

# Natural Language Processing in the Post Covid-19 World

First Annual Transform4Europe PhD conference

Sofia, 9 December 2021

David Tomás  d\_tomas



Universitat d'Alacant  
Universidad de Alicante

Co-funded by the  
Erasmus+ Programme  
of the European Union



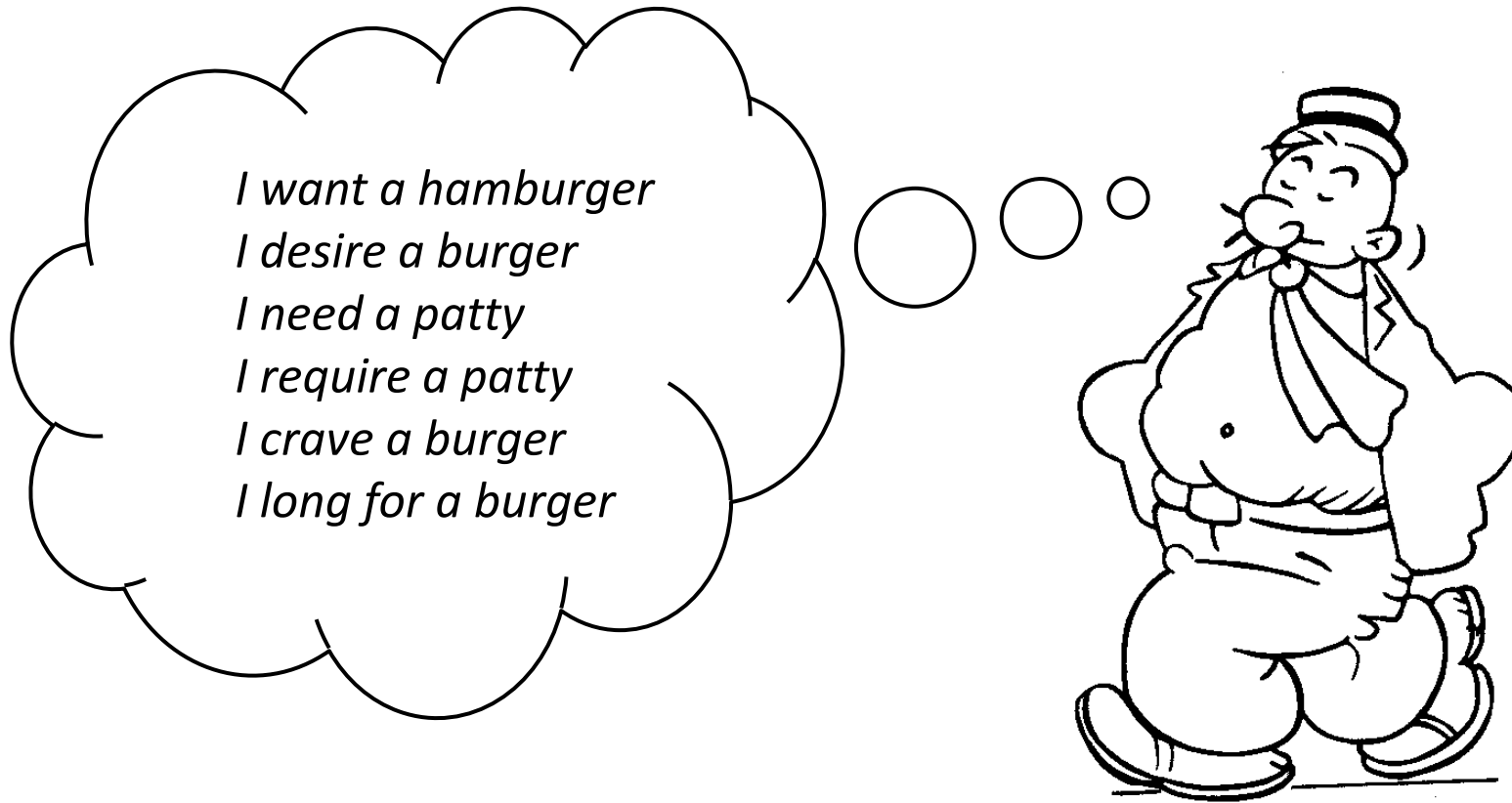
**TRANSFORM  
4EUROPE**







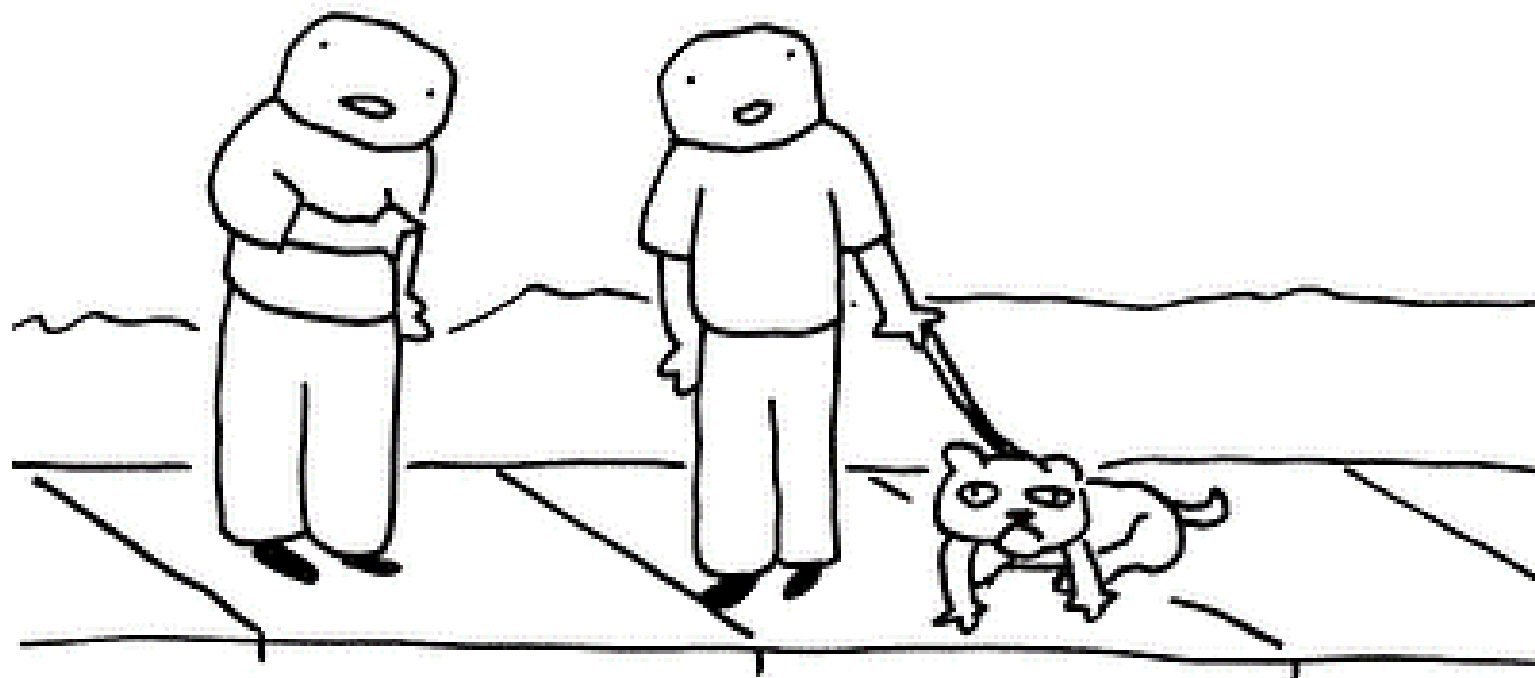
# The problem of **synonymy**



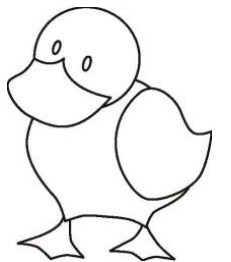
# The problem of polysemy

nice dog, man...  
you pick up a lot  
of girls with him?

