Syntactic phenomena in light of prosody-oriented segmentation in spoken Brazilian Portuguese

Heliana Mello, Giulia Bossaglia & Tommaso Raso (UFMG)

C-ORAL BRASIL · UFMG · FAPEMIG · CNPq Conselho Nacional de Desenvolvimento Científico e Tecnológico · LEEL

Summary

- L-AcT

- C-ORAL-BRASIL

- Some figures about conjunctions

- Coordination

- Subordination

- Insubordination

- Challenges

L-AcT

- Language into Act Theory (L-AcT): an extension of the Speech Act theory (Austin, 1962), derived from corpus-based research carried out on large spontaneous speech corpora, developed mainly by Emanuela Cresti at LABLITA, University of Florence (Cresti, 2014).

- Utterance is assumed as the reference unit for speech; it is pragmatically defined and corresponds to a speech act; it is identified through prosodic cues.

Cresti (2014)

- Syntactic properties of spontaneous speech in the Language into Act Theory: data on Italian complements and relative clauses.
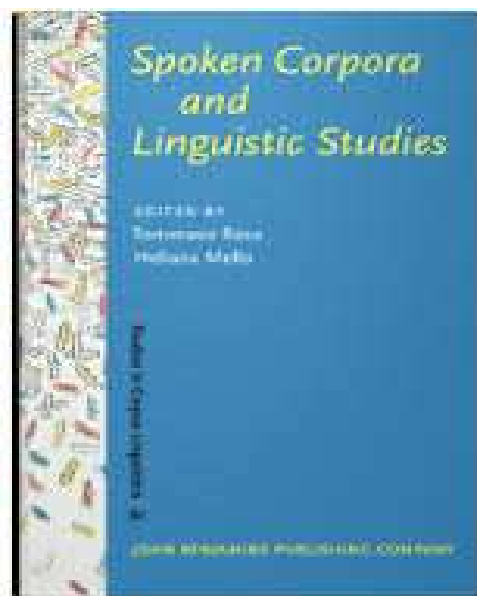
L-AcT

- In L-AcT, syntactic reduction finds its explanation functionally, on pragmatic information, because the final linguistic nature of the utterance corresponds to the combination of semantic/syntactic islands, going hand in hand with information units (IU). IUs partake in Information Patterns; each IU has a pragmatic function.

- Thus, the study of syntax in L-AcT is carried out considering the utterance (and its information patterning).

L-AcT: Textual Information Units

Type	Tag	Function
Comment	COM	It accomplishes the illocutionary force of the utterance.
Topic	TOP	It establishes the domain of application of the illocution expressed by the Comment.
Appendix of Comment	APC	It integrates the text of the Comment and concludes the utterance.
Appendix of Topic	APT	It gives a delayed integration of the information given in Topic.
Parenthesis	PAR	It provides instructions about how the utterance, or a part of it, has to be interpreted; it has a backward or forward scope.

L-AcT: Textual Information Units

Type	Tag	Function
Locutive Introducer	INT	It signals that the subsequent locutive space has to be interpreted by means of different coordinates (generally, it introduces metaillocutions).
Multiple Comments	CMM	They constitute a chain of Comments forming an illocutionary prosodic pattern, which is interpreted holistically.
Bound Comment	COB	It constitutes a chain of Comments, produced by progressive adjunctions which follow the flow of thought. Chains of COB are called Stanzas.
List of Topic	TPL	List of two or more TOP units semantically and syntactically connected, forming one prosodically marked major unit.
List of Parenthesis	PRL	List of two or more PAR units semantically and syntactically connected, forming one prosodically marked major unit.

L-AcT: Dialogic Information Units

Type	Tag	Function
Incipit	INP	It opens the communicative channel in order to start a dialogic turn or an utterance, bearing contrastive value.
Conative	CNT	It pushes the interlocutor to participate or to stop a non-collaborative behaviour in a dialogue.
Phatic	PHA	It controls the communicative channel, contributing to its maintenance and stimulating the interlocutor towards social cohesion.
Allocutive	ALL	It specifies the addressee of the message, keeping his attention and having a social cohesive function.
Expressive	EXP	It provides emotional support to the speech act, marking social cohesion.
Discourse Connector	DCT	It connects different parts of the discourse, marking continuity.

