Designing a new platform for AD analysis: the VIW project

Anna Matamala, Marta Villegas

Universitat Autònoma de Barcelona

anna.matamala@uab.cat

CILC 2016, 8th International Conference on Corpus Linguistics

Málaga, 02-04/03/2016.

Funding: FFI2015-62522-ERC

2014SGR0027

CC BY NC

Overview

- Audio description (AD) research
- The VIW (Visuals Into Words) project
- Compiling the corpus: strategies
- Corpus description
- Exploitation needs and annotation levels
- Sample analyses
- Next steps

AD research

- AD, intersemiotic translation (visuals into words)

- AD research: case-studies, descriptive research on various modalities

Corpus research in AD

- Small corpus (Piety 2004)
- TIWO (Salway 2007)
- TRACCE (Jiménez Hurtado et al. 2010)
- Pear Tree Project (Mazur and Kruger 2012), inspired by Chafe (1980)

The VIW project

- Funded by the Spanish Gov. under “Europa Excelencia” funding scheme (Oct 15-Sept 16)
- Rationale
  - one small stimulus, expanding research
  - open access
  - multilingual
  - multimodal

Compiling the corpus: strategies

- Authentic materials/materials created ad hoc?
- Copyright issues: open access
- Multimodal: audio and video, and text
- Multilingual: English, Spanish, and Catalan (possibility to expand it even more)

Corpus description: the short film

- Commission a short film to a film director
- Create copyright agreement

- Literature and guidelines' analysis on AD challenges:
  - characters and action (including gestures and facial expressions)
  - spatio-temporal settings
  - film language
  - sound effects and silence
  - text on screen
  - intertextual references.

Instructions to the film director

- Standard narrative
- Various actions
- At least four characters, speaking in English except for one
- At least three different spatio-temporal settings
- Incorporate some text on screen and credits.
- Include one non-identifiable sound
- Silent passages for artistic purposes
- 12-15 minutes long
- Awareness about the AD.

What happens while---

- Result is a 14-minute film directed by Núria Nia (trailer & open access link)

- Dubbed into Catalan and into Spanish (professional dubbing studio)

Corpus description: the ADs

- 10 ADs in Catalan , 10 in English and 10 in Spanish

- Service providers contacted with the same instructions, paid by the project
- Copyright issues solved through an agreement

The corpus: textual data

AD	Versions	Words
English	10	6,814
Catalan	10	5,702
Spanish	10	5,292
Global:	30	17,808

