university of groningen

UNIVERSITAT DE BARCELONA

Writing TV series scripts to measure learning from extensive audio-visual input

Anastasia Pattemore and Maria-del-Mar Suárez

university of groningen

UNIVERSITAT DE BARCELONA

Audiovisual input for language learning

– Beneficial for vocabulary, grammatical constructions, speech segmentation, listening comprehension
(see Montero Perez, 2022)

– On-screen text (L1 subtitles, L2 captions, TE captions, no captions) supports learning of different features

university of groningen

UNIVERSITAT DE BARCELONA

Measurement issue

– Majority of audiovisual input and captioning studies used written test (Jelani & Boers, 2018)

– Those tests are rarely productive

– Need to have more ecologically valid productive measures/activities (e.g. Kusyk & Sockett, 2012)

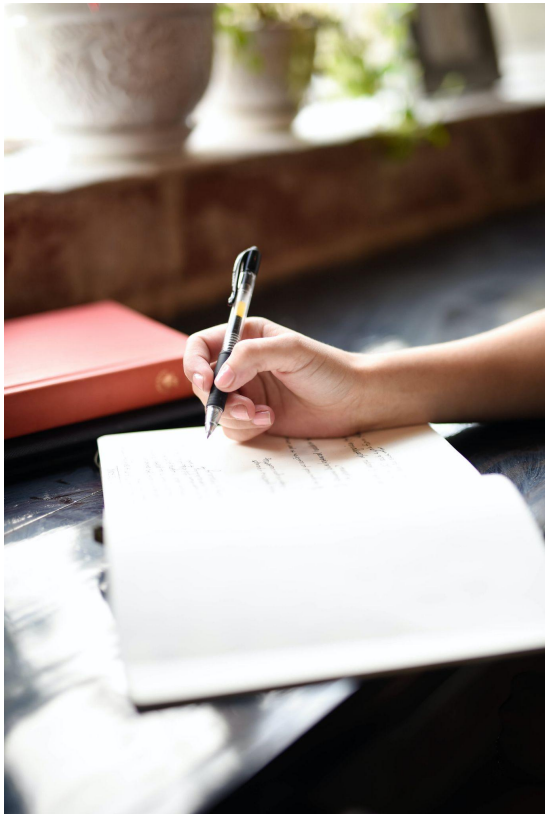
university of groningen

UNIVERSITAT DE BARCELONA

Written production: Fan fiction

– Fan fiction for language learning (*FanTales*, e.g. Cornillie et al., 2021)

– Frequent viewers produced more TV series constructions than infrequent viewers (Sockett, 2014)



university of groningen

UNIVERSITAT DE BARCELONA

Aim of the study

A follow-up study on Sockett (2014):

–larger participant group
–more controlled environment with various types of exposure
–natural language toolkit analysis

university of groningen

UNIVERSITAT DE BARCELONA

Research question

To what extent does prolonged viewing of TV series under different viewing conditions affect language uptake (as demonstrated by written output)?

university of groningen

UNIVERSITAT DE BARCELONA

The study

university of groningen

UNIVERSITAT DE BARCELONA

Participants

134 Audiovisual Communication undergraduate students

Catalan/Spanish bilinguals

17-32 y.o. (M=19.32)

Captions (n=71), TE captions (n=37), no captions (n=26)

A1–C2 proficiency levels (Mean B2)

university of groningen

UNIVERSITAT DE BARCELONA

TV series

university of groningen

Instruments

– Oxford Placement Test (Allan, 2004)

– Script writing (40 min)

Episode scene-writing competition!

The producers of *The Good Place* are looking for new ideas!

How is Tahani going to keep Eleanor in The Good Place? What is her plan? Write a conversation between the main characters describing what they are going to say to each other and what they are going to do to save Eleanor.

Write between 200-220 words. Try to **use as many phrases as you can** from what you **heard/read/learnt** while you were watching ***The Good Place*** show.

Write the first conversation of the next episode using the background story given.

You have 20 minutes to complete the task.

