

Stance detection: a sociolinguistic approach

Emilia Milano
University of Luxembourg

outline

- Data
- Theoretical framework
- First study: ideological ground
- Second study: multilevel analysis
- Next steps

my PhD

Goals

- Theoretical definition of stance
- Ground NLP research in this definition

Initial questions:

- how do stance and identity combine in multilingual environments?
- which are the different pieces combining to form the stance activity?
- can an interactional phenomenon be analysed in this kind of data?

data

RTL.lu comments

- 1.7M posts
- 2008 - today
- anonymised
- Luxembourgish
- thumbs up and down (from 2023)
- date and time
- published database
- unpublished database
- comments in isolation

De_fliiger - 12.05.2024 18:29

 30  8

An datt huet eis 7 Milliounen kascht..

Merci Letzebuerg....

Moni Mich ★★ - 13.05.2024 08:00

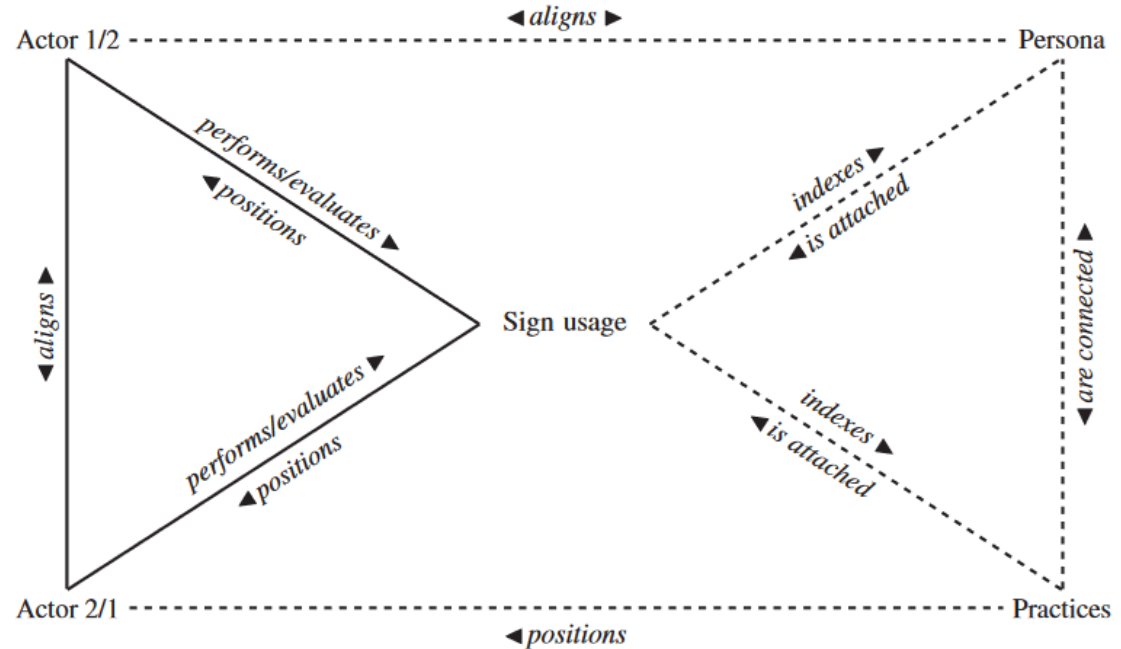
 5  0

@De_fiiger

Vill manner wéi déi 30 Milliounen deen eis de Pavillon fir déi näischt Welt-Expo zu Osaka kascht, an deen ausser e puer Japaner kee wärt gesinn !

theoretical framework

- interactional properties
(evaluation, alignment,
investment)
- ideological properties



theoretical framework

stance as a **multilevel speech activity**:

- Interactional level
- Ideological level
- Language material

ideological level

Ideologies in the discourse about language and multilingualism in Luxembourg:

- 150 manually analysed comments
- focus on the images *figures* use to express the ideological motifs
- ideological motifs: identity, belonging, vitality, recognition, responsibility

ideological motifs

- **Identity:** Strong connection between language, nation (building) & culture; master narrative of resisting German occupation during WW2 through language
- **Vitality:** Threats to and preservation of the language; political policy making vs. people's experiences in practice
- **Belonging:** Social integration through legitimate language skills
- **Responsibility:** Attribution of responsibility for problems and necessary solutions; foreigner-blaming (us vs. them), self-blaming (us for them) or elite-blaming (them vs. us)
- **Recognition:** Hierarchization of languages based on nationalist or pragmatic motives; LB as a dialect of German vs. LB as a language; international standing of LB (EU); ...

