Multilingual Codeswitching in Letters

Gisella and Lena
Istanbul, 1951

Sept. 10th 1962

Milan tam Italyan avec grand dolce far niente et les gens qui sortent le soir pour le plaisir de flaner.

'Milan fully Italian with big sweet do nothing and the people who go out in the evening for the joy of strolling.'
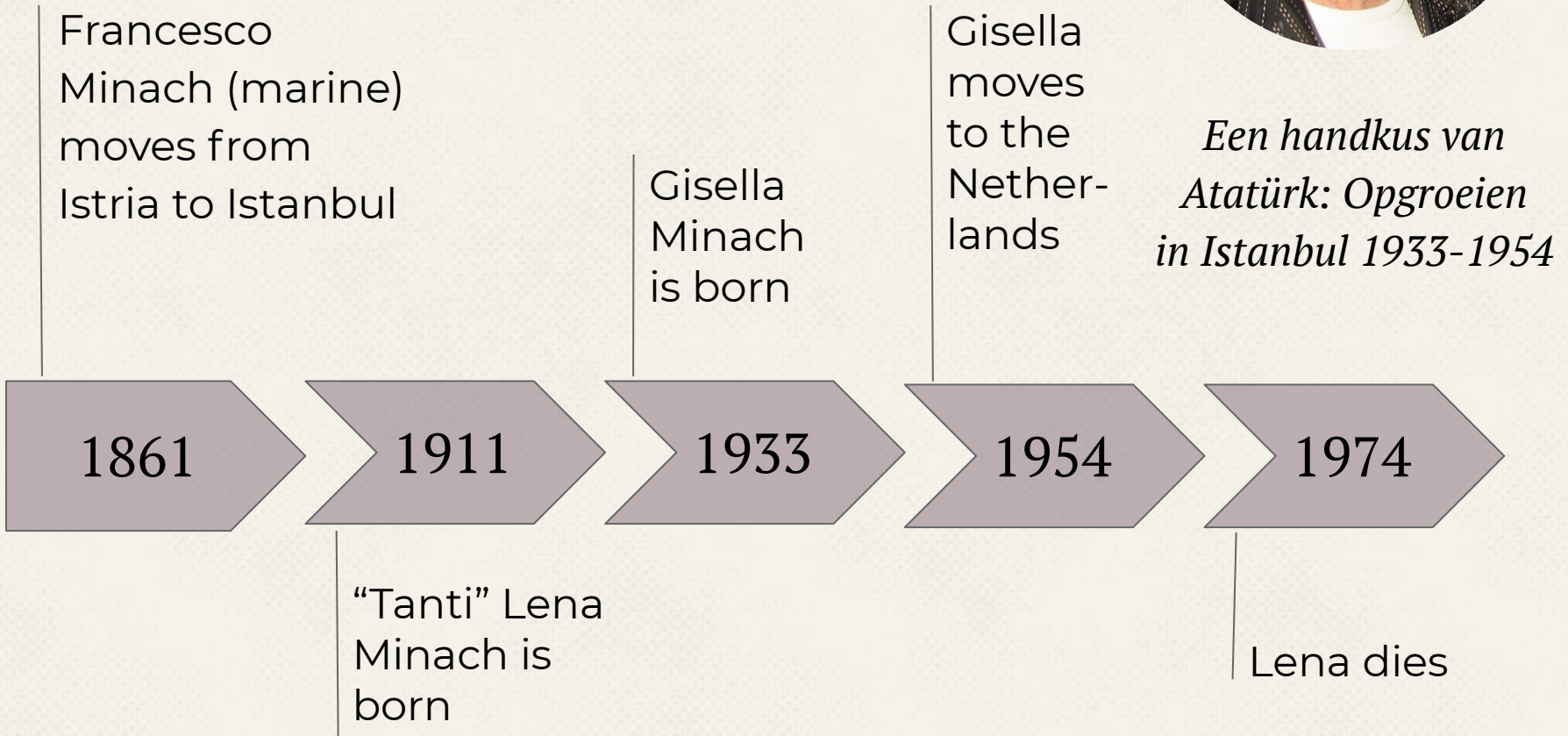
Minach family
Burgaz, 1968

Sept. 24th 1962

Manitchka avec entertainment de guest pendant
15 jours est definitely over-tired.
'Manitchka with entertainment of guests during
15 days is definitely over-tired.'

History of *Tanti's letters*

Francesco Minach (marine) moves from Istria to Istanbul

Gisella Minach is born

Gisella moves to the Netherlands

Een handkus van Atatürk: Opgroeien in Istanbul 1933-1954

| 1861 | 1911 | 1933 | 1954 | 1974 |

"Tanti" Lena Minach is born

Lena dies

333 letters

From 1958 to 1974

333 letters

From 1958 to 1974

225k words

About 230 per page

333 letters

From 1958 to 1974

225k words

About 230 per page
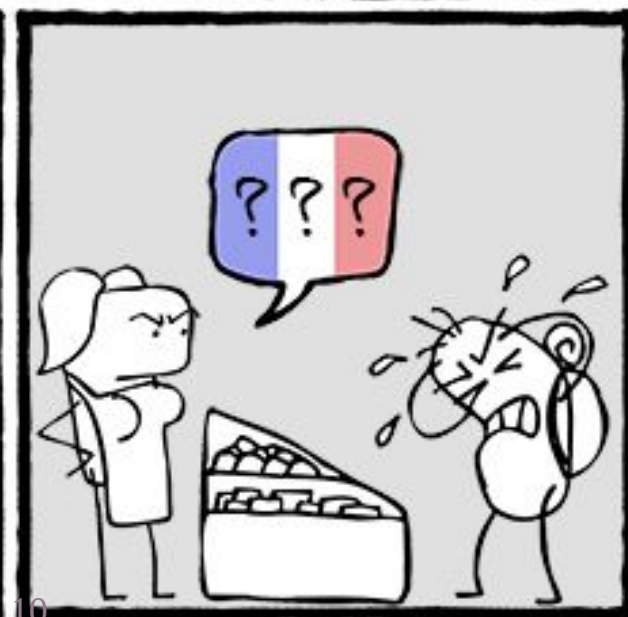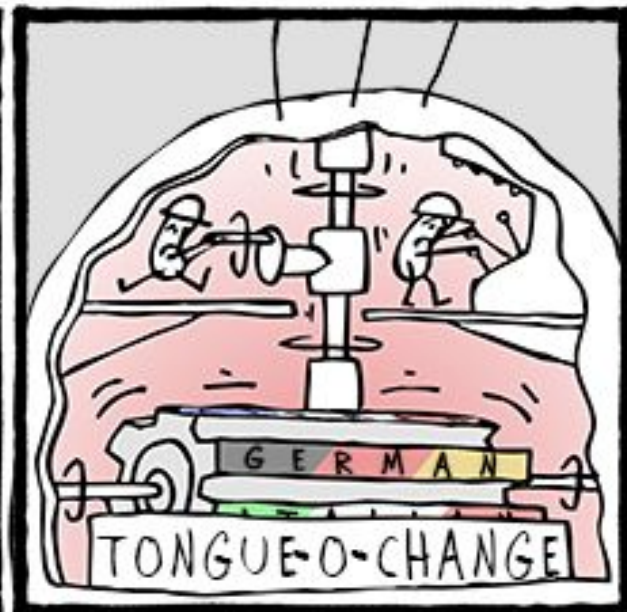
> 5 languages

