

How good is AI at Natural Language Understanding and Inferencing?

Olga Kellert

Universidade da Coruna, CITIC, Grupo LyS, Depto. de Ciencias de la Computación y Tecnologías de la Información

AIAI

16th of May 2024

Outline

1. Study 1 on semantic similarity of sentences: SBERT
2. Study 2 on implicit knowledge: ChatGPT
3. Conclusion and outlook

Outline

- 1. Study 1 on semantic similarity of sentences: SBERT**
 - a. Overview/Goals
 - b. Methodology/Data
 - c. Results
 - d. Conclusion and Discussion
2. Study 2 on implicit knowledge: ChatGPT
3. Conclusion and outlook

Outline

1. **Study 1 on semantic similarity of sentences: SBERT**
 - a. Overview/Goals
 - b. Methodology/Data
 - c. Results
 - d. Conclusion and Discussion
2. Study 2 on implicit knowledge: ChatGPT
3. Conclusion and outlook

Sociolinguistic and demographic factors influencing the transmission of COVID-19 information in rural indigenous and non-indigenous populations in Peru and Ecuador.

Olga Kellert (PI) and Stavros Skopeteas (Co-PI), project partners: Claudia Crespo (Perú), Marleen Haboud (Ecuador), Fernando Ortega (Ecuador)



Objective: measure the influence of sociolinguistic and demographic factors on the transmission of information about COVID-19 in rural indigenous and non-indigenous populations in Peru and Ecuador.

Fernando Ortega

Kellert et al. 2023 Deutscher Hispanistentag XXII. Universität Graz. 22.2.2023-26.2.2023, Kellert et al. in prep.

Data and Methodology

- Interviews with indigenous and non-indigenous rural inhabitants of Peru and Ecuador on COVID-19:
 - 20 questions about COVID-19, e.g. **Do you know how the COVID-19 virus is transmitted?**
 - 10 questions on metadata (what languages do you speak, what profession do you have, etc.)
- Estimate agreement with government information or “normative answers”:
 - **Manual analysis:** students and medical experts annotate the semantic similarity between participants’ answers and government information, e.g. **COVID-19 is transmitted through close social contact** (government), **I can get the virus being exposed to a crowd of people** (strong semantic similarity)
 - **Automatic analysis** of measuring semantic similarity between normative answers and participants answers using LLMs (BERT) (Kellert & Zaman 2023, AmericasNLP, ACL 2023)
- Test a statistical model of the social factors influencing the responses. The social factors are:
 - **Country** (Peru, Ecuador), **Education** (Sp. “primaria” = 10 years school, “secundaria”), **Source of information** (social networks, TV, radio, etc.), **Gender, and Language** (monolingual Spanish, monolingual Quechua/Kichwa, bilingual Spanish-Quechua/Kichwa). (Kellert et al 2023, Kellert et al. in prep.)

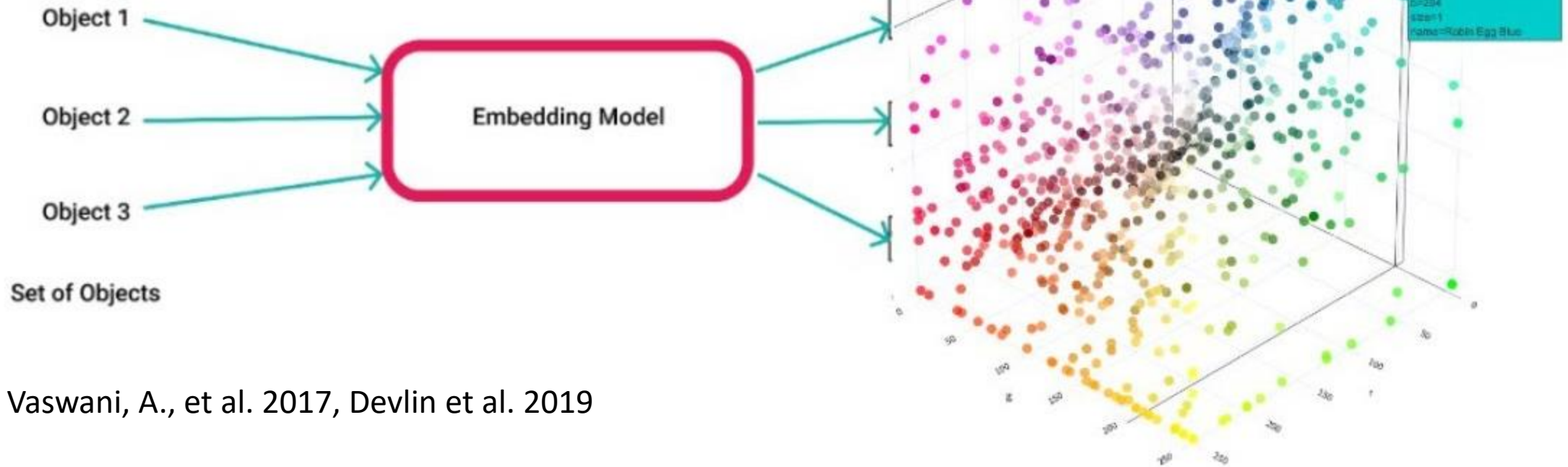
- Estimate agreement with government information or “normative answers”:
 - **Manual analysis:** students annotate the semantic similarity between participants’ answers and government information, e.g. **COVID-19 is transmitted through close social contact** (government), **I can get the virus being exposed to a crowd of people** (strong semantic similarity)
 - **Automatic analysis** of measuring semantic similarity between normative answers and participants answers using LLMs (BERT) (Kellert & Zaman 2023, AmericasNLP, ACL 2023)