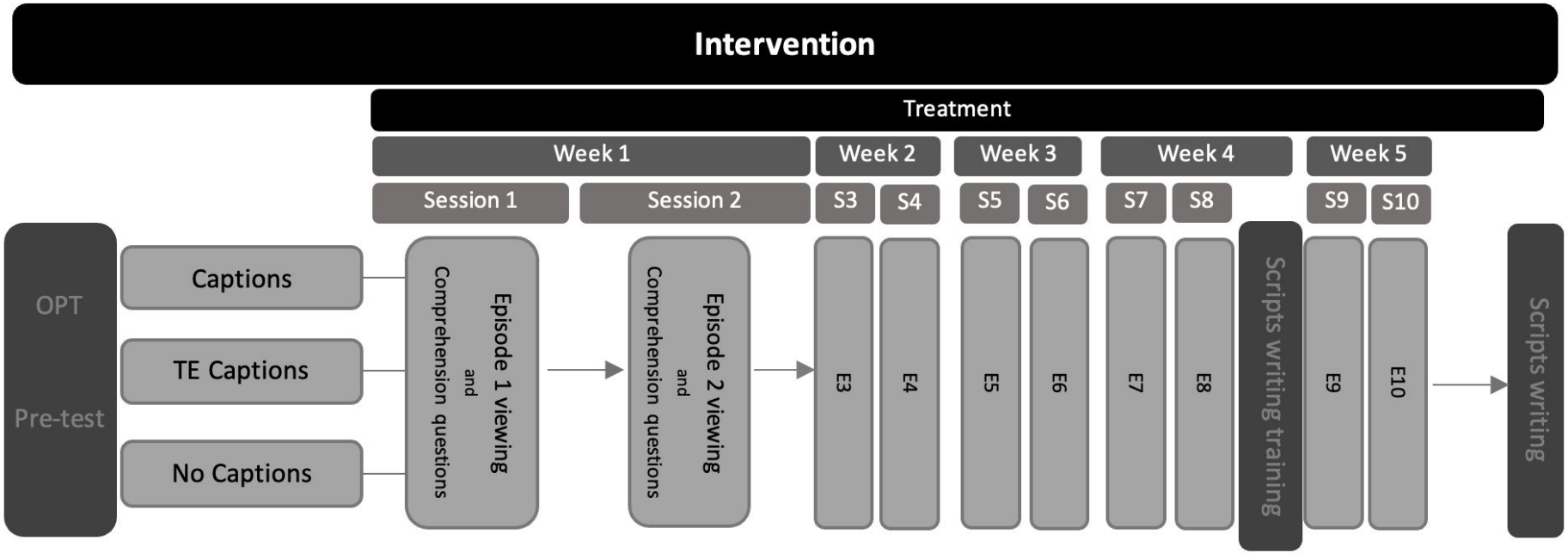
INT. MICHAEL'S OFFICE AT THE GOOD PLACE - ….(TIME?)……

Tahani, Chidi, Jason, and both Eleanors break into Michael's office without knocking and interrupt his conversation with Janet. Tahani impatiently sits on the chair and starts explaining her plan. Everyone is listening to her and discusses her plan when she finishes.

university of groningen

UNIVERSITAT DE BARCELONA

Procedure



university of groningen

Analysis: Corpora

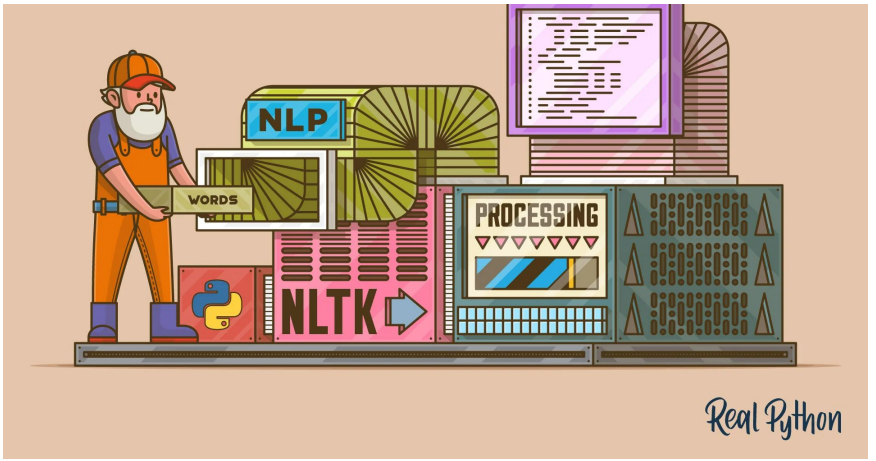
– *The Good Place* 10 episodes scripts corpus
(30,334 words)

