

## El Paso Bilingual Corpus

Margot Vanhaverbeke, Renata Enghels, Ghent University  
 Anavictoria Dominguez, University of Texas at El Paso  
 María del Carmen Parafita Couto, Leiden University  
 Iva Ivanova, University of Texas at El Paso



From left: Margot Vanhaverbeke, Iva Ivanova, Lawrence DeBacker



Buildings from the middle-ages on the Graslei Quay, City of Ghent

## How using language in conversation innovates language

Renata Enghels, Robert Hartsuiker, Ghent University  
 Iva Ivanova, University of Texas at El Paso

### INTRODUCTION

Research on language contact faces theoretical and methodological challenges, because it is difficult to compare findings across studies that use different data collection methods

-> Need for publicly available and comparable bilingual corpora

### CODE-SWITCHING ON THE BORDER

**Code-switching** (aka Spanglish) is a typical, and fascinating mode of bilingual language use

- Obeys grammatical rules of both languages
- Increases expressive power
- Challenging to study -> few corpora of spontaneous code-switched conversations

### AIMS

**Collect, transcribe and annotate a corpus of spontaneous code-switched conversations**  
**Comparable to Bangor-Miami corpus**

### CORPUS DESCRIPTION

Recorded April 2022 – January 22-23 by **Anavictoria Dominguez** (Language and Communication Lab, UTEP)  
 Funded by Research Foundation – Flanders  
 42 spontaneous, informal conversations, 84 bilingual speakers living in El Paso (TX).  
 Each conversation lasted 30-45 minutes, 31 total hours of speech / 270,000 words of text

### METHODS

- Participants recruited from the UTEP community (flyers, social networks, friend-of-a-friend method)
- Adult (18+) bilinguals, highly proficient in Spanish and English, living in El Paso for >4 years
- Eligible participants invited to bring friend or relative -> informal conversations more likely
- Location of their choice (coffee shop, office, home, campus)
- Recorded with a Marantz PMD 661 MKII, some post-processed with Audacity.
- Participants were instructed to talk about any topic they liked (e.g., work-related events, family gatherings, hobbies, weekend plans)
- bilingual research assistant switched language to “cue” participants into this language mode
- Assistant then left for about 45 mins, so participants could talk spontaneously
- Participants filled out sociolinguistic background questionnaire
- Conversations transcribed and annotated in ELAN following established conventions

### CURRENT USES

**Use of diminutives:** little casa? hous-ita? (Margot Vanhaverbeke Ph.D. project)  
**Use of intensifiers:** muy, -ísimo, -azo, -on, hiper-, super- (Simon Claassen Ph.D. project)