Study 1 today

Vector embeddings

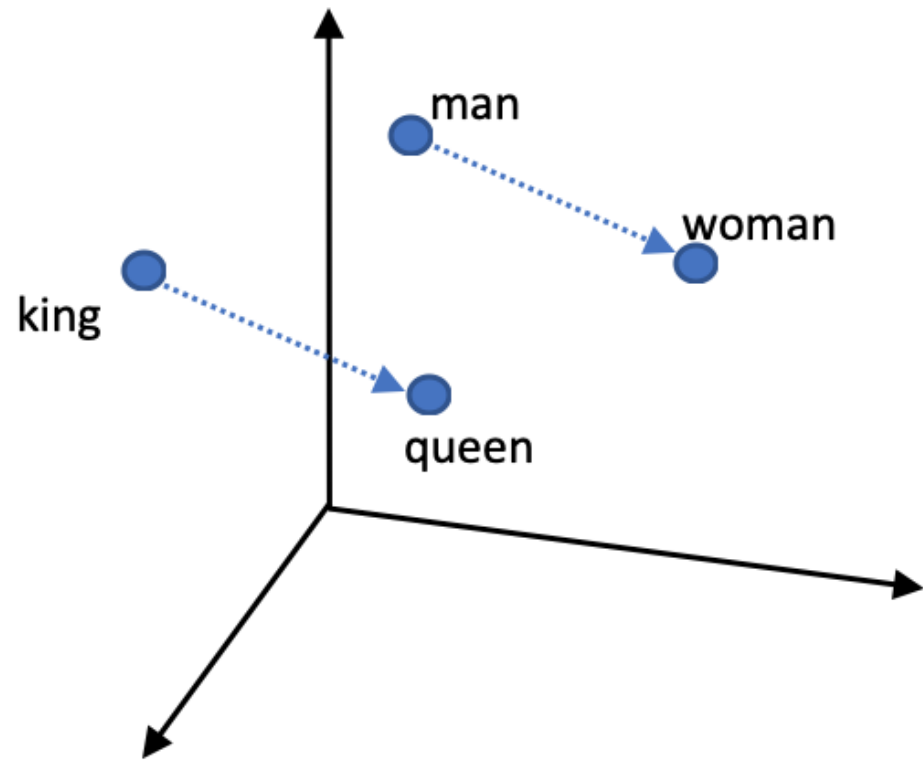
Color represents vector similarity \rightarrow similarity between objects (words, sentences, etc.)

Objects can be Words, Sentences, etc.