– Students' scripts corpus
(32,090 words)

– Comparison corpora (*Friends* and *The Good Doctor*)
(23,756 and 44,351 words)

university of groningen

Analysis procedure



1) Natural language toolkit (nltk)
2) The text is tokenized and lemmatised
3) a. Stopwords (e.g. I, me, you, yourself, am, is, are, have, has, between, 127 in total) are filtered

   b. Stopwords are kept
4) The word units are vectorised
5) Cosine similarity is calculated

university of groningen

General Linear Models

Dependent variable: Cosine similarity scores (with or without stopwords)

Independent variable: Viewing group

Covariate: Proficiency

university of groningen

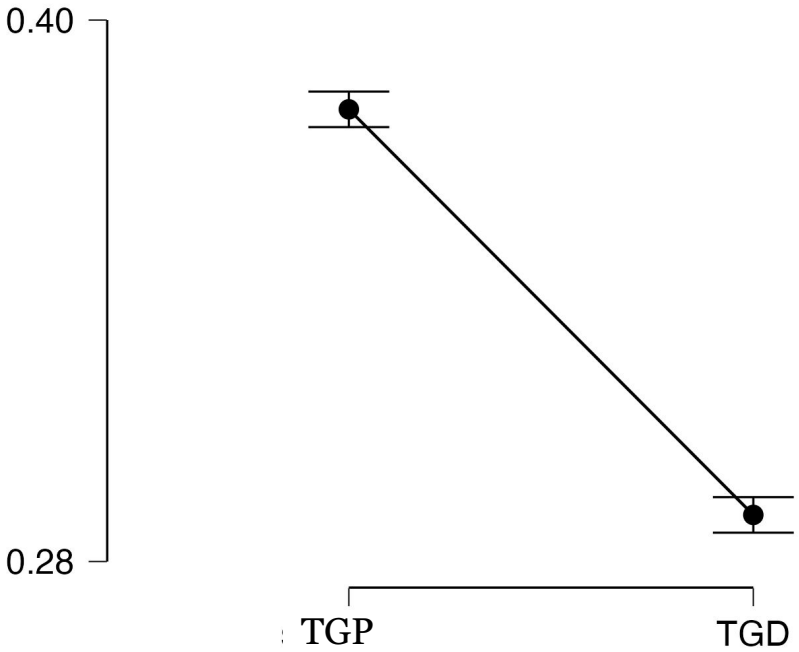
UNIVERSITAT DE BARCELONA

The results

university of groningen

UNIVERSITAT DE BARCELONA

Control comparison: *The Good Doctor*

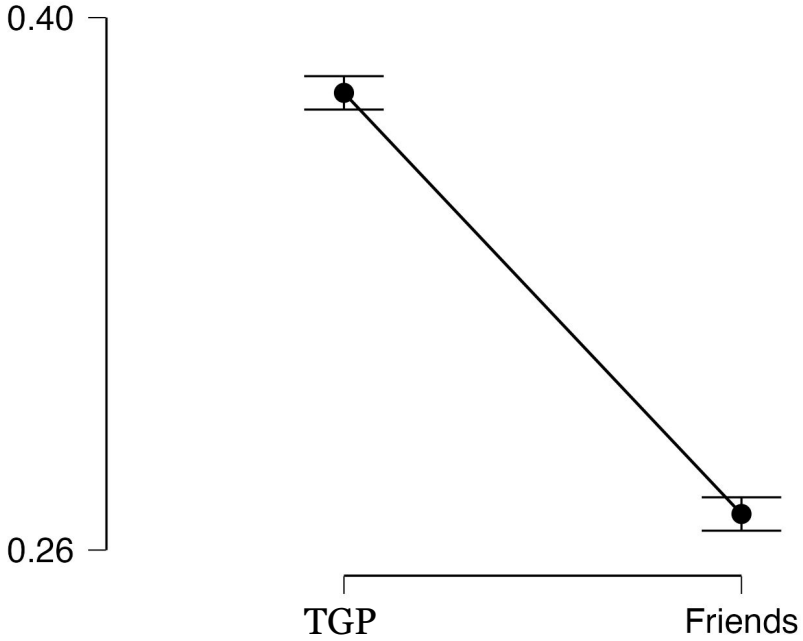