L-AcT: Units without information value

Type	Tag	Function
Scanning Unit	SCA	It is a tonal part of a bigger information unit, whose locutive content needs to be scanned, normally for expressive or speech limits reasons.
Interrupted Unit	i-[TAG]	It is part of an information unit which is interrupted by a PAR or dialogic unit.
Empty Unit	EMP	Unit whose locutive content is not to be considered as part of the utterance, as it happens in case of 1) retracting; 2) interrupted last unit of an utterance.
Time taking Unit	TMT	It corresponds to the so-called filled pause.
Unclassified	UNC	Unit to which it is not possible to attribute another tag for some reason.

C-ORAL-BRASIL
www.c-oral-brasil.org
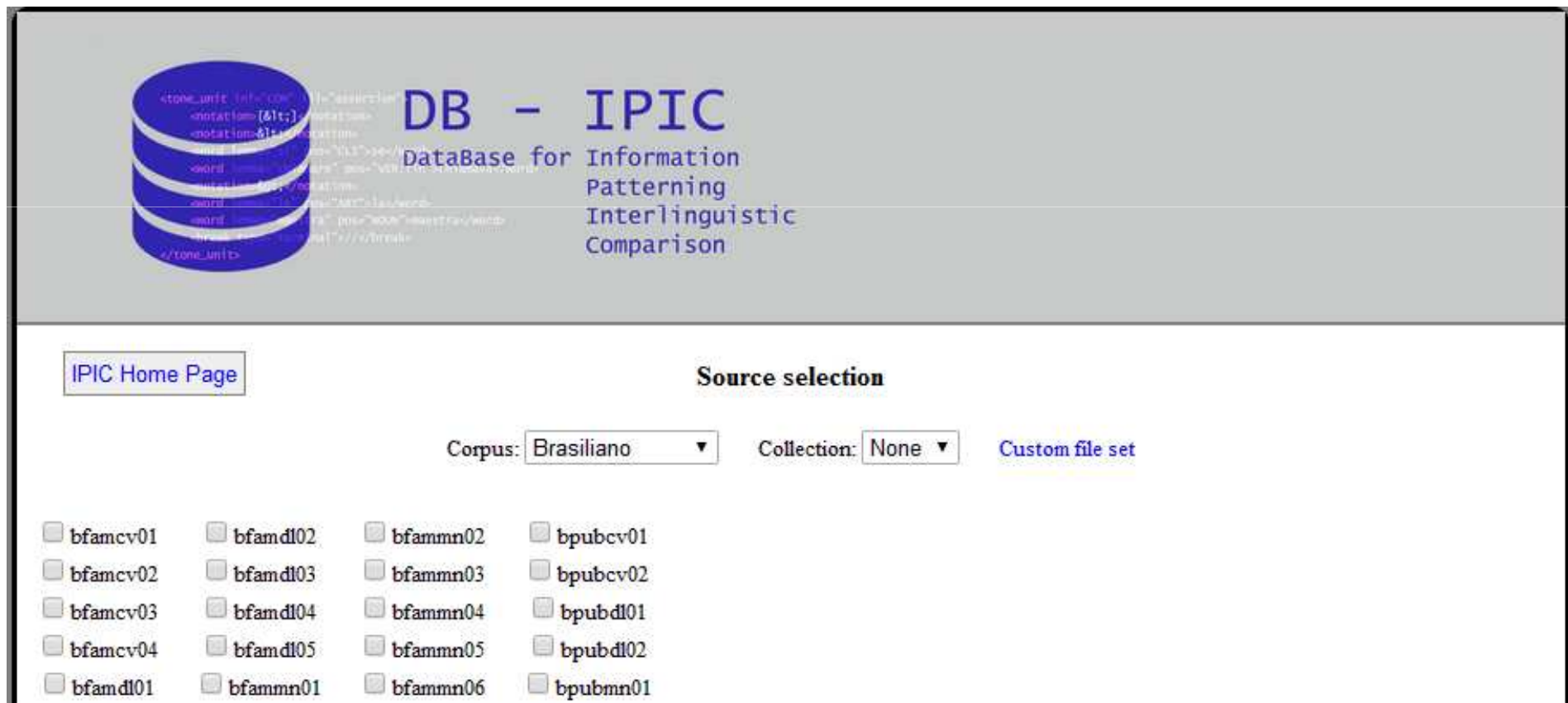
- Brazilian Portuguese spontaneous speech corpus, representative of the state of Minas Gerais diatopy. The corpus has been compiled following the same architecture and segmentation criteria adopted by the C-ORAL-ROM (Cresti & Moneglia, 2005) as well as its alignment software, the WinPitch (Martin, 2004).