Englisch, French, Turkish, Greek, Dutch

CONTENT

- Codeswitching
- Language Choice
- Language Identification
- Corpus Analysis
- Discussion

CODESWITCHING

Terminology and theory

"

The most **general definition of codeswitching** is this: the use of two language varieties in the same conversation.
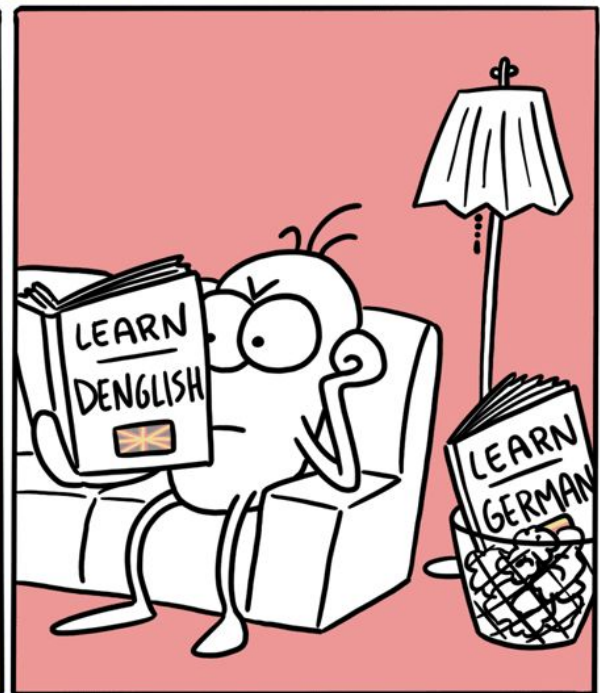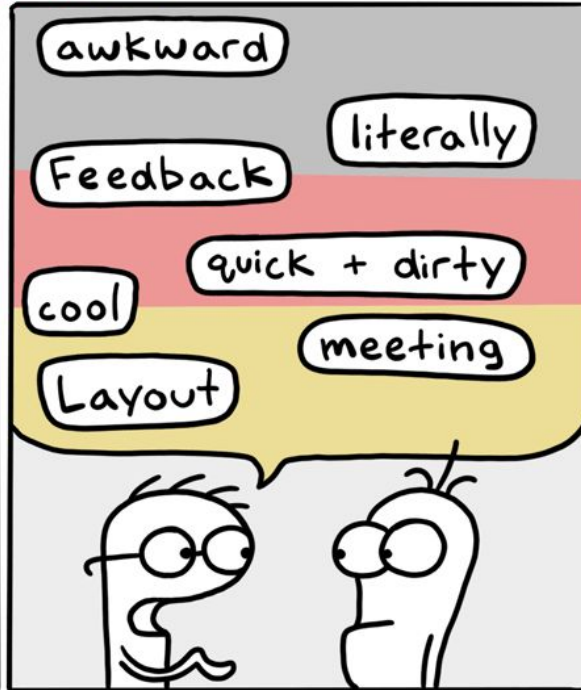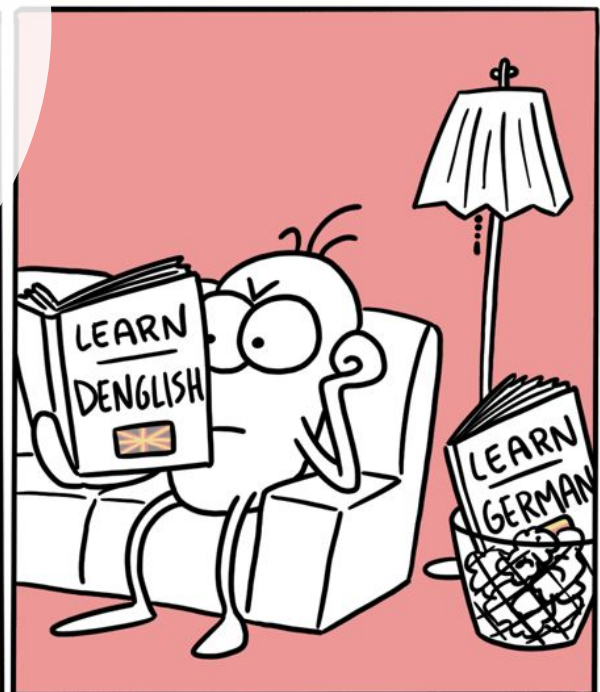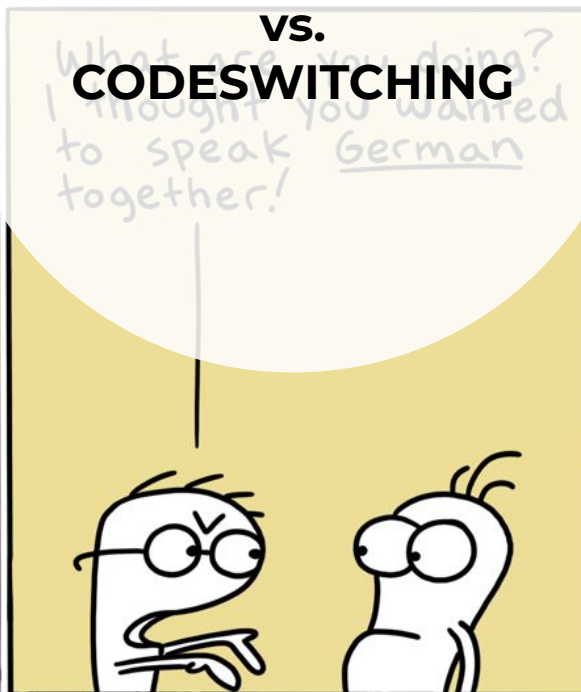
Myers-Scotton 2006

Code-Switching

Code-Mixing*

Borrowing

13

(Sinha, Takur 2005)