example

ID 468569, Y 2017

@ username, username, username, username, ... I can't understand how one can make such generalizations, if it's not to incite, to provoke. And then I wonder why one does that.	-
I have nothing against French, or any other language. Quite the opposite. We have 3 official languages. Then one must be able to talk in all three languages. What's so bad about putting the 3 languages on the same podium?	recognition
Our multilingualism certainly won't be lost if the Luxembourgish language is given a little more importance.	recognition
Why is the Luxembourgish language a thorn in the eye of so many people? Nation branding is desperately sought, but Luxembourgish is condemned.	vitality
And those who stand up for their mother tongue are insulted, condemned/judged.	responsibility
By the way: The author's comment is in flawless Luxembourgish. That makes me happy.	recognition
But I don't understand the connection between right-wing extremism/language.	identity

ideological level – follow up study

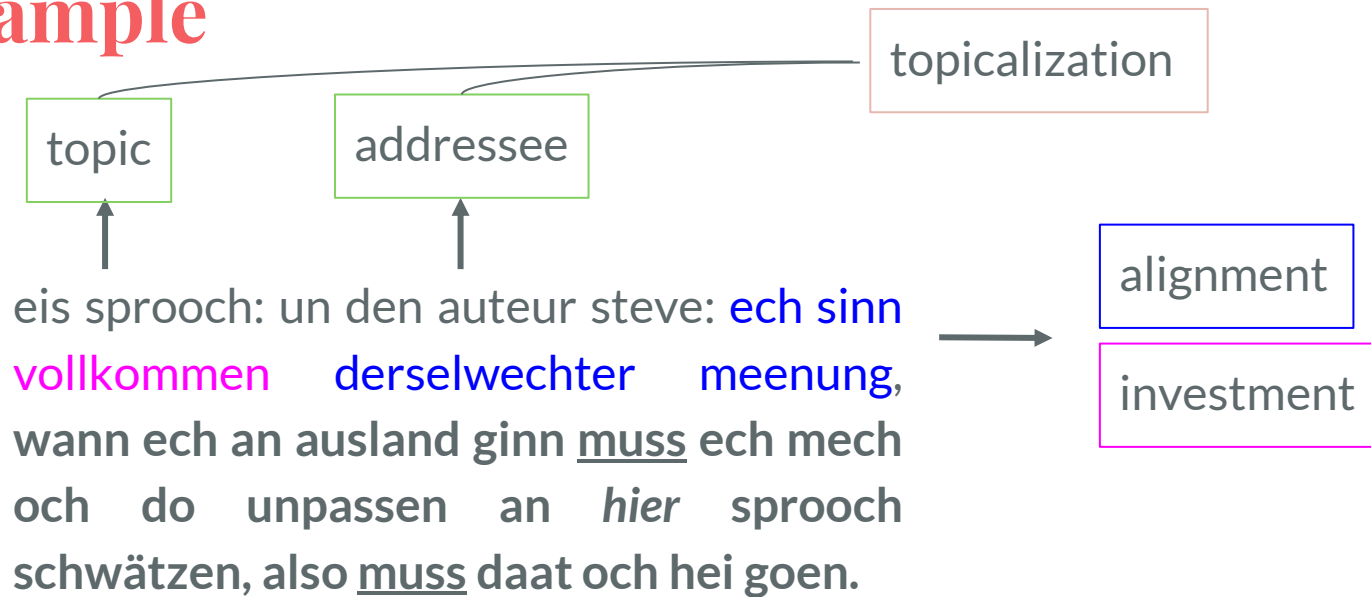
1. consistency of the ideological motifs
 2. annotate a higher number of comments
- annotation with gpt-4o
 - different types of prompting :
 - provision of labels (zero-shot)
 - description of the categories (few-shot)
 - examples of the categories (few-shot)
 - description and examples (full prompt)
- + explanation of the labels

multilevel analysis

1. **linguistic level:** syntactic structure + lexical style + semantic structure + rhetoric
2. **pragmatic level:** topic + addressee
3. **social grounding:** sentiment + knowledge (historical/external events) + language ideology
4. **interactional level:** evaluation, alignment, investment
 - main stance + stance supports



example



ideology: responsibility

evaluation: implicit negative evaluation

next steps

→ annotation study

→ stereotypes detection:

word embedding of 'us' and 'them'

generate sentences about different 'them'

prompt LLM towards different nationalities/'them' (perplexity?)

→ stance as a multilevel speech activity in NLP?

stance direction

stance quality

social grounding

A large red square with a white border, centered on a white background. Inside the square, the text "Tusen takk!" is written in a white, italicized, sans-serif font.

Tusen takk!