University of Stuttgart
Germany

Institut für
Maschinelle
Sprachverarbeitung

Modeling language variation across time and communities

Filip Miletic

July 21, 2022

ellis-stuttgart.eu

Filip Miletic

Institute for Natural Language Processing

vector space models, lexical semantics, language variation

filip.miletic@ims.uni-stuttgart.de ↗

PhD: University of Toulouse (June 2022)

Postdoc: SemChangeMWE project (since July 2022)

PI: Sabine Schulte im Walde

DETECT LANGUAGE   ITALIAN   FRENCH   **ENGLISH**   ⌄

GERMAN   **FRENCH**   ENGLISH   ⌄

Artificial intelligence (AI) is intelligence demonstrated by machines, as opposed to the natural intelligence displayed by animals including humans. AI research has been defined as the field of study of intelligent agents, which refers to any system that perceives its environment and takes actions that maximize its chance of achieving its goals.

347 / 5,000

L'intelligence artificielle (IA) est l'intelligence démontrée par les machines, par opposition à l'intelligence naturelle affichée par les animaux, y compris les humains. La recherche en IA a été définie comme le domaine d'étude des agents intelligents, qui fait référence à tout système qui perçoit son environnement et entreprend des actions qui maximisent ses chances d'atteindre ses objectifs.

If you have any questions about my rate or if you find it necessary to increase or decrease the scope for this project, please let me know.

Clarity · Rephrase sentence

If you have any questions about my rate or ~~if you~~ find it necessary to increase or decrease ~~the scope for~~ this ~~project~~ project's scope, please let me know.

Rephrase   Dismiss

①

Hi, there are no Member Support Associates available to chat right now - they'll be back at 6 AM EST.

I can help you with anything self-service.

Technical Support

What type of problem are you having?

Hardware (e.g. Power issues, Noises)
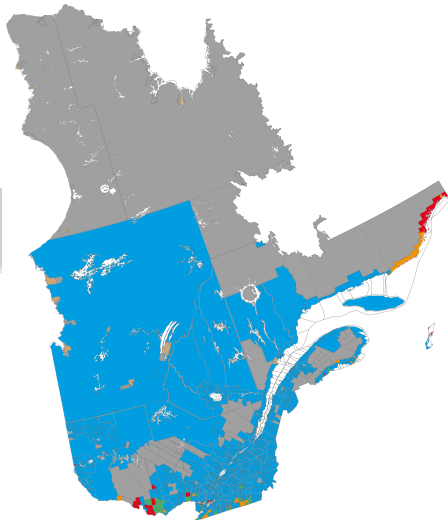
Software (e.g. Bluetooth Pairing, Connection Issues)

App + TV

[1] https://translate.google.com
[2] https://www.grammarly.com
[3] https://support.onepeloton.com

e l l i s
European Laboratory for Learning and Intelligent Systems

2

Implementing and evaluating quantitative empirical methods on precise linguistic issues

Contact-induced semantic shifts in Quebec English

Montreal   'déception'
Big **deception**... you were not present in the Pride Parade [...] Conservatives keep disappointing.
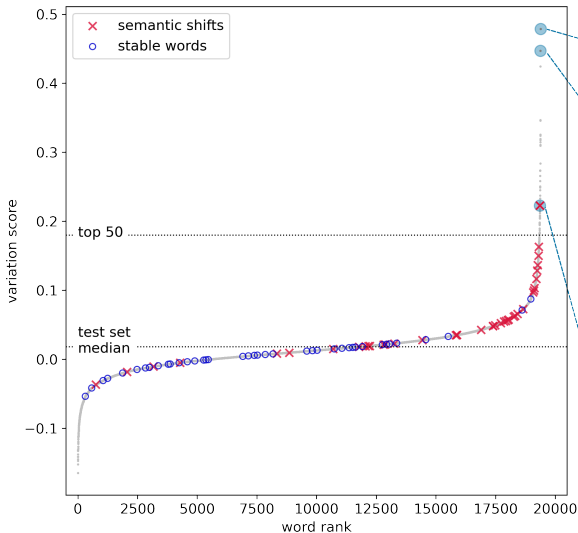

Majority language communities
■■ English   ■■ French
Source: Statistics Canada (2006)

3

- Data: 600