© 2018 Malachi Ray Rempen   www.itchyfeetcomic.com

BORROWING vs. CODESWITCHING

Monolingual language mode

Bilingual (or multi~) language mode

L*a*

L*b*

L*a*

L*b*

L*c*

16
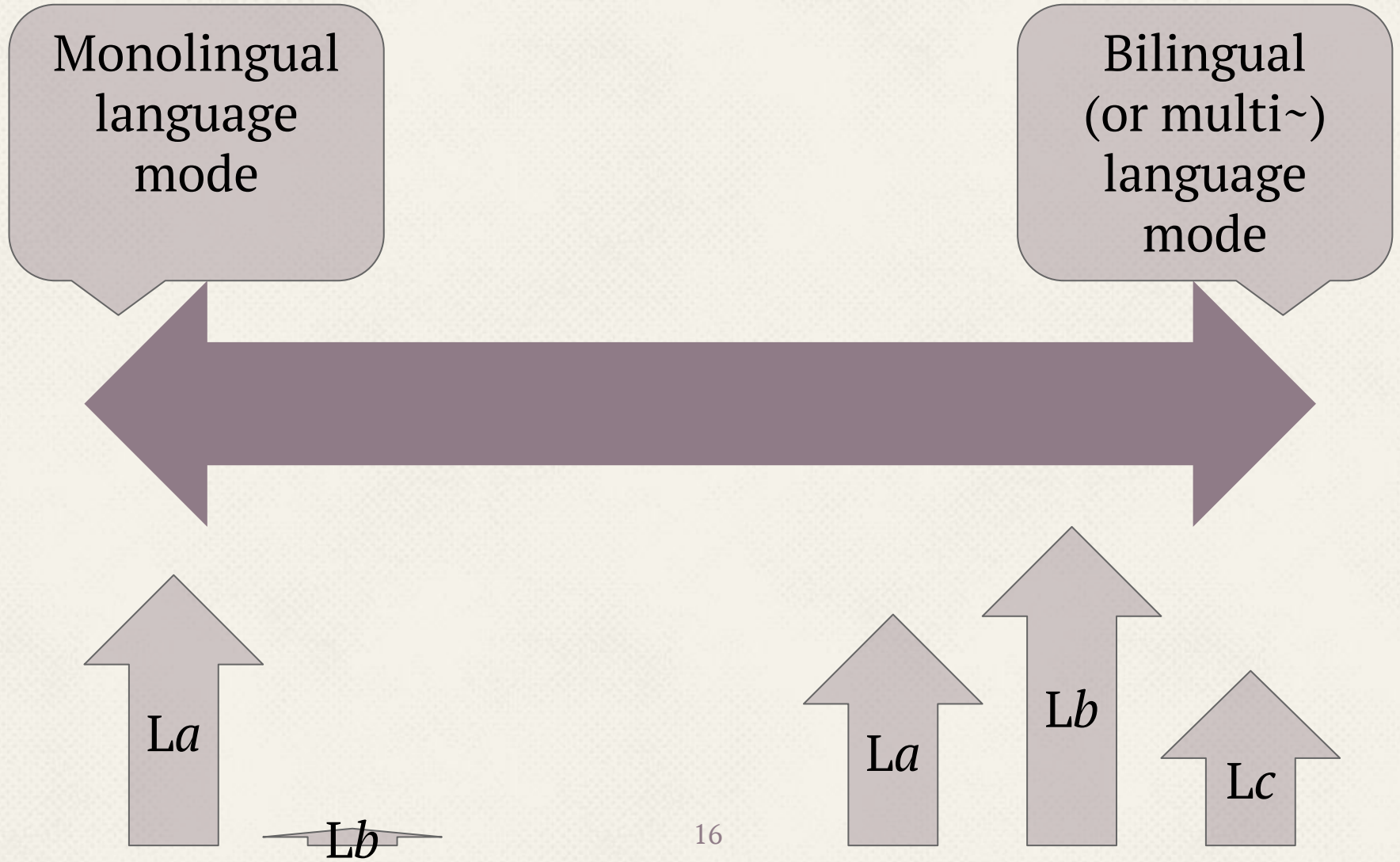
LANGUAGE CHOICE

Theory and application

Participants

Proficiency of both

Language history

Attitude towards language or group

(Grosjean 2010)

Participants

Always Gisella

Who is being talked about

Quotes

Participants

Who is being talked about

Quotes

Situation

Location

Formality

Presence of monolinguals

(Grosjean 2010)

Participants

Who is being talked about

Quotes

Situation

Location: at home/in Istanbul

Presence of monolinguals

Participants

Who is being talked about

Quotes

Le *mot juste*

Frequent lexical items in one language

Rare in the other language(s)

Specific connotation

(Grosjean 2010)

Participants

Who is being talked about

Quotes

Le *mot juste*

Frequent lexical items in one language

Rare in the other language(s)

Specific connotation

Content of discourse

Cooking

Housework

Shopping

Participants

Who is being talked about

Quotes

Le *mot juste*

Frequent lexical items in one language

Rare in the other language(s)

Specific connotation

Content of discourse

Cooking

Housework

Shopping

Function of the interaction

Exclude someone

Request, command, etc.

(Grosjean 2010)

LANGUAGE CHOICE

Participants

Who is being talked about

Quotes

Le *mot juste*

Frequent lexical items in one language

Rare in the other language(s)

Specific connotation

Content of discourse

Cooking

Housework

Shopping

Function of the interaction

Request to write/send something

Express affection

Domain- or semantic specificity

(Backus 2001, Grosjean 2010)

Motivation for CS

Domain- or semantic specificity

(Backus 2001, Grosjean 2010)

Individual preference, creativity

(Ben-Rafael 2001, Myers-Scotton 2006)

Motivation for CS

Domain- or semantic specificity

(Backus 2001, Grosjean 2010)