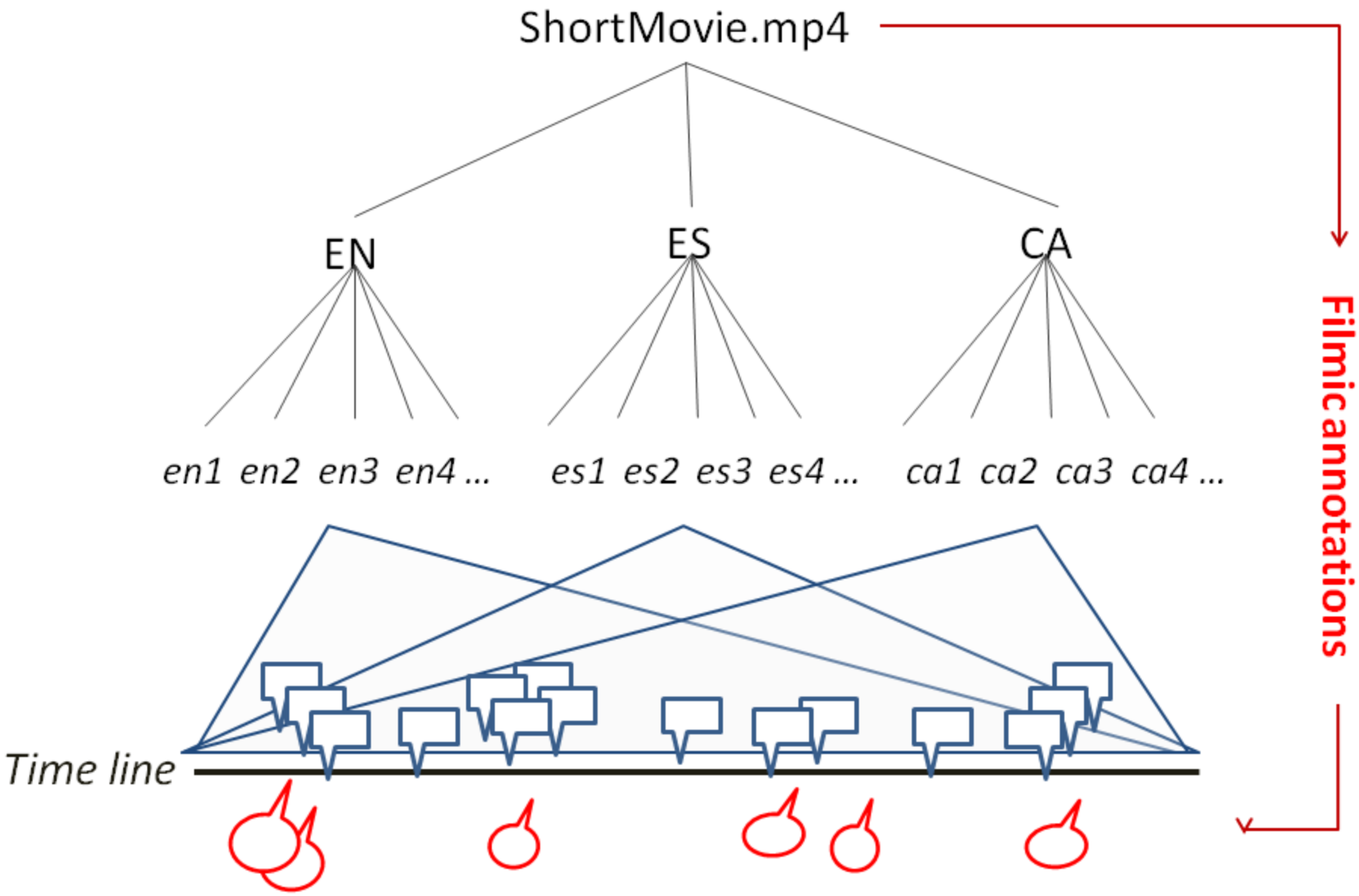
Exploitation needs

- Comparative analyses
  - Monolingual
  - Multilingual
- Linguistic elements: frequency, POS, semantic?
- AD distribution along the timeline
- AD elements: how is X rendered in the AD? Divergences and commonalities

Annotation levels

- ELAN (Sloetjes and Wittenburg 2008)