nah, he can only  
lift a few pounds

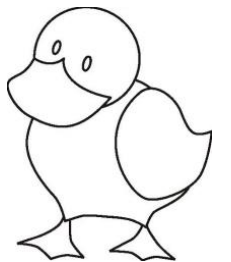


Find at least 5 meanings of the sentence *I made her duck*



Find at least 5 meanings of the sentence *I made her duck*

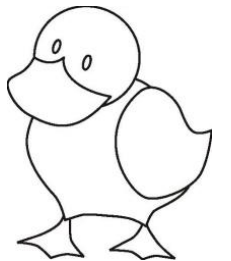
- *I cooked duck for her benefit (to eat)*





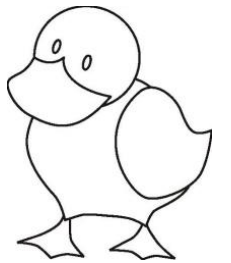
Find at least 5 meanings of the sentence *I made her duck*

- *I cooked duck for her benefit (to eat)*
- *I cooked duck belonging to her*



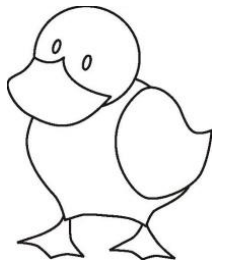
Find at least 5 meanings of the sentence *I made her duck*

- *I cooked duck for her benefit (to eat)*
- *I cooked duck belonging to her*
- *I created the (plaster?) duck she owns*



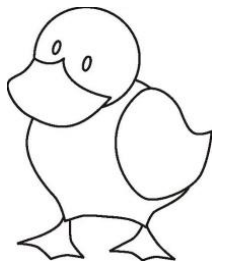
Find at least 5 meanings of the sentence *I made her duck*

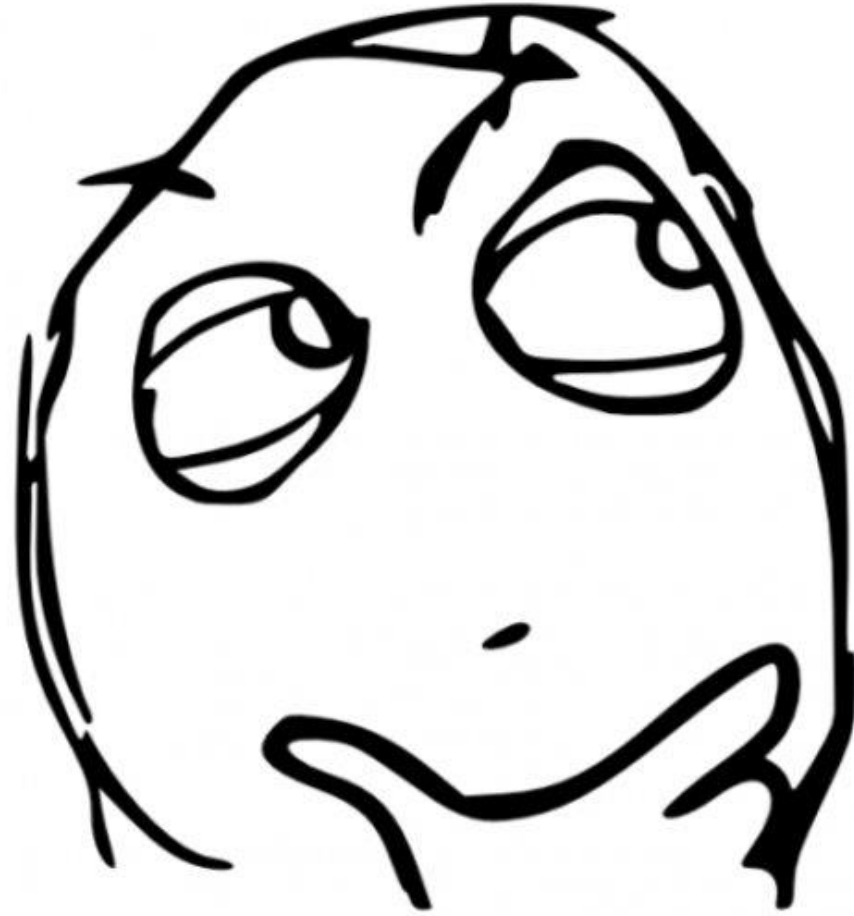
- *I cooked duck for her benefit (to eat)*
- *I cooked duck belonging to her*
- *I created the (plaster?) duck she owns*
- *I caused her to quickly lower her head or body*