Individual preference, creativity

(Ben-Rafael 2001, Myers-Scotton 2006)
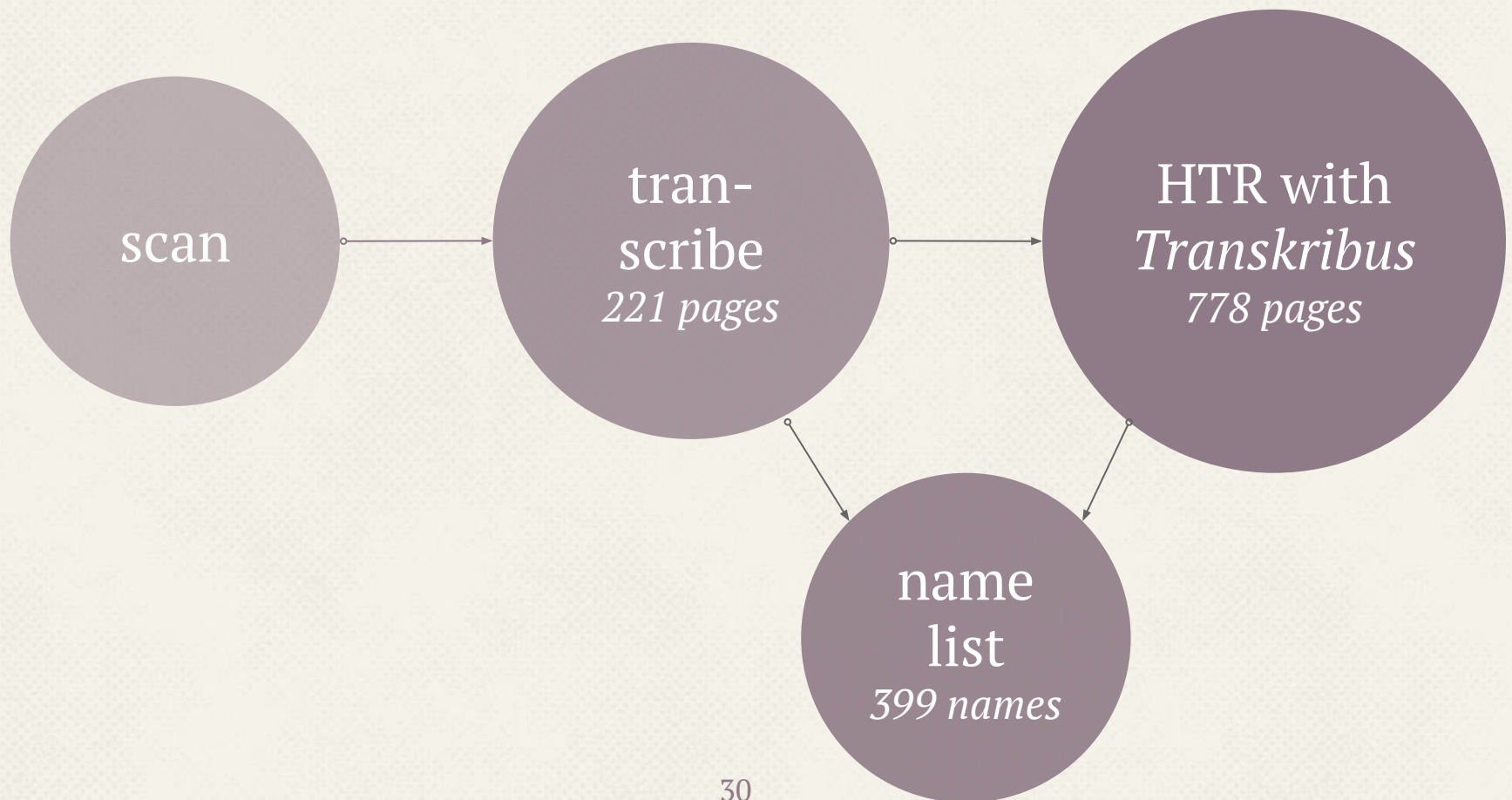
Cognates as triggers

(Broersma 2009)

Motivation for CS

LANGUAGE IDENTIFICATION

Preparation of the corpus

scan

tran-
scribe
221 pages

HTR with
Transkribus
778 pages

name
list
399 names

CORPUS BUILDING

Unfortunately j'ai fait un peu de pisbojazlık. pisboğazlık. The porters wofe brouglhet porter's wife brought me some lovely nuts sent to her from her village. I took two which were just delicions. delicious. Je me retiens pour ne pas continuer. Wweight Weight goes down with the die diet and then flies up to Holt 140lb again at the week-end, so conte coute que coute il faut se retenie retenir pendant la semaine. Tmorrow Tomorrow
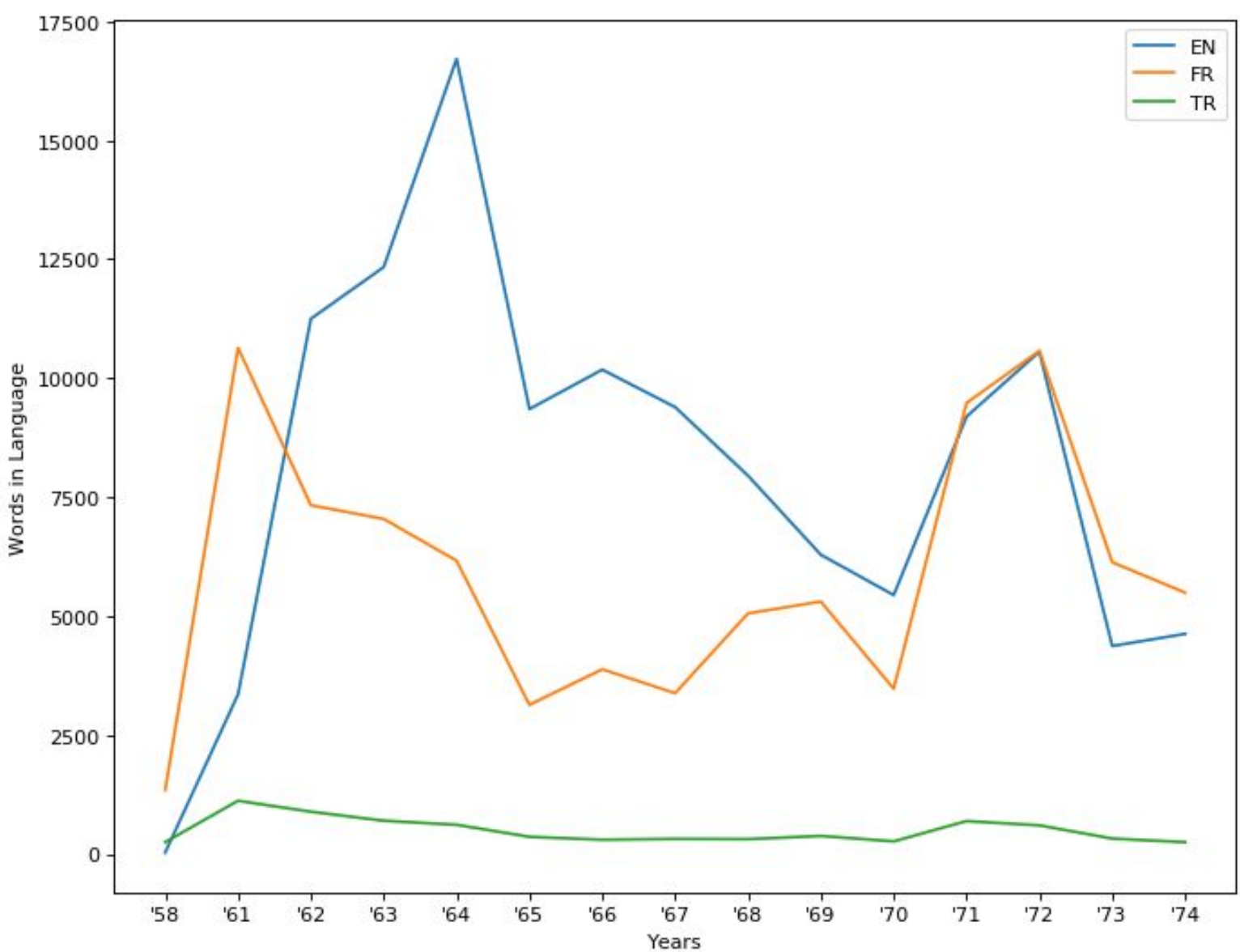
LID in *Tanti's Letters*