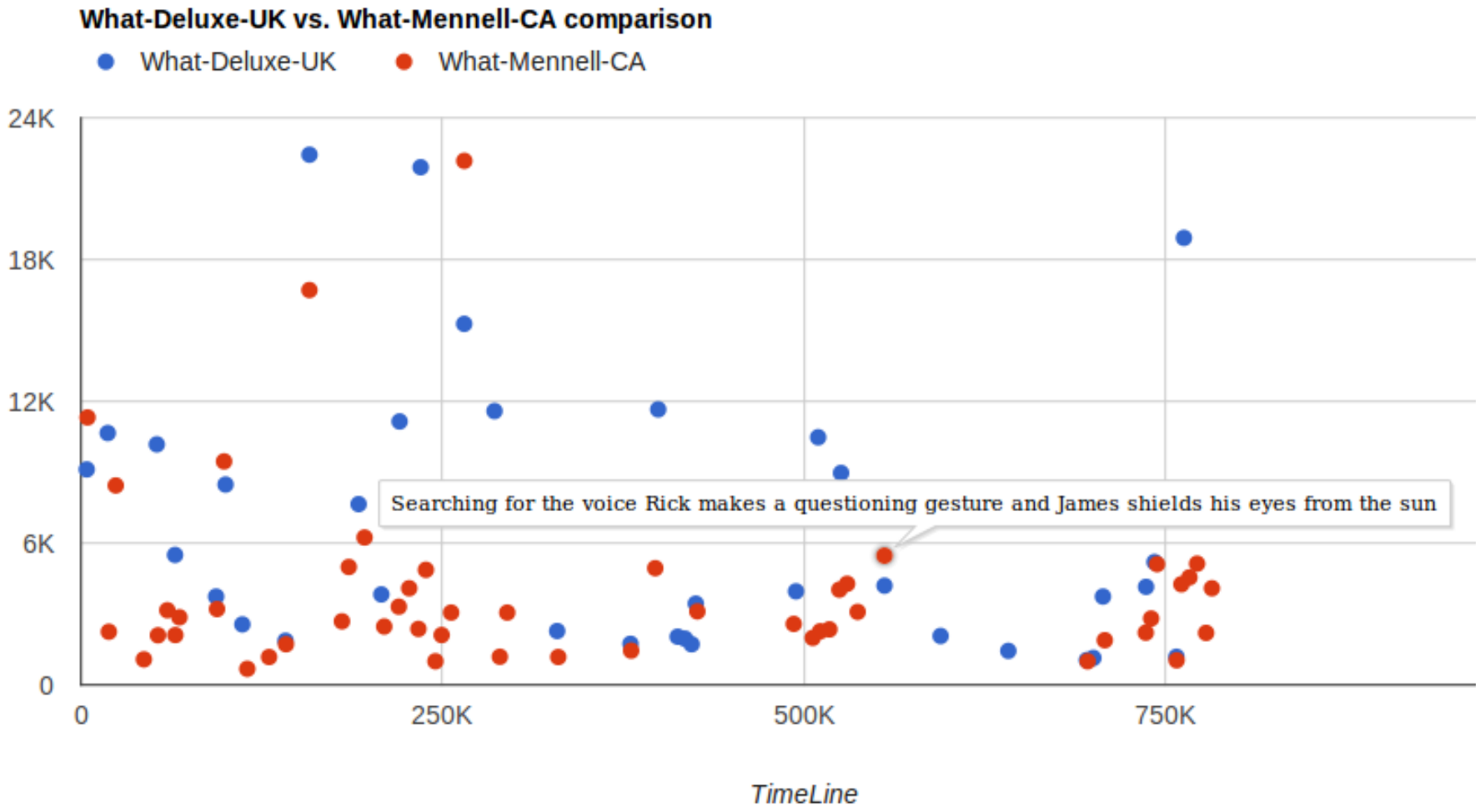
- Linguistic annotations: sentences, chunks, tokens, part of speech, lemma, and semantic annotations.
- Cinematic annotations: text, sound, and camera.


ShortMovie.mp4
EN
ES
CA
en1 en2 en3 en4 …
es1 es2 es3 es4 …
ca1 ca2 ca3 ca4 …
Time line
Filmicannotations

Data visualisation

- Web application providing access to:
  - XML annotation file conforming to EAF format, visualized in HTML
  - some graphical visualizations using Google Charts API

