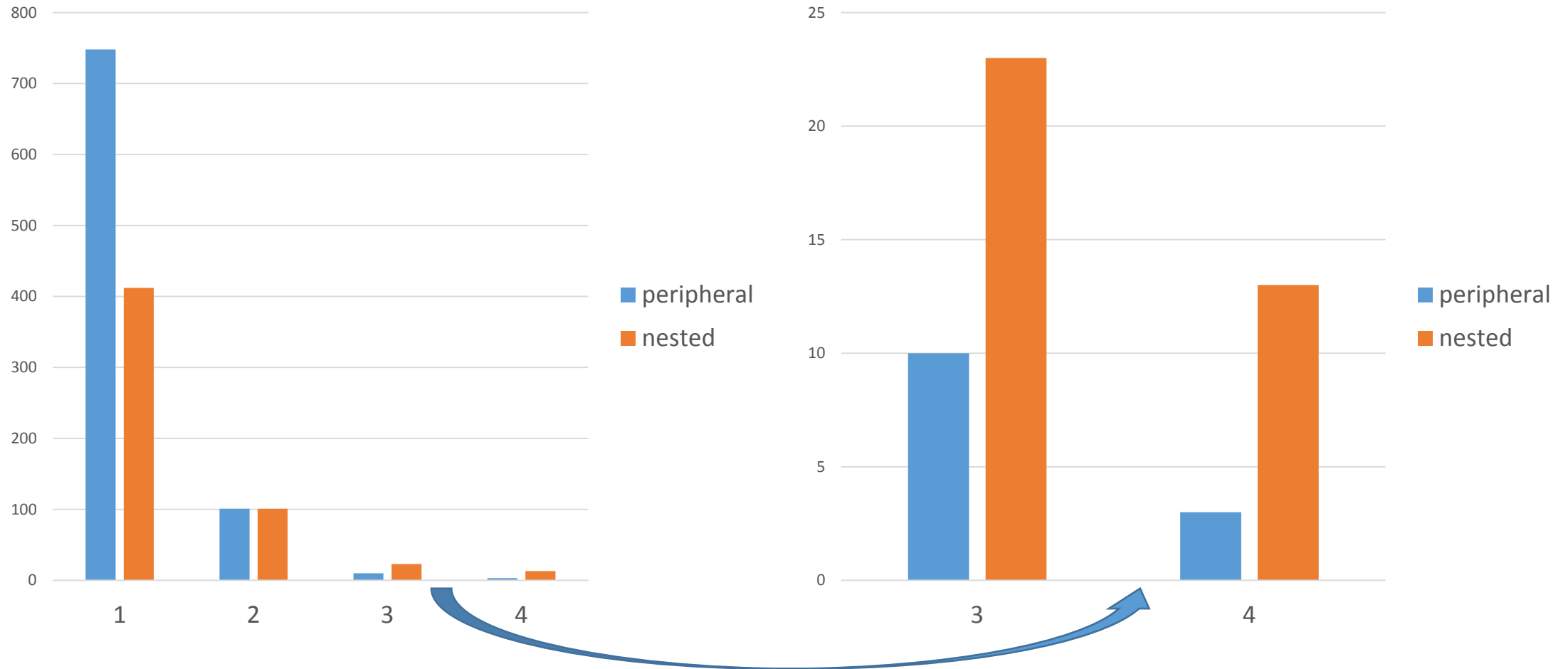*The purchase of the century was that, because I think…*

$\rightarrow$ nested insertion: noun, hapax in the corpus

dai kimm her zwischn zwåa mandår # madonna

Come on! Come here between to men, bloody hell!

$\rightarrow$ two peripheral insertions: interjections (freq 213 and 18)

Insertional type and n. of switches/utterance

dånn jaz muscht di ve_tre få di effe_ventitré in hundårt comuni fåscht italiani schikkn

Then now you must send the V-3 of the F-23 (forms) to almost one hundred Italian municipalities.

→ lexical insertions with referential function denoting specific referents trigger

- further lexical insertions that may lead to
- language alternation, i.e. the change of the base language or
- congruent lexicalization

Factors favoring the activation of the L2
Logistic regression with VARBRUL

Dependent variable:

- two or more switches in the utterance

Factor groups: ✅
1. type of insertion
2. direction of switch ❌
3. POS of inserted word ❌
4. frequency of inserted word ❌
5. speaker's dialect area ❌

Input: 0.171
Factor group # 1
Peripheral insertion: 0.425
Nested insertion: 0.617 ←

Log likelihood = -645.181 Significance = 0.000

Dependent variable:

- one switch only in the utterance

Factor groups: ✅
1. type of insertion
2. direction of switch ❌
3. POS of inserted word ❌
4. frequency of inserted word ❌
5. speaker's dialect area ❌

Input: 0.829
Factor group # 1
Peripheral insertion: 0.575
Nested insertion: 0.383 ←

Log likelihood = -645.181 Significance = 0.000

Two insertional styles were detected

Nested insertions

- have a lexical content, involve low frequency words
- activate the «embedded language» and trigger alternation
- pertain the domain of code-mixing and more generally bilingual speech

Peripheral insertions

- have a discursive function, involve high frequency words
- do not activate the «embedded language»
- have much in common with loanwords in monolingual speech

Thanks for your attention

http://kontatti.projects.unibz.it/