○ Preprocessing and NE (name list)

○ LangDetect and Langid

  ◦ Manual annotation of Greek, Dutch, German, Spanish, Italian, Minach terms

  ◦ Set languages to EN, FR (TR)

  ◦ Word n-grams

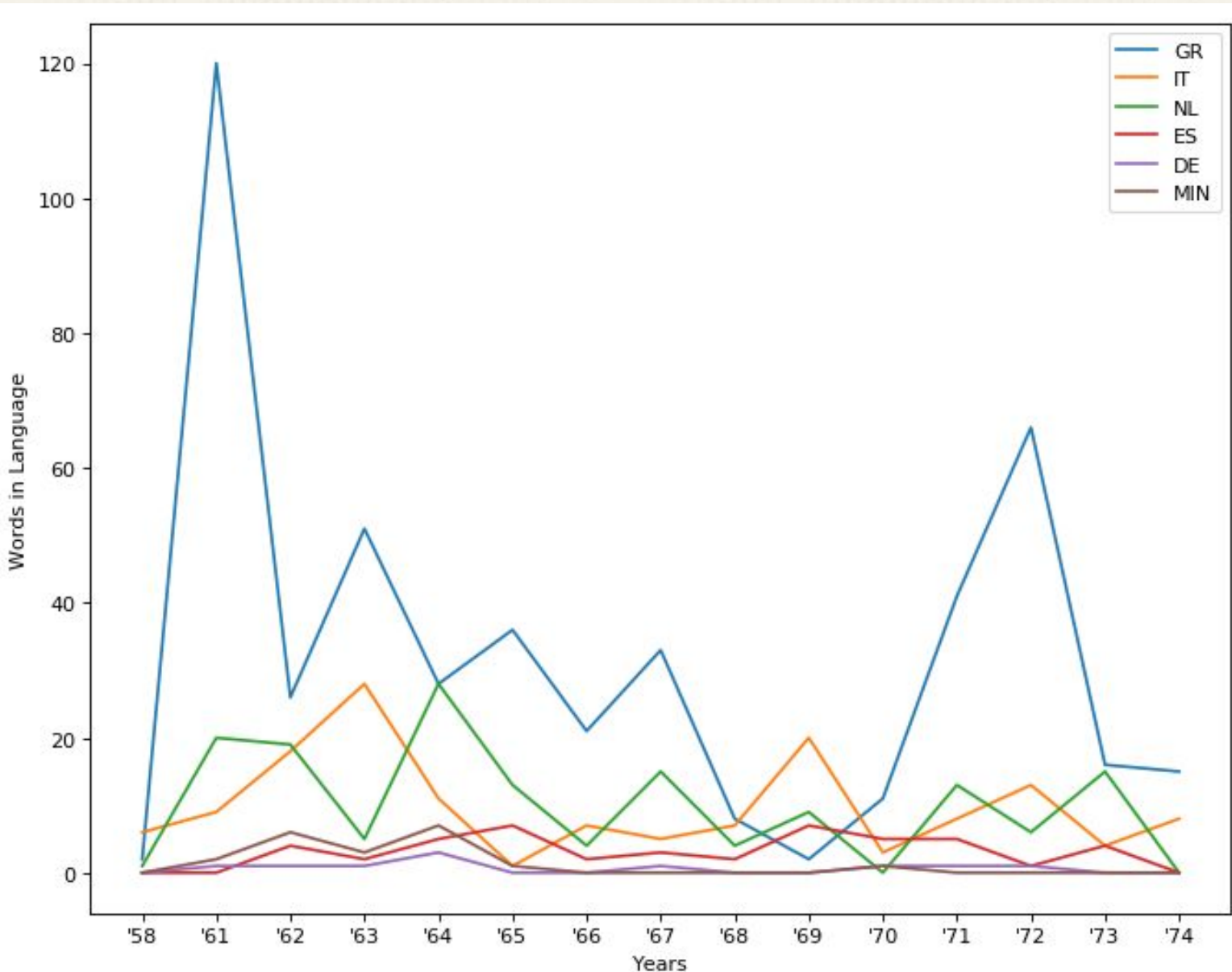 ◦ Classifier combination (highest confidence)
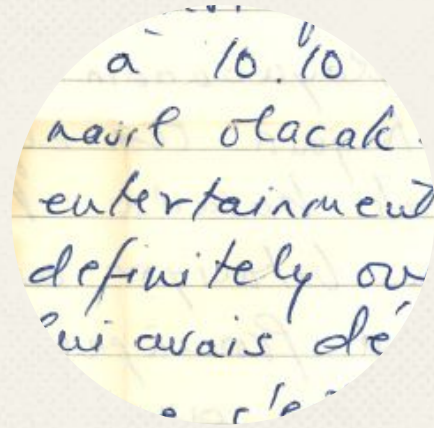
CORPUS ANALYSIS

Results

OBSERVATIONS

Languages	Words	Mean Length	Max Length
English	121 035	6.35	41
French	88 422	6.07	42
Turkish	7 342	1.59	9
Greek	476	3.26	23
Dutch	152	1.52	8
Italian	148	1.50	9
Spanish	47	1.04	3
German	10	1	1
Minach	20	1.05	2
total	217 652	5.64	42