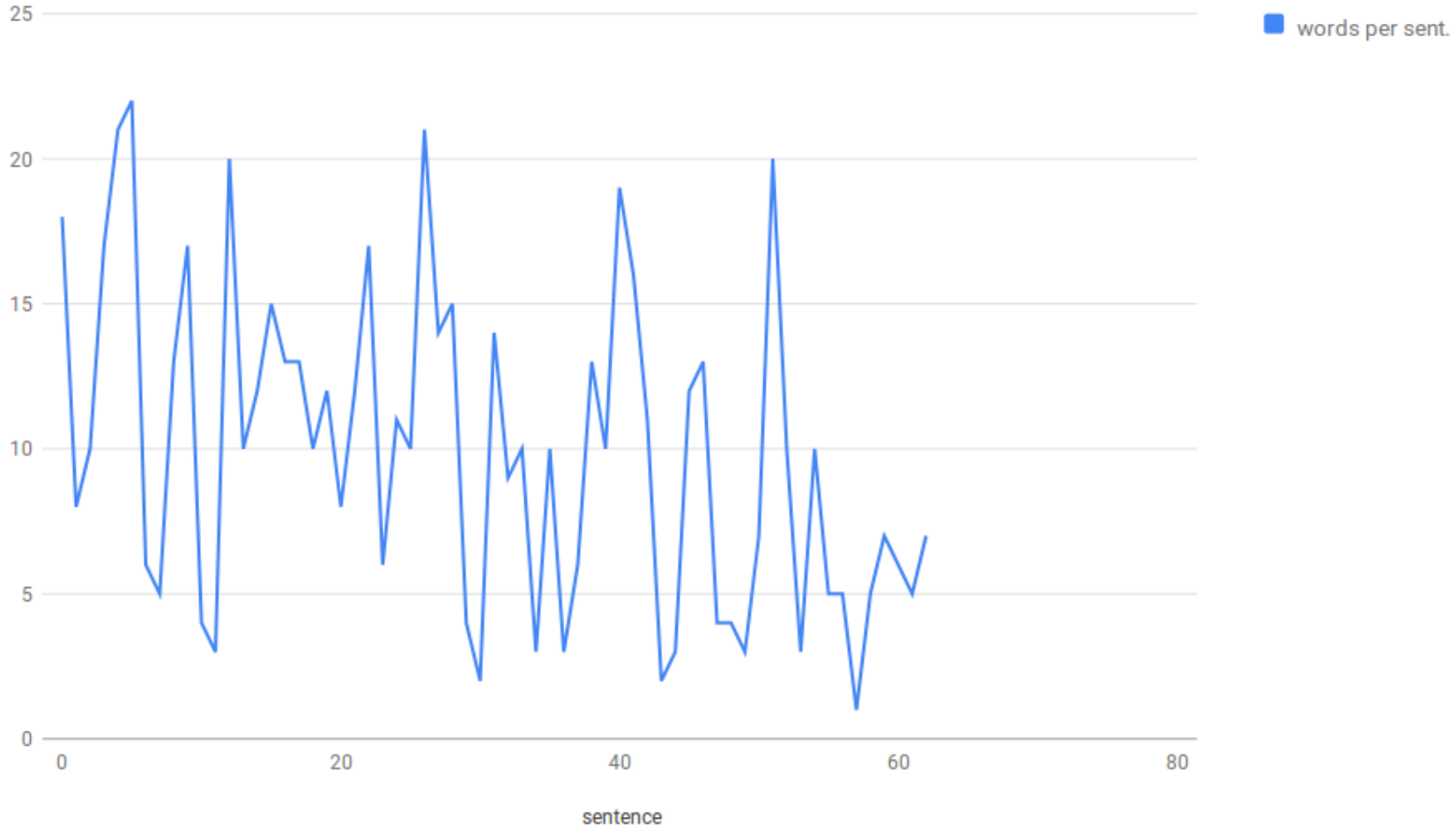
Sample analyses: comparing two ADs



Sample analyses: numbers of words

(2) Number of words per senteces:

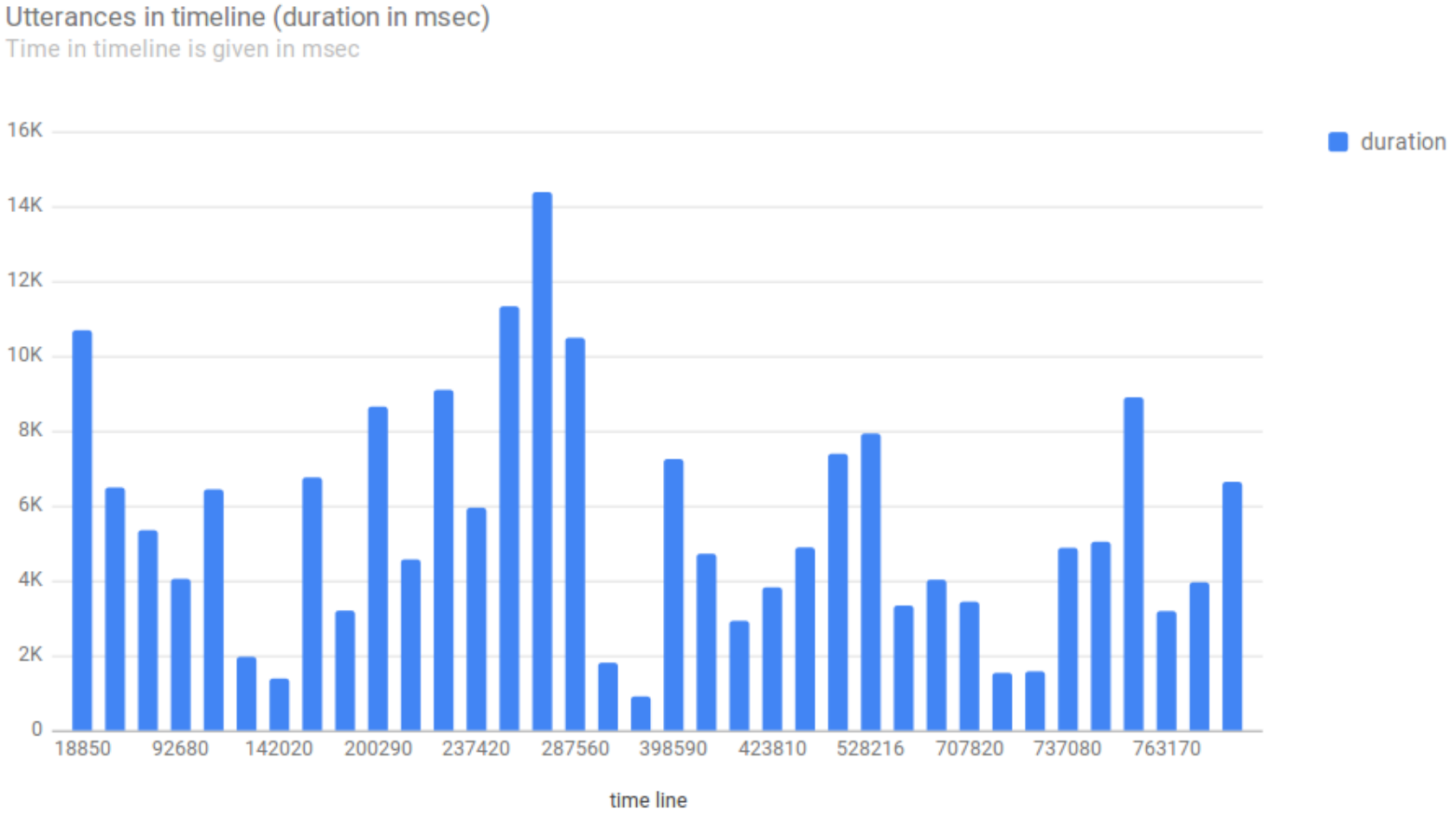
Num of sentences	63	Avg. words per sentence	9.92063492063	Min. words per sentence	1
Num of Words	623	Median words per sentence	10	Max. words per sentence	22
		Mode words per sentence	13	Range. words per sentence	21



Sample analyses: density in timeline

Density Graph (duration in msec) for What-Able-NZ-Hits.txt:

Mean duration	5567.8888888889	Min duration	950	Variance	9715490.8765432
Median duration	4929	Max duration	14420	Standard deviation	3161.182742291



Next steps

- Creating a website, a platform and uploading materials: pagines.uab.cat/viw/

- Expanding the corpus: students? Process data? Other languages? Project sustainability

- Another level of annotation: AD annotation (who, what, why, where, when)

Designing a new platform for AD analysis: the VIW project

Anna Matamala, Marta Villegas

Universitat Autònoma de Barcelona

anna.matamala@uab.cat

CILC 2016, 8th International Conference on Corpus Linguistics

Málaga, 02-04/03/2016.

Funding: FFI2015-62522-ERC

2014SGR0027

CC BY NC