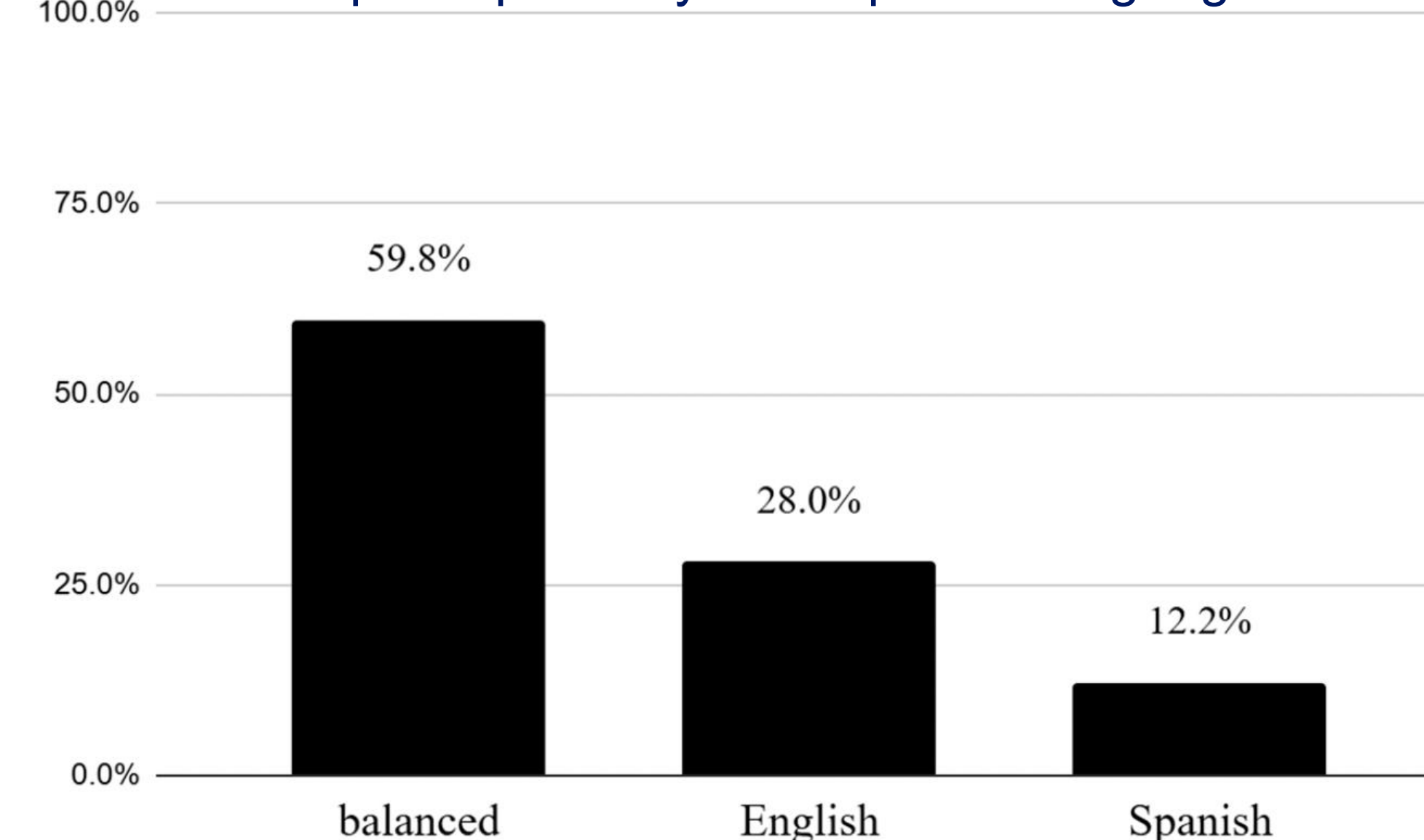
### EXAMPLES

DN7658: *Sí nos falta un minutito un minutito*  
 'Yeah we need one more minute, one little minute'  
 KV9880: *Vamos a decirles un secreto vamos va chiquito*  
 'We're gonna tell them a secret, let's tell them a little one'

IB6767: *Dijo que ahorita como a los cuarenta y cinco regresaba*  
 'She said that now like at forty five she would return'  
 GZ5538: *En efecto whatever*  
 'Actually'  
 IB6767: *Yo digo que le dejemos cinco minutos más*  
 'I say we leave it about five minutes more'

### LANGUAGES

Distribution of participants by self-reported language dominance



Dominant Language in the Conversations	#	%
Spanish	20	47.6
English	12	28.6
both	10	23.8
Total	42	100.0

### DIMINUTIVES

**Diminutives** are linguistic elements that indicate diminution of

- the size or quantity of the referent they are modifying (Ponsonnet, 2018)
- affective connotations, conveying the speaker's positive or negative attitude toward the referent (e.g., a lovely little gift vs. the hideous little beast)

*Gusta un cafecito? Tú tienes un problemita también ahí*

Their grammatical formation is different across languages

**English:** mostly analytic forms (*little house*); some synthetic (*fishie, famoso-ish*)  
**Spanish:** mostly synthetic form (*casita*); some analytic (*espacio chiquito*; *es un pequeño pocket*)

Diminutive formation strategies and language of the diminutive marker in the El Paso Bilingual Corpus and Bangor Miami Corpus.

Diminutive Strategy	El Paso Bilingual Corpus		Bangor Miami Corpus	
	n	%	n	%
synthetic	754	75.8	563	63.2
Spanish	745	98.8	531	94.3
English	9	1.2	32	5.7
analytic	241	24.2	328	36.8
Spanish	89	36.9	50	15.2
English	152	63.1	278	84.8
Total	995	100.0	891	100.0

### ALIGNMENT

**Linguistic alignment:** The tendency of speakers to repeat words, style-match or mirror an accent

- It characterizes conversation but also happens in non-interactive situations
- It may happen because you pick up the rhythm
- You want to help understanding
- You want to affiliate

Can you think of any words you have picked up from people you know?

### LANGUAGE CHANGE

Languages constantly evolve!

- Adapt to shift the changing needs of speakers, reflecting social, cognitive and cultural changes
- **Innovation** (first use of novel form) – more likely in looser social networks
- **Diffusion** (spread through the community) – more likely in tighter social networks
- The mechanisms of grammaticalization and analogy

Can you think of any words that appeared recently in English or Spanish?

### THE IDEA

Combine methods from linguistics and psychology to provide a comprehensive mechanistic account of how linguistic alignment drives and shapes language change

- Predictions informed from corpus analysis and apparent time method
- Controlled experiments
- Naturalistic environments
- Wide range of linguistic units

Are there any differences between how you speak, compared to older and younger people?

lexicon, clitics, possessives, pragmatic markers (e.g., *es que* [it's that], *sabes* [you know]), terms of address (e.g., *tío/tía* [lit. uncle], *hijo/hija* [son/daughter]), intensifiers (*super-[extremely]*, *puñetero* [freaking!]) and inchoative constructions (e.g., *ponerse a estudiar* [lit., put oneself to study], *lanzarse a hablar* [lit., launch oneself to speak]) (De Latte & Enghels, 2021; Enghels & Roels, 2024; Van Den Driessche & Enghels, 2025)

### THE PLAN

**Theme 1:** Identify the linguistic properties of expressions that are more versus less likely to spread through a community

**Theme 2:** Identify factors specifically related to the process of diffusion of linguistic expressions through a community

**Theme 3:** Identify the role of attentional focus and cognitive load for the process of diffusion of linguistic expressions through a community

### PRODUCTS

- R21 to NICHD in revision
- Postdoctoral project to FWO (Fien de Late), pending
- MSCA (Ioli Baroncini), in revision
- Theoretical proposal article to the high-impact *Behavioral and Brain Sciences*, partial drafts