t (133) = 31.920, $p < .001$, $d = 2.758$

university of groningen

UNIVERSITAT DE BARCELONA

Control comparison: *Friends*



t (133) = 35.282, $p < .001$, $d = 3.048$

university of groningen

UNIVERSITAT DE BARCELONA

Participants' and TV series' scripts comparison

	Stopwords filtered Mean (SD)
Original TV series scripts	1.00
All participants	0.380 (.043)
Captions	0.391 (.040)
TE captions	0.365 (.043)
No captions	0.372 (.041)

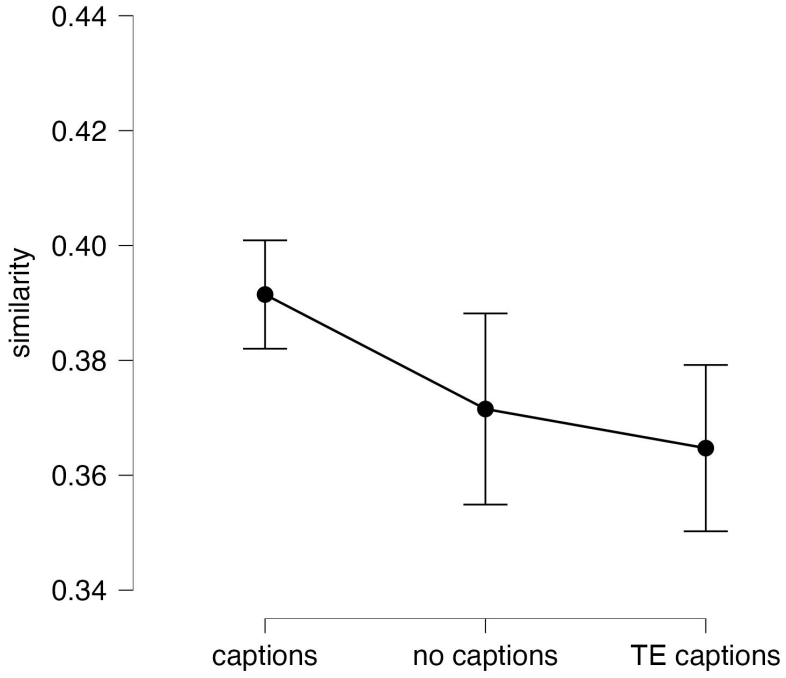
university of groningen

UNIVERSITAT DE BARCELONA

Participants' and TV series' scripts comparison

	Stopwords filtered Mean (SD)	Stopwords not filtered Mean (SD)
Original TV series scripts	1.00	1.00
All participants	0.380 (.043)	0.614 (.073)
Captions	0.391 (.040)	0.621 (.072)
TE captions	0.365 (.043)	0.588 (.077)
No captions	0.372 (.041)	0.632 (.065)

Effect of viewing group (stopwords filtered)