Vaswani, A., et al. 2017, Devlin et al. 2019

Semantic Similarity via Cosine Similarity

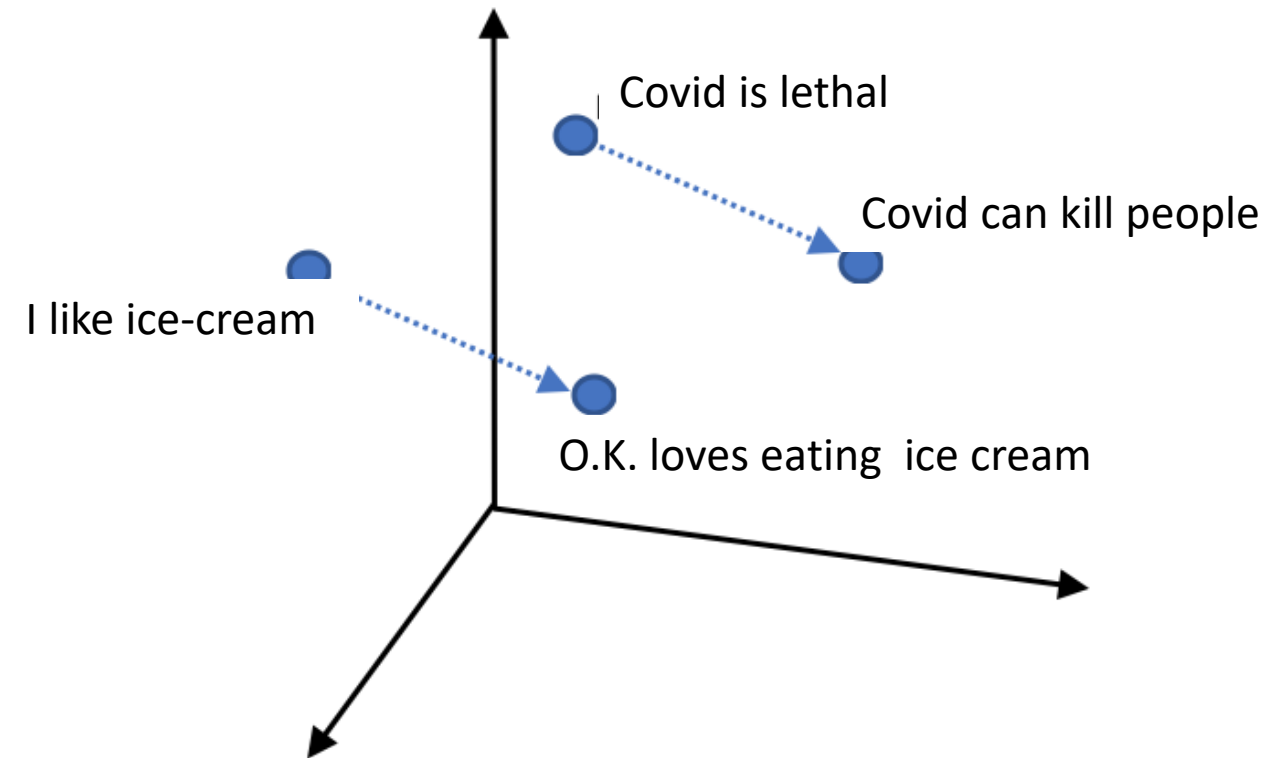


Semantic Similarity (SemS) via Cosine Similarity:

- words are converted into vectors
- SemS is a measure of the distance between vector representations.
- If two words have **similar** meanings like *king* and *queen*, their vectors will be **close** together in the high-dimensional space, resulting in a **small distance** between them.
- Conversely, if the words have different meanings like *king* and *radio*, their vectors will be farther apart.

Semantic Similarity of Sentences

S(entence)BERT Reimers and Gurevych, 2020), <https://sbert.net/>



Semantic Similarity of sentences:

- sentences with **similar** meanings → **close** vectors/smaller distance.
- sentences with **different** meanings → **distant** vectors

Semantic Similarity between Sentences and Images

S(entence)BERT Reimers and Gurevych, 2020), <https://sbert.net/>

● A cat on a table



● Image Embedding

● Text Embedding

● Two dogs in the snow

● London at night

Methodology

Goal: Using SBERT to calculate the semantic similarity via cosine similarity between normative and participants answers:

Question: When should a mask be used?

Normative answer: Especially in closed public places, but it is also useful in outdoor public places.

Answer by participant 1:

Whenever we are in contact with another person.