Methods
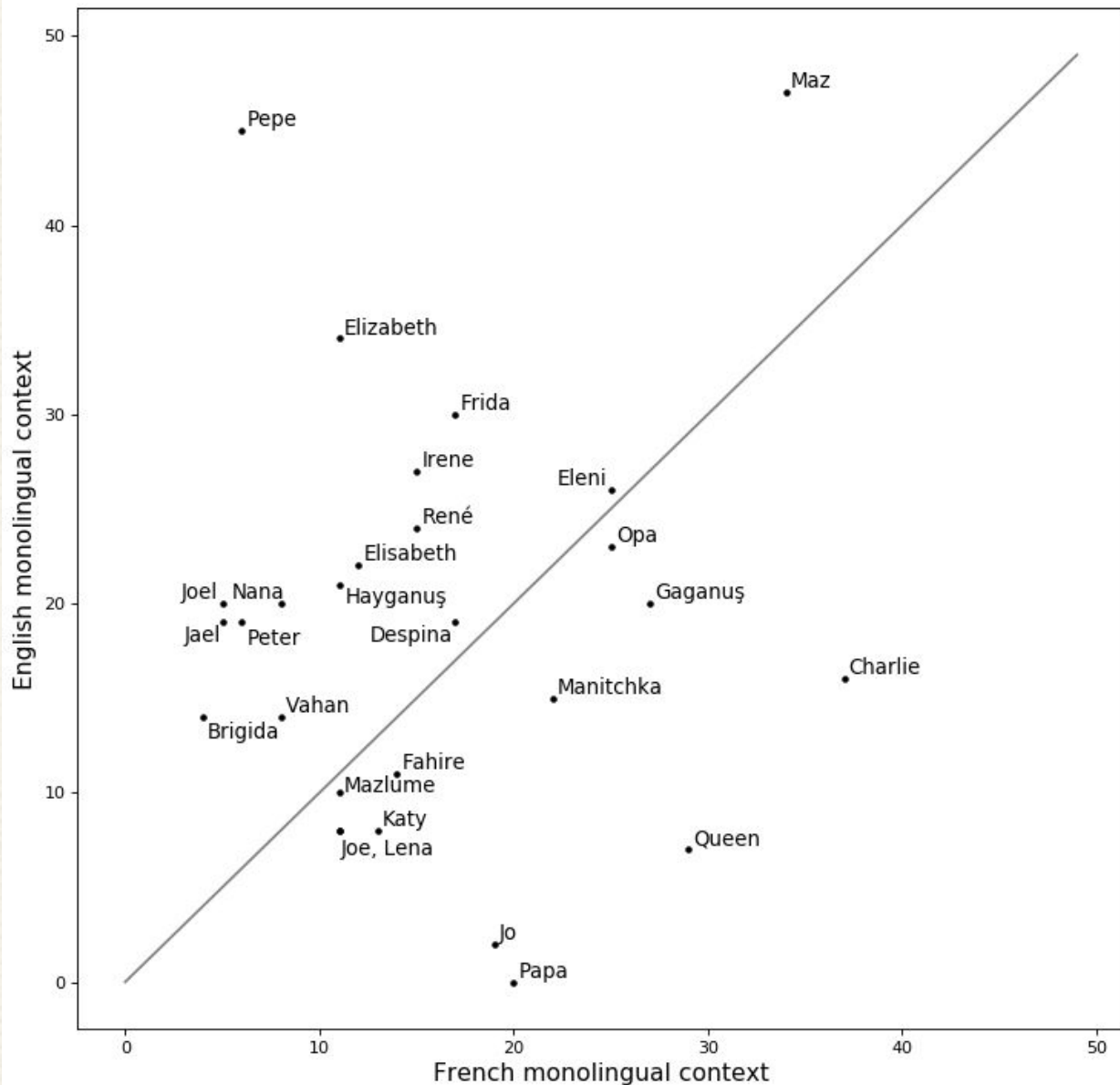
▷ visual inspection of names per dominant context language