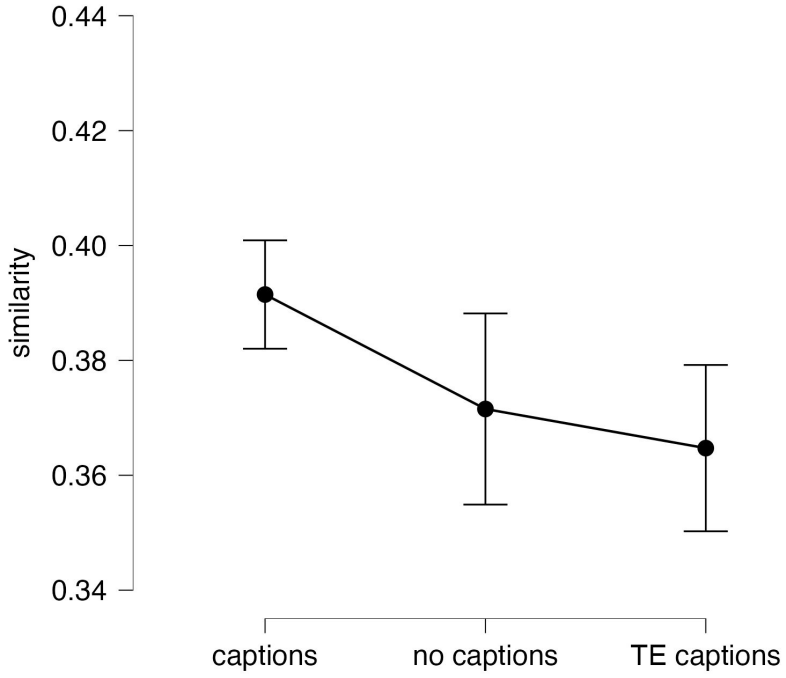


	F	*p*	η^2
Group	4,582	.012	.063
Proficiency	3,264	.073	.022

university of groningen

UNIVERSITAT DE BARCELONA

Post hoc comparison (stopwords filtered)

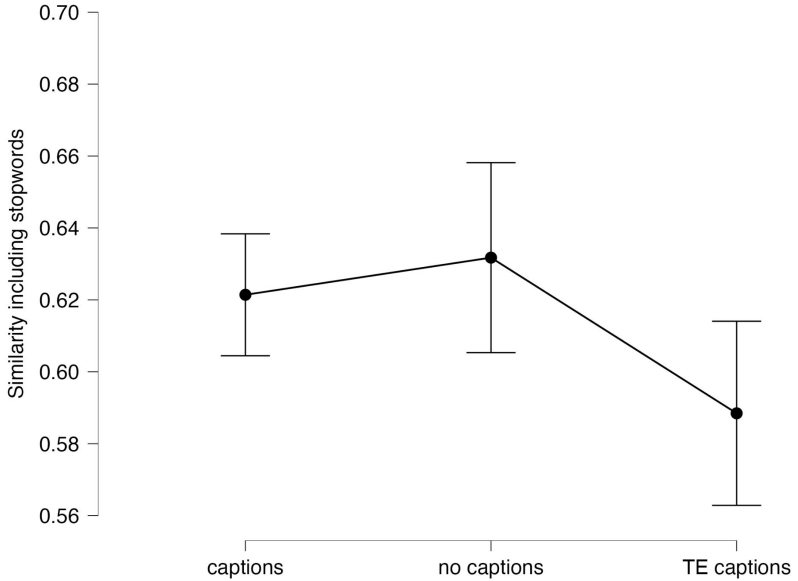


	Mean difference	*t*	*p*
Captions – NC	.018	1.868	.152
Captions – TE	.024	2.854	.014
NC – TE	.006	0.611	.815

university of groningen

UNIVERSITAT DE BARCELONA

Effect of viewing group (including stopwords)

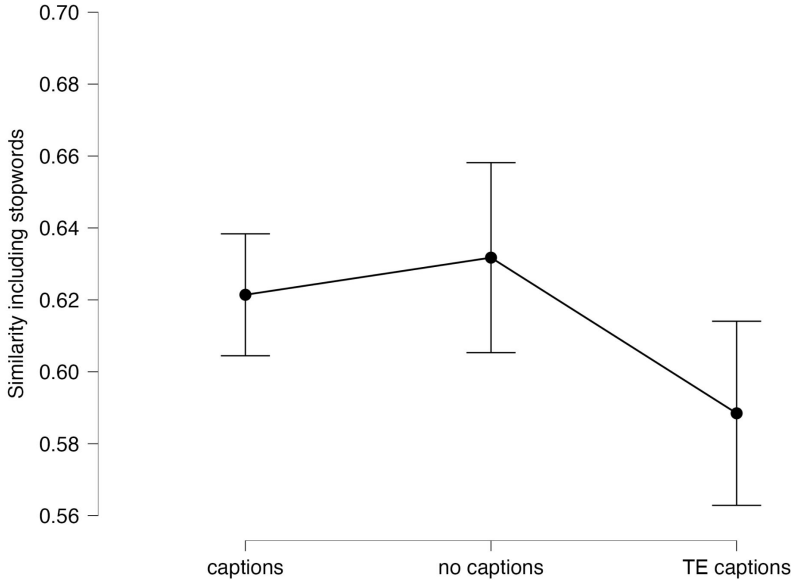


	F	*p*	η^2
Group	3,183	.045	.046
Proficiency	1,037	.310	.008

university of groningen

UNIVERSITAT DE BARCELONA

Post hoc comparison (including stopwords)