- The informal part of the corpus, C-ORAL-BRASIL I, has been published (RASO & MELLO, 2012) and is available through a dvd as well as the Linguateca and CorpusEye platforms (will be available through TalkBank soon).

C-ORAL-BRASIL

- C-ORAL-BRASIL I comprises 139 informal spoken texts, 208,130 words, 21:08:52 hours of recording (6.1 GB wav files). The mean word number per text is 1,500. The recordings were carried out with high resolution, non-invasive wireless equipment, generally with clip-on, monodirectional microphones, and a mixer whenever there were more than two interactants.

- Texts are transcribed following the CHAT format, implemented for prosodic annotation. The main goal for the corpus architecture is the documentation of diaphasic variation in Brazilian Portuguese spontaneous speech (the broadest variation in speech acts); however, it is diastratically balanced as well.

Data analysed: minicorpus

- DB-IPIC: http://lablita.dit.unifi.it/app/dbipic/index.php

Minicorpus

- Representative sample

- Morphosyntactically annotated (PALAVRAS - Bick, 2012)

- Informationally annotated

- Ready access through DB-IPIC to transcription, information annotation, audio and aligned (text-sound XML) files

Accessing data: DB-IPIC

Linearized and patterned syntax

- **Linearized syntax:** constrained by the limits of an Information Unit, considered to be the proper locus for speech syntax to take place.

Eg.:

a gente não perdeu por causa disso não //=COM_r=

This is not why we have lost //=COM_r=

Linearized and patterned syntax

- **Patterned syntax:** pattern formed across the limit of a single infomation unit. Interpreted as a pragmatic pattern, not a 'true' syntactic relation

- Eg.:

< porque o Durepox > /=TOP= pelo menos jogava bola //=COM=
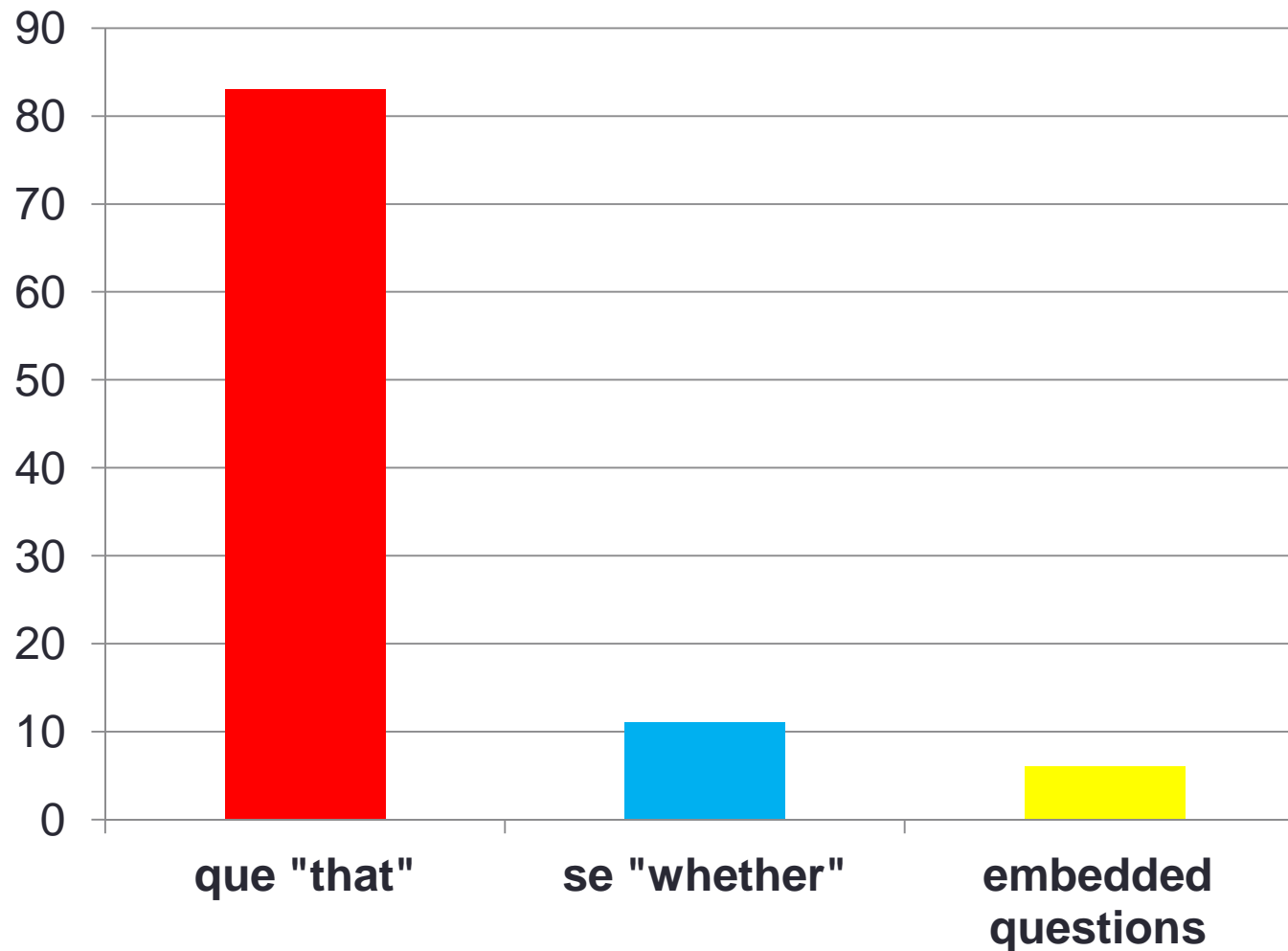
< because Durepox > /=TOP= at least played well //=COM=