Outline

- 1. Study 1 on semantic similarity of sentences: SBERT**
 - a. Overview/Goals
 - b. Methodology/Data**
 - c. Results
 - d. Conclusion and Discussion
2. Study 2 on implicit knowledge: ChatGPT
3. Conclusion and outlook

Methodology

```
from sentence_transformers import SentenceTransformer, util

model = SentenceTransformer("all-MiniLM-L6-v2")
```

```
# Two lists of sentences
sentences1 = [
    "The cat sits outside",
    "A man is playing guitar",
    "The new movie is awesome",
]
```

```
sentences2 = [
    "The dog plays in the garden",
    "A woman watches TV",
    "The new movie is so great",
]
```

```
# Compute embedding for both lists
embeddings1 = model.encode(sentences1, convert_to_tensor=True)
embeddings2 = model.encode(sentences2, convert_to_tensor=True)
```

```
# Compute cosine-similarities
cosine_scores = util.cos_sim(embeddings1, embeddings2)
```

```
# Output the pairs with their score
for i in range(len(sentences1)):
    print("{} \t\t {} \t\t Score: {:.4f}".format(
        sentences1[i], sentences2[i], cosine_scores[i][i])
    )
```



A big challenge, if you want to apply this code on your data!

Data preprocessing/filtering

Problem/Challenge: Choosing the right pair for comparison!

Solution: Choosing **the same (aboutness) topic (necessary condition!),** a similar **length (structure?)**

Why it's important ? Differences in topics → Differences in content → Difference in similarity!
Differences in length can have an effect on similarity too!

Normative answer: Especially in closed public places, but it is also useful in outdoor public places.

Answer by participant 1:

Whenever we are in contact with another person. (Pragmatic Inferences: **physical and close contact**)

Necessarily implies both situations



Our approach: Q-A-pairs fulfill the same aboutness topic requirement. Filtering relevant from detailed information, e.g. looking for **main sentences (main information!), same or similar words,** e.g. „distance.“

Methodology

- <https://github.com/mahmuduzzamanDE/ACLAmericaNLP>
- Computing cosine similarity between normative and given answers:
- highest similarity is >0.9 between literally identical answers:
 - Especially in closed public places, but it is also useful in outdoor public places-S1
Similarity: tensor([[0.9961]])
 - **Normative answer**: Especially in closed public places, but it is also useful in outdoor public places.
 - **Answer by participant 1**: Whenever we are in contact with another person.
 - Similarity: tensor([[0.1775]]), # cosine similarity between normative and given answer

Outline

1. **Study 1 on semantic similarity of sentences: SBERT**
 - a. Overview/Goals
 - b. Methodology/Data
 - c. **Results**
 - d. Conclusion and Discussion
2. Study 2 on implicit knowledge: ChatGPT
3. Conclusion and outlook

Results

Task	Semantic Similarity
<i>Value</i>	0.2 Cosine Similarity
<i>Human Annot.</i>	0.7 Human annotation
<i>Accuracy loss</i>	0.5

(Kellert & Zaman 2023, AmericasNLP, ACL 2023)

Outline

- 1. Study 1 on semantic similarity of sentences: SBERT**
 - a. Overview/Goals
 - b. Methodology/Data
 - c. Results
 - d. Conclusion and Discussion**
2. Study 2 on implicit knowledge: ChatGPT
3. Conclusion and outlook

Conclusion and Discussion

- Challenges in selection of sentence pairs, preprocessing
- Low accuracy in computing semantic similarity by SBERT (R&G 2019)
- Pragmatic inference is a big problem:
 - avoid close physical contact to prevent virus infection → avoid crowds of people to prevent virus infection (**0.2** according to SBERT)
 - However, being exposed to a crowd implies close physical contact. SBERT does not capture this implication, but human annotators do!

Conclusion and Discussion

- Future work: fine-tuning SBERT by data from the NLI research (Stanford NLI, SNLI) <https://nlp.stanford.edu/projects/snli/>
- [Samuel R. Bowman](#), [Gabor Angeli](#), [Christopher Potts](#), and [Christopher D. Manning](#). 2015. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Text	Judgments	Hypothesis
A man inspects the uniform of a figure in some East Asian country.	contradiction C C C C C	The man is sleeping
An older and younger man smiling.	neutral N N E N N	Two men are smiling and laughing at the cats playing on the floor.
A black race car starts up in front of a crowd of people.	contradiction C C C C C	A man is driving down a lonely road.
A soccer game with multiple males playing.	entailment E E E E E	Some men are playing a sport.
A smiling costumed woman is holding an umbrella.	neutral N N E C N	A happy woman in a fairy costume holds an umbrella.