	Mean difference	*t*	*p*
Captions – NC	-.013	-.759	.729
Captions – TE	.030	2.037	.107
NC – TE	.043	2.327	.056

university of groningen

UNIVERSITAT DE BARCELONA

Discussion

university of groningen

UNIVERSITAT DE BARCELONA

TV series scripts as a measurement of uptake

– Students' produced similar scripts to *TGP*

– Scripts' similarity depended on their viewing mode (cf. previous findings with less productive, more controlled tests, e.g. Pattemore & Muñoz, 2022)

→ Promising measure

– Mixed findings on TE captions (cf. Montero Perez, 2022)

university of groningen

UNIVERSITAT DE BARCELONA

Group difference

- Captions group performed better for content words and showed less variability in scores (levelled the playing field, Pattemore & Muñoz, 2020)

- Once the function words were included, the advantage disappeared

university of groningen

Group difference

- TE captions always performed worse

→ possible cognitive overload

→ enhanced expressions appeared isolated (scripts→ contextualized)

university of groningen

UNIVERSITAT DE BARCELONA

Limitations and Future research

- preliminary analysis – more data cleaning required

- decision on filtering the stopwords

→ creation of a stopword list for LL

- more complex NLP vectorisers (e.g. text2vec)

- chunking (e.g. Sockett, 2014)

university of groningen

UNIVERSITAT DE BARCELONA

Conclusion

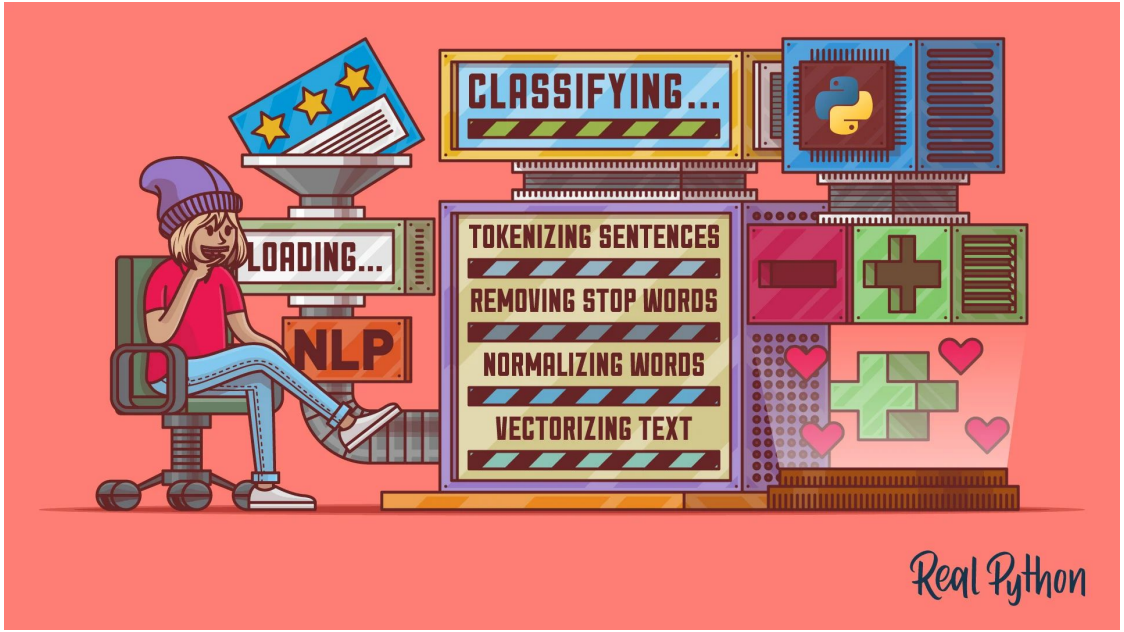
- TV series viewing leads to uptake

- Of productive language, not mere receptive language

- Captions are beneficial for this form of learning (for content words and levelling the playing field)

university of groningen

UNIVERSITAT DE BARCELONA



a.pattemore@rug.nl
mmsuarez@ub.edu