Find at least 5 meanings of the sentence *I made her duck*

- *I cooked duck for her benefit (to eat)*
- *I cooked duck belonging to her*
- *I created the (plaster?) duck she owns*
- *I caused her to quickly lower her head or body*
- *I waved my magic wand and turned her into duck*





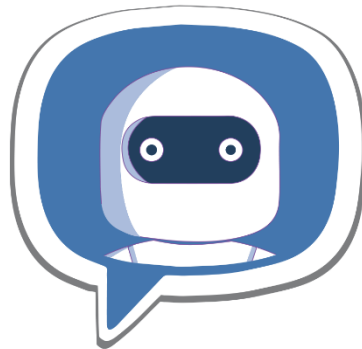
# Natural Language Processing to the rescue!



# Natural Language Processing to the rescue!



Google



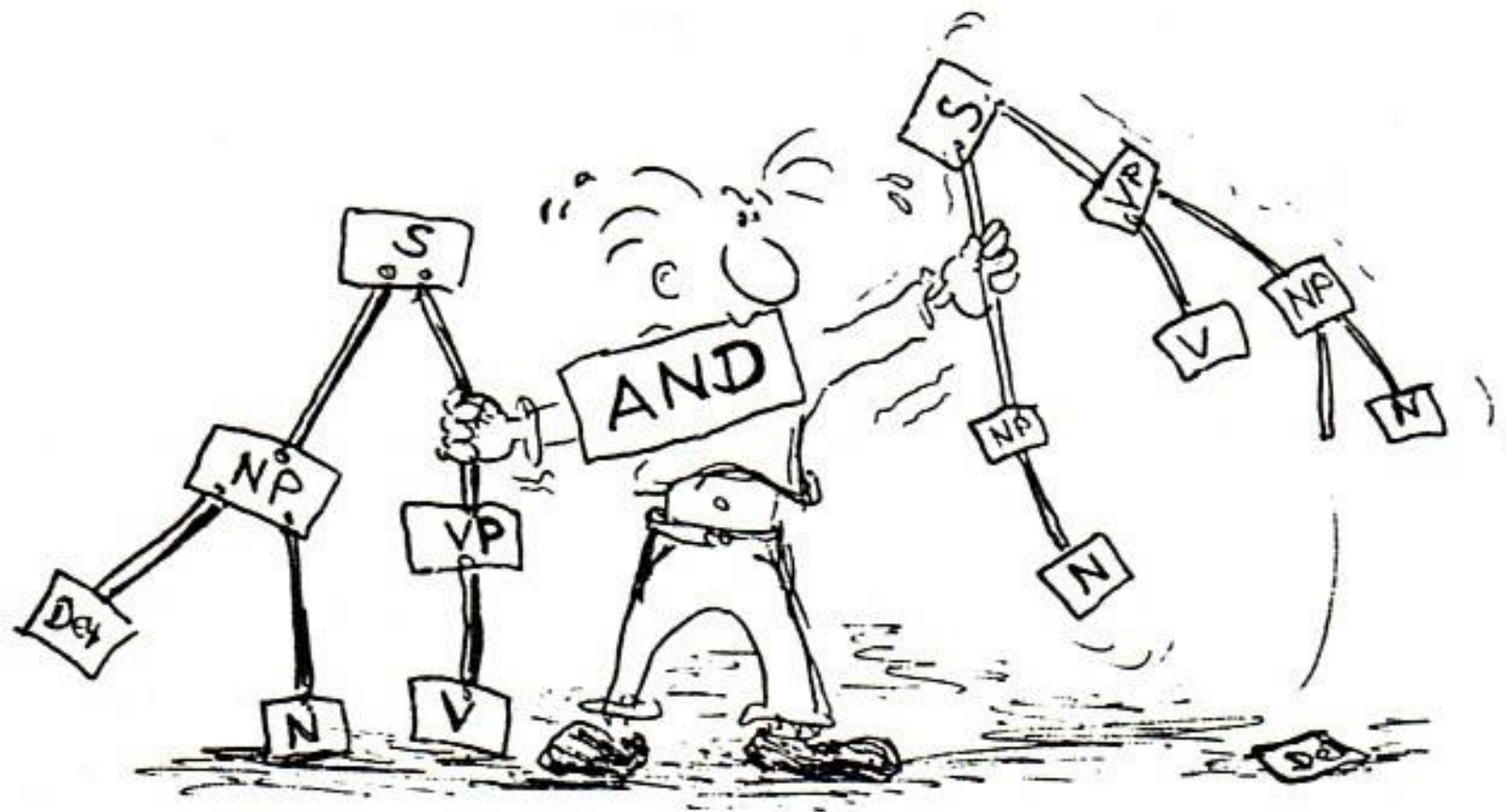
Linguistic approach



Statistical approach



# The linguistic approach

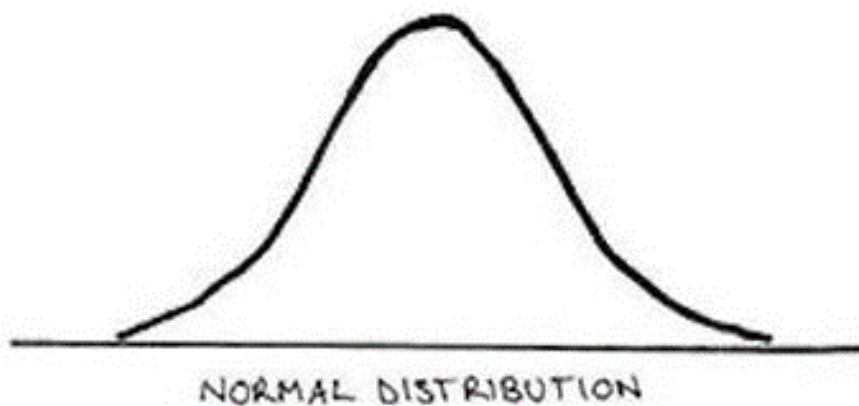


*The grand jury  
commented on a  
number of other topics.*



The the DT 1  
grand grand JJ 0.832524  
jury jury NN 1  
commented comment VBD 0.954545  
on on IN 0.971769  
a 1 Z 0.99998  
number number NN 0.998704  
of of IN 0.999898  
other other JJ 0.632399  
topics topic NNS 1  
. . Fp 1

# The **statistical** approach



zurpojo



# zurpojo

*A bottle of **zurpojo** is on the table.*

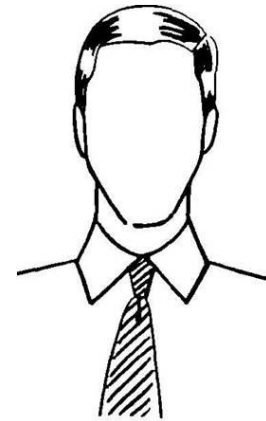
*Everyone likes **zurpojo**.*

***Zurpojo** makes you drunk.*

*We make **zurpojo** out of corn.*

# The distributional hypothesis

*Words that occur in the  
same contexts tend to have  
similar meanings*



Zellig Harris (1954)

So... statistics or linguistics?