Outline

1. Study 1 on semantic similarity of sentences: SBERT
 - a. Overview/Goals
 - b. Methodology/Data
 - c. Results
 - d. Conclusion and Discussion
2. **Study 2 on implicit knowledge: ChatGPT**
3. Conclusion and outlook

Study 2 on implicit knowledge: ChatGPT

- **Goal:** test whether ChatGPT can provide the implicit message of a sentence.
- **Methodology:** Testing several speech acts that do **not coincide** with the literal/compositional meaning of the linguistic sentence act such as **rhetorical questions**, questions used as **imperatives**, etc. (see Grice 1989).
- some example prompts:

Study 2 on implicit knowledge: ChatGPT

- User:
- What does "Can you pass me the salt?" mean? What kind of speech act is it?
- ChatGPT:
 - “The sentence "Can you pass me the salt?" is a **request**. In terms of **speech act theory**, a request is a type of **illocutionary act** where the speaker expresses a desire for the listener to perform a specific action. In this case, the speaker is asking the listener to pass them the salt. The illocutionary force of the sentence is the speaker's intention to have the listener take a particular action, and the **perlocutionary** effect is the expected outcome of the listener actually passing the salt.” **Could be a paragraph from a linguistic book on speech acts!!!**

Study 2 on implicit knowledge: ChatGPT

- User:
- What does "What the hell are you doing here?" mean? What kind of speech act is it?
- ChatGPT:
- “The sentence "What the hell are you doing here?" is an interrogative sentence that can be interpreted as expressing surprise, disbelief, or suspicion. In terms of speech act theory, it functions as a question, specifically an **information-seeking question**. The speaker is seeking an explanation or clarification regarding the presence or actions of the person being addressed. **The illocutionary force** is the speaker's intention to gather information, and the **perlocutionary** effect is the expected response or clarification from the person being questioned.” **The same as before: could be a paragraph from a linguistic book on speech acts!!!**

Study 2 on implicit knowledge: ChatGPT

- User: the virus Covid propagates through contact. Do you think that it also propagates through a crowd of people?
- ChatGPT:
 - As of my last knowledge update in January 2022, COVID-19 primarily spreads through respiratory droplets produced when an infected person talks, coughs, or sneezes. [....]. Being in a **crowded** environment can **increase** the **risk** of exposure to respiratory droplets from infected individuals, which may lead to the spread of the virus. **Close contact** with people who are infected, especially in **crowded** places where physical distancing is challenging, can contribute to the **transmission** of COVID-19.

Conclusion and outlook

- How good is AI at NLU and NLI?
- It depends on the AI. ChatGPT (January 2024) better than SBERT (R&G 2019, K&Z 2023)
- AI is as good at NLU and NLI as its training data. If the training data contains NLI data (e.g. Stanfords corpus data), the AI is better at NLI.
- Future work?
- Testing NLI and NLU of AI in a communication context (future work) in contrast to testing its implicit knowledge representation (Study 2).

Thank you for your attention!

References

- Reimers, N., Gurevych, I.: Sentence-BERT: sentence embeddings using siamese BERT-networks. In: EMNLP 2019 (2019). [arXiv:1908.10084](https://arxiv.org/abs/1908.10084)
- Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. [arXiv:1810.04805v2](https://arxiv.org/abs/1810.04805v2) [cs.CL], 24 May 2019
- Kellert, O. & Mahmud Zaman (2023) Use of NLP in the Context of Belief states of Ethnic Minorities in Latin America. In Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP), pages 1–5, Toronto, Canada. Association for Computational Linguistics.
- Vaswani, A., et al.: Attention is all you need. [arXiv:1706.03762v5](https://arxiv.org/abs/1706.03762v5) [cs.CL], 6 December 2017
- Conneau, A., et al.: Unsupervised cross-lingual representation learning at scale. [arXiv:1911.02116](https://arxiv.org/abs/1911.02116) [cs.CL], 5 November 2019
- SQuAD2.0 The Stanford Question Answering Dataset. <https://rajpurkar.github.io/SQuAD-explorer/>
- Semantic Textual Similarity Wiki. <http://ixa2.si.ehu.es/stswiki>