▷ linear mixed effects regression model with word fields (*affection, request, cooking, work, etc.*) and dominant context language
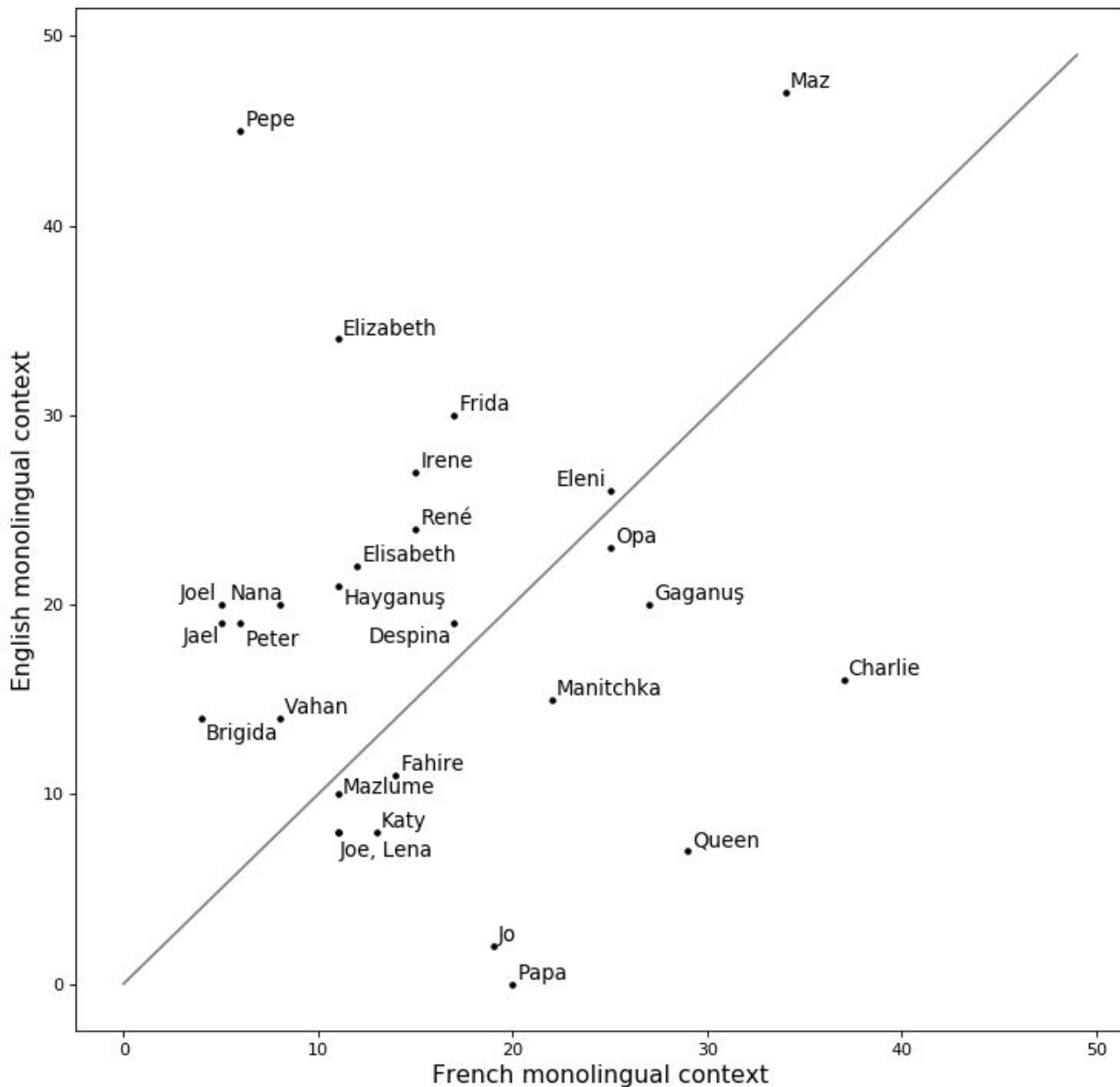
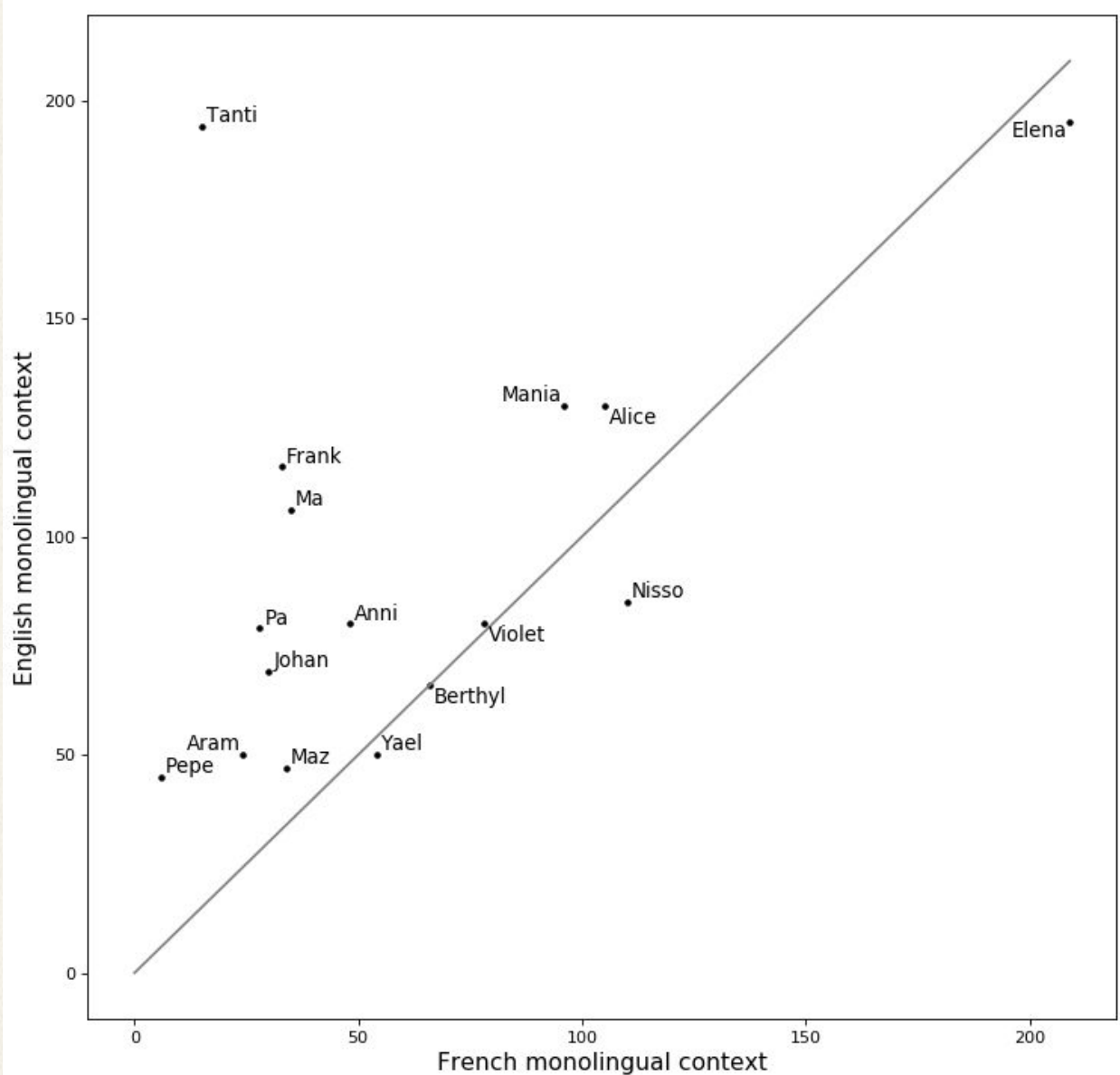▷ inspection of the use of minor languages

Tendencies:

○ nicknames FR

Tendencies:

- nicknames FR
- locals EN
- Europeans EN

Tendencies:

- nicknames FR
- locals EN
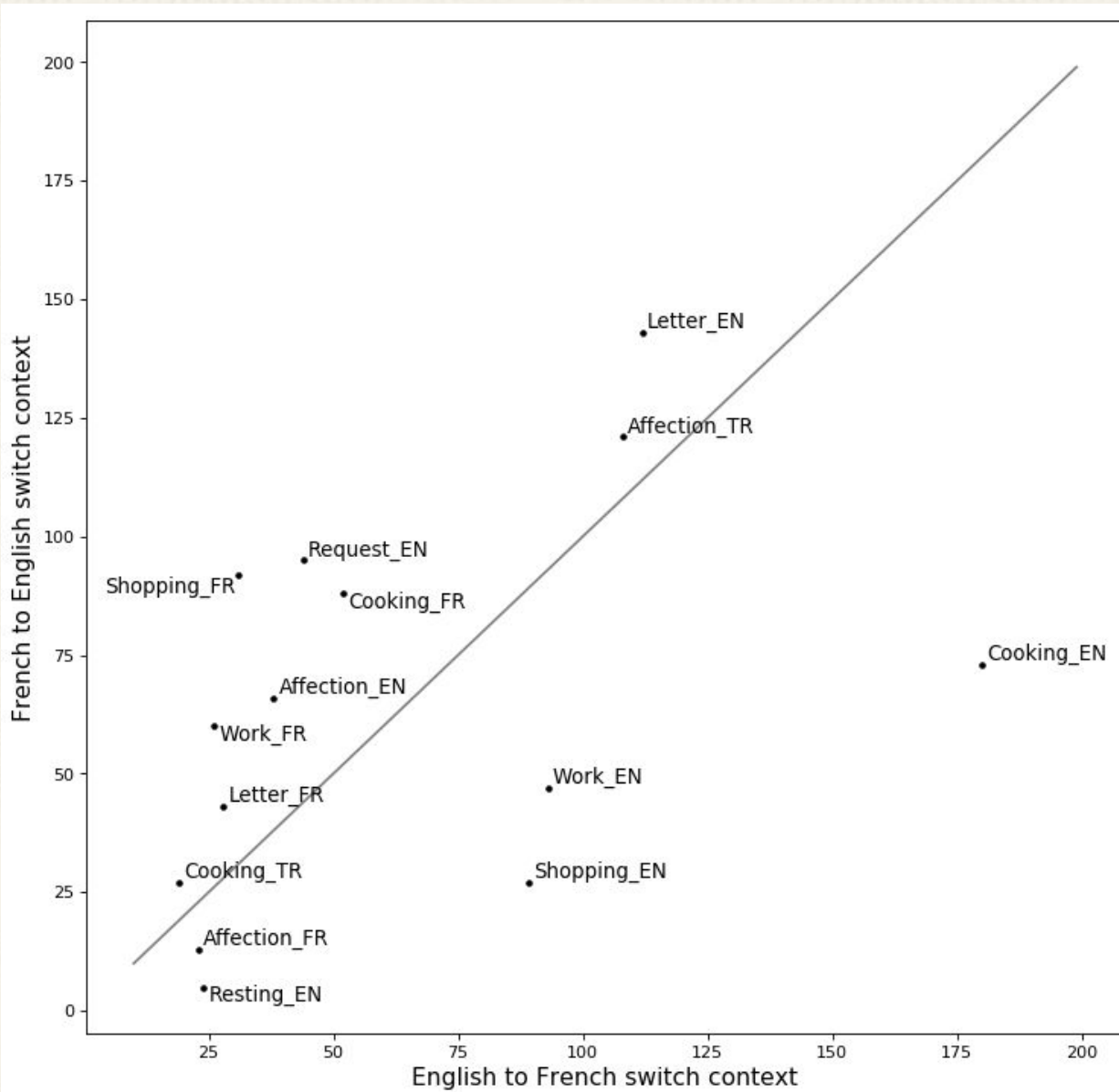- Europeans EN
- most frequent EN & FR

What influences language choice?

Participants

- ○ either dominant with one language or indifferent
- ○ only motivation for language choice, not CS

What influences language choice?

Content of the discourse &
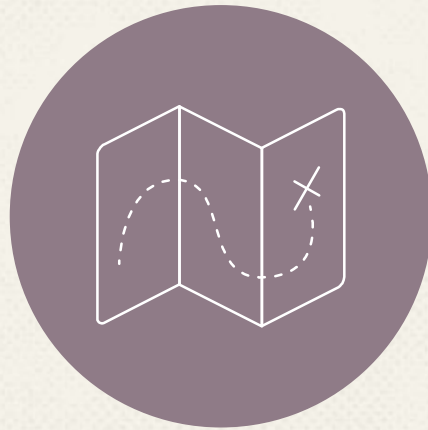
Function of the interaction

- ○ domain- and culture-specific concepts or objects
- ○ personal preference
- ○ preference for monolingual passages

What influences language choice?

🎯 *Le mot juste*

- connotations, lack of translational equivalents
- culture-specific concepts or objects
- personal preference (*cervello* for *cerveau*)
- can trigger longer CS

DISCUSSION

Limitations and outlook

Limitations

Imbalanced corpus
○ language use of major and minor languages not comparable
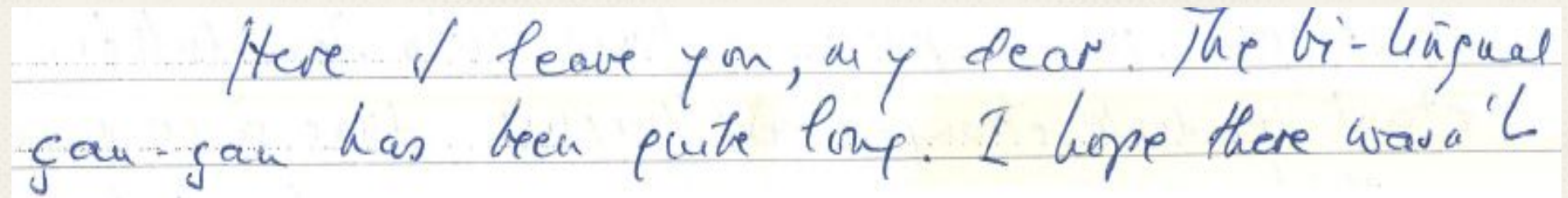
One author
○ consistent, yet specific language use

Standardized method missing
○ manual inspection
○ multitude of factors

Outlook

- CS in writing vs. speech

- incorporate findings in automatic LID/NLP methods

"Here I leave you, my dear. The bi-lingual çan-çan has been quite long. [...]" May 22nd 1962

Franka Weber
M.A. Speech and Language Processing
University of Konstanz
franka.weber@uni-konstanz.de

Bibliography

Backus, Ad 2001. The role of semantic specificity in insertional codeswitching: Evidence from Dutch-Turkish. In Rodolfo Jacobson (ed.), *Trends in Linguistics: Codeswitching Worldwide II,* 125-154. Berlin: Mouton de Gruyter.

Ben-Rafael, Miriam 2001. Codeswitching in the language of immigrants: The case of Franbreu. In Rodolfo Jacobson (ed.), *Trends in Linguistics: Codeswitching Worldwide II,* 251-307. Berlin: Mouton de Gruyter.

Broersma, Mirjam. 2009. Triggered codeswitching between cognate languages. *Bilingualism: Language and Cognition* 12 (4), 447-462. DOI: 10.1017/S1366728909990204.

Grosjean, François. 2010. *Bilingual: Life and reality*. London: Harvard University.

Lui, Marco and Timothy **Baldwin**. 2012. langid.py: An off-the-shelf language identification tool. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012) Demo Session, 25-30. Jeju, Republic of Korea. https://www.cs.cmu.edu/~ralf/langid.html.

Minach, Gisella and Pieter Stokvis. 2011. *Een handkus van Atatürk: Opgroeien in Istanbul 1933 - 1954*. Uitgeverij Tilia Levis.

Myers-Scotton, Carol. 2006. *Multiple voices: An introduction to bilingualism*. Malden: Blackwell.

Nakatani, Shuyo. 2010. Language detection library. Slides. http://www.slideshare.net/shuyo/ languagedetection-library-for-java/. https://pypi.org/project/langdetect/.

Sinha, R. Mahesh K. and Anil **Thakur**. 2005. Machine translation of bi-lingual Hindi-English (Hinglish) text. *Conference Proceedings: the tenth Machine Translation Summit*, 149-156.

Transkribus. https://transkribus.eu/Transkribus/.