# So... statistics or linguistics?

*Whenever I fire a linguist our  
system performance improves*



Frederick Jelinek (1988)

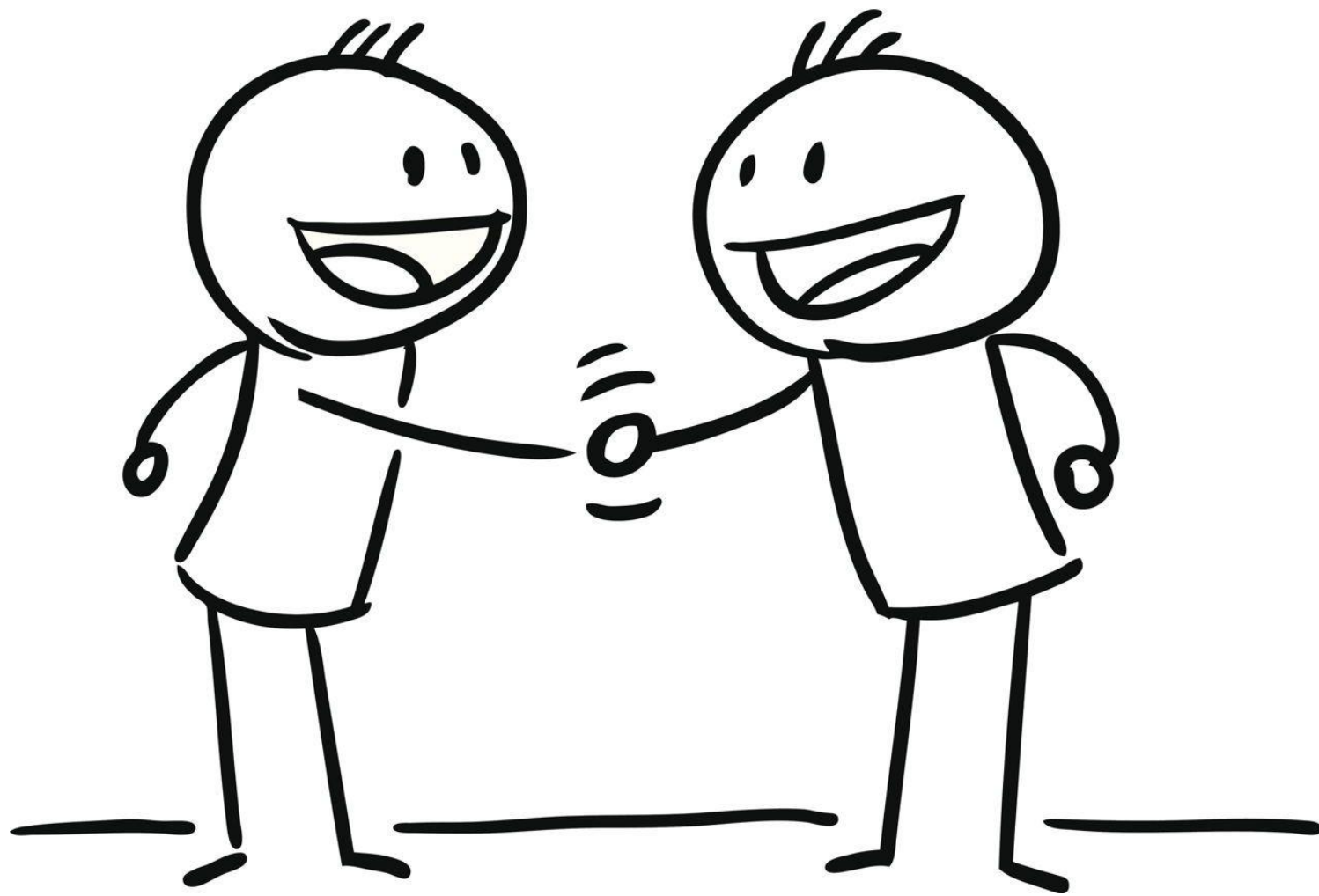


# So... statistics or linguistics?

*Some of my best friends are  
linguists*



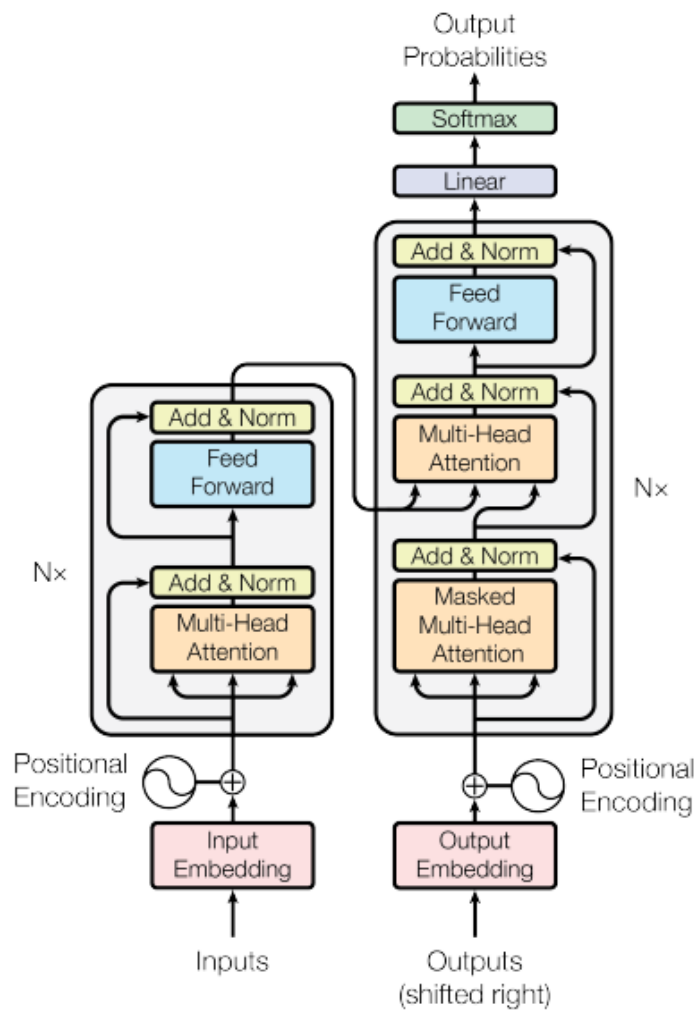
Frederick Jelinek (2004)



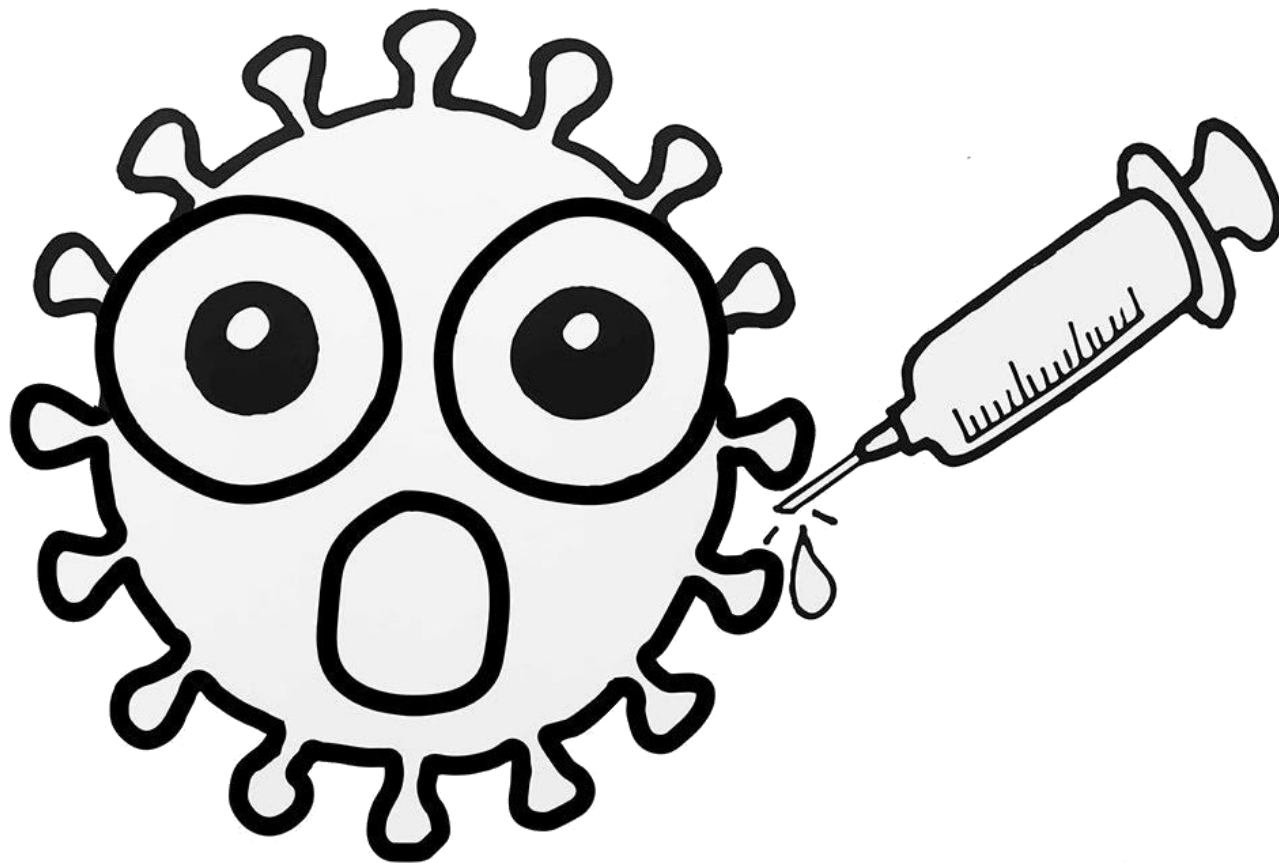
And then they arrived... Transformers!



# And then they arrived... Transformers!



# Applications of NLP for COVID-19



Assist with **healthcare** and **rehabilitation**



# Monitoring social media



# Scientific papers mining





Let's practice!

<https://bit.ly/transform4europe>

# References

- DeCaprio, D., Gartner, J. A., Burgess, T., Kothari, S., Sayed, S., and McCall, C. J. (2020). *Building a covid-19 vulnerability index*. medRxiv. Available at: <https://arxiv.org/abs/2003.07347>
- Carriere J., Shafi H., Brehon K., Pohar Manhas K., Churchill K., Ho C. and Tavakoli M. (2021) *Case Report: Utilizing AI and NLP to Assist with Healthcare and Rehabilitation During the COVID-19 Pandemic*. *Front. Artif. Intell.* 4:613637. doi: 10.3389/frai.2021.613637
- Liu, Y., Whitfield, C., Zhang, T. et al. (2021) *Monitoring COVID-19 pandemic through the lens of social media using natural language processing and machine learning*. *Health Inf Sci Syst* 9, 25. <https://doi.org/10.1007/s13755-021-00158-4>
- Cury, R. C., Megyeri, I., Lindsey, T., Macedo, R., Batlle, J., Kim, S., Baker, B., Harris, R., and Clark, R. H. (2021) *Natural Language Processing and Machine Learning for Detection of Respiratory Illness by Chest CT Imaging and Tracking of COVID-19 Pandemic in the United States*. *Radiology: Cardiothoracic Imaging* 3:1
- COVID-19 Open Research Dataset Challenge (CORD-19): <https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge>
- How Natural Language Processing (NLP) Can Help Us Understand the Landscape of COVID-19 Information: <https://www.copyright.com/blog/natural-language-processing-information-covid-19/>
- Mahlberg, M. and Brookes, G. *Language and Covid-19: Corpus linguistics and the social reality of the pandemic* (2021) *International Journal of Corpus Linguistics* 26:4, pp. 441-443. <https://doi.org/10.1075/ijcl.00043.mah>

Questions? [dtomas@dlsi.ua.es](mailto:dtomas@dlsi.ua.es)