Some focal results stemming from in-depth complement and adverbial clause studies

- Lexical indexes (complementizers, adverbial subordinators)

- Linearized vs. Patterned syntax
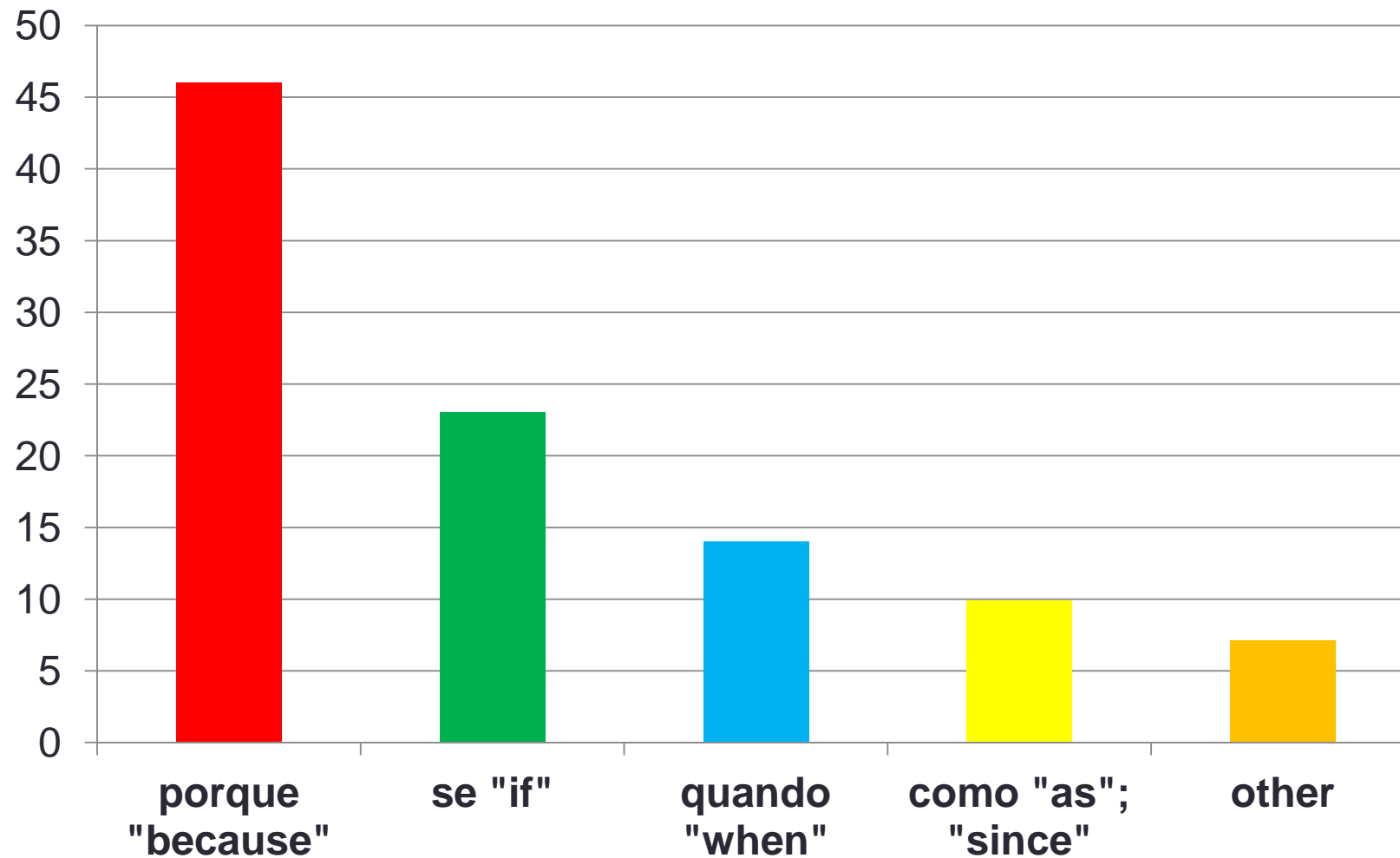
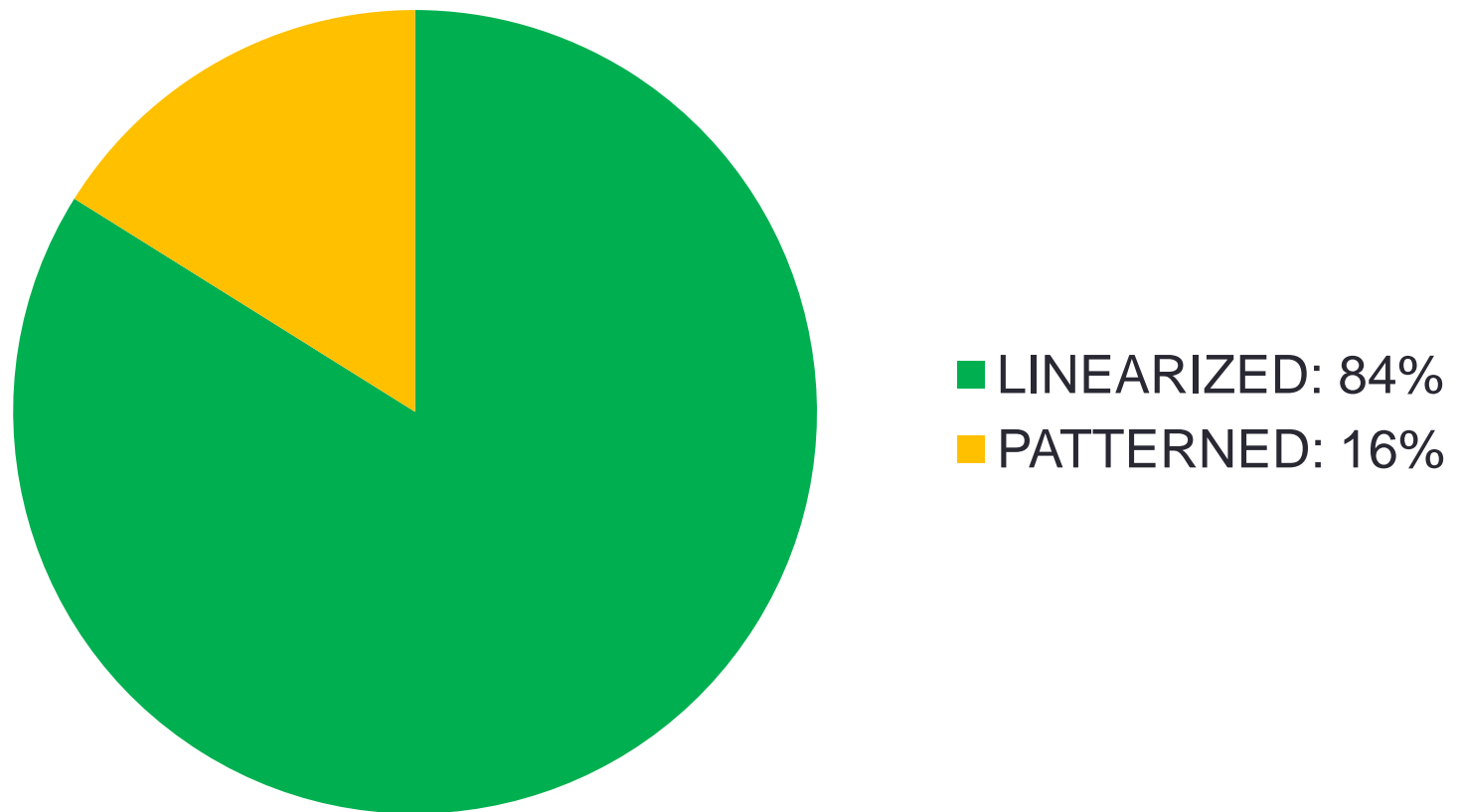# Minicorpus: Complementizers

Minicorpus: adverbial subordinators

|	
porque	'because'
se	"if"
quando	'when'
como	'since' / 'as'
caso	'if' (lit. 'in the case that')
já que	'since'
enquanto que	'while'
depois que	'after'
se bem que	'although'
apesar que	'although'
igual	'as'

Minicorpus: adverbial subordinators

Linearized vs. Patterned Syntax: Complement clauses

- LINEARIZED: 84%
- PATTERNED: 16%

Linearized vs. Patterned Syntax: Adverbial clauses

- LINEARIZED: 5.9%
- NON-LINEARIZED: 94.1%
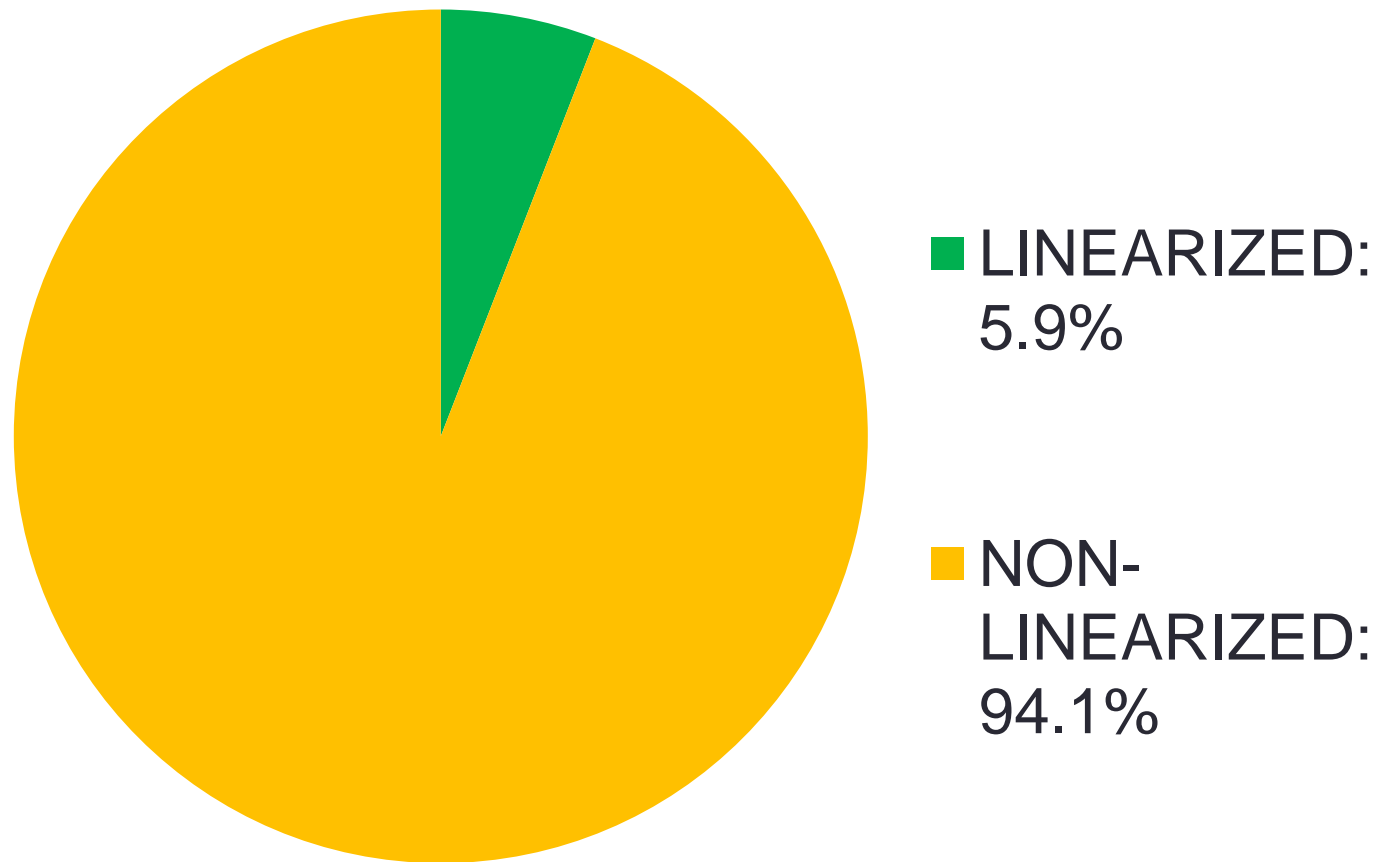
Coordination

Linearized coordination

1. a porta tá fechada **mas** cê pode abrir//=COM=
   *the door is closed **but** you can open it //=COM=*

2. aí ela pegou **e** falou que não queria aquilo não //=COM=
   then she said that she didn't want it //=COM=

3. agora imagina cê pega um balde **e** joga dentro //=COM=
   *now imagine that you take a bucket (full of water) **and** pour it all inside //*

Coordination

Patterned coordination

3. aí ele nos chamou /=COB= e falou /=INT= toma //=COM_r= 🔊
   then he called us / and said / take it //

4. ou é vinho bom caro /=CMM= ou é cerveja //=CMM= 🔊
   it's either expensive good wine / or it's beer //

5. na verdade eu queria levar as duas /=COM= né //=PHA= 🔊
   mas eu vou levar essa aqui //=COM=

   actually I wanted to buy both of them / you know //
   but I'm buying this one here //

Subordination

Linearized subordination

6. eu falei com Deus que eu queria uma criança saudável //=COM=
   I told God that I wanted a healthy child //

7. não falo porque acho muito pesado //=COM=
   I don't say it because I find it very heavy //

8. quando vê tá a pia cheia de trem //=COM=
   when he looks at it there's the sink full of stuff //

Subordination

Patterned subordination

9. **como** o Chub fala /=INT= es vão pegar os caras **que** /=i-COB= tipo /=PAR= tavam reclamando /=COB= e tal /=COB= es vão pegar /=COB= es vão pegar tentar fazer um negócio desse /=COB= e eu aposto **que** cê vai ver os caras **que** já conhecem a gente há mais tempo /=COB= tipo José [/1]=SCA= Zé Mourinho /=PAR= falando assim /=INT= não /=CMM_r= o /=SCA= campeonato d' ocês é bem melhor //=COM_r

*as Chub says / they're going to catch the guys **that** / I mean / were complaining / and so on / they're catching / they're catching them in order to try to make a deal like that / and I bet **that** you'll see the guys **that** already have known us for a longer time / like José [/1] Zé Mourinho / talking like that / no / the / championship of yours is pretty better //*

Subordination

Patterned subordination

10. quando ele chegou co bebê pra mim /=TOP= ele falou comigo assim /=COB= que /=DCT= era uma criança /=SCA= de [/1]=EMP= filha de mãe negra //=COM=

when he came with the baby for me / he told me so / that / she was a child / of [/1] child of black mother //

Insubordination

"the conventionalised main-clause use of what, on prima facie *grounds, appear to be formally subordinate clauses"*

(EVANS, 2007, p. 367)

Insubordination

11. essa aqui não fecha no meu pé //=COM=

apesar que meu pé tá meio sujo /=COM= né //=PHA=

então não fecha //=COM=

this one doesn't fit my foot //
although my foot is a little dirty / isn't it //
then it doesn't fit //

Insubordination

12. **se** cê quiser comprar as duas //=COM=
eu fico mais feliz /=COM= viu //=PHA=

if you want to buy both of them //
I'll be happier / you know //

Insubordination

13. A: mas se for só o dela então /=TOP= eu tiro //=COM=
    B: se for só o dela //=COM=
    B: se for mais gente numa situação pior /=TOP= aí nós vamos ter que pensar //=COM=

    A: but *if it's only hers then* / *I remove it* //
    B: *if it's only hers* //
    B: *if there are more people in a worst situation* / *then we'll have to think about it* //

Insubordination

14. A: eh /=PHA= se cê não tiver um carrinho que [/1]=SCA= que sobe
      aqui //=COM=
   B: uhn //=COM= é //=COM= isso não é muito bom //=COM=

   A: *well / if you don't have a good car that [/1] that climbs here //*
   B: *uhn // yeah // this is not very good //*

Insubordination

15. A: dá muito trabalho agora desativar //=COM=

    andando /=TOP= nũ dá não //=COM=

    B: mas e **se** ela explodir //=COM=

    A: explode não//=COM=

    A: *now it's really difficult to turn it off //*

    while driving / it doesn't work //=COM=

    B: *but and **if** it explodes //=COM=*

    A: *it doesn't explode //=COM=*

Challenges

The analysis of spoken BP syntax led us to find some issues that are worth being pursued because they challenge the conception that no syntactic relationship can exist across more than one IU or utterance.

- Syntax across IUs: Subordination

- Syntax across Utterances: Coordination

- Syntax across IUs/Utterances: the case of *porque*

Syntax across IU: Subordination

16. <u>que</u> eu /=INP= peguei ela com uma pessoa /=COB= eu pedi ele até pra arrumar pra mim /=PAR= <u>que</u> eu falei com ele assim que /=PAR= é um senhorzinho /=PAR= ele falou comigo que nũ ia [/3]=SCA= pa nós esperar /=COM= né //=PHA=

<u>that</u> I / took her through a man / I even asked him to arrange her for me / <u>that</u> I told him so that / he's kind of a gentleman / he told me that he wouldn't [/3] for us to wait / you know //

Syntax across Utterances: Coordination

17. na verdade eu queria levar as duas /=COM= né //=PHA=
    mas eu vou levar essa aqui //=COM=

    actually I wanted to buy both of them / you know //
    but I'm buying this one here //

18. pô /=EXP= Mailton /=ALL= eu nũ entendo muito de cobra não /=COB=
    mas essa história daí /=TOP= eu acho que quem matou o cara foi a
    mulher dele /=COM= hein //=PHA=
    e inventou tudo //=COM=

    god / Mailton / I really don't understand much about snakes / but that
    story / I think that who killed the guy was his wife / uh //
    and she invented it all //

Challenges: *porque* 'because'

19. agora /=INT_r= cê tá pagando porque ocê quer //=COM_r=

   now / you are paying because you want to //

20. não que isso fosse me confundir //=COM=
    porque eu nunca confundo Letras com Informática /=COB= não tem
    nem como //=COM=

   not that this could confuse me //
   because I never mistake Arts for Informatics / it's really impossible //

Challenges: *porque* 'because'

21. aí /=DCT= entrou lá dentro /=TOP= tem /=INT= uma incubadora //=COM=
    ocupada //=COM=
    porque tinha um neném que já tinha nascido //=COM=

    then / she entered there / there was / one incubator //
    occupied //
    because there was a baby that had already been born //

Conclusions

- In this talk we reported on BP spoken syntax within the L-AcT framework, in the light of Cresti (2014).

- Aims and results of studies and on-going investigations based on C-ORAL-BRASIL and DB-IPIC BP minicorpus were presented, in order to show the actual development of our research group on spoken syntax.

- We think that some of the properties of spoken syntax observed in the C-ORAL-BRASIL deserve further attention and will certainly gain from discussion and in-depth cross-linguistic comparisons (currently being undertaken: Italian, English).

References

AUSTIN, L. J. *How to do things with words*. Oxford: Oxford University Press, 1962.

BOSSAGLIA, G. 2014. Interface entre sintaxe e articulação informacional na fala espontânea: uma comparação baseada em corpus entre português brasileiro e italiano. *Caligrama*, v. 19, n. 2.

CRESTI, E. *Corpus di Italiano parlato*. Firenze: Accademia della Crusca, 2000.

CRESTI, E. 2014. Syntactic properties of spontaneous speech in the Language into Act Theory: Data on Italian complement and relative clauses. In Raso, T. & Mello, H. (eds), *Spoken Corpora and Linguistic Studies*. Amsterdam/Philadelphia: John Benjamins Publishing Company. 365-410.

CRESTI, E.; MONEGLIA, M. (Ed.). *C-ORAL-ROM*. Integrated Reference Corpora for Spoken Romance Languages. Amsterdam-Philadelphia: John Benjamins Publishing Company, 2005.

EVANS, N. Insubordination and its uses. In: NIKOLAEVA, I. (ed.) *Finiteness*: Theoretical and Empirical Foundations. Oxford: Oxford University Press, 2007, p. 366-431.

MARTIN, Ph. WinPitch Corpus, a Text to Speech Alignment Tool for Multimodal Corpora. In: *LREC*. 2004.

RASO, T.; MELLO, H. (Ed.). *C-ORAL-BRASIL I*. Corpus de referência do português brasileiro falado informal. Belo Horizonte: UFMG, 2012.

Thank you!

heliana.mello@gmail.com

giulia.bossaglia@gmail.com

tommaso.raso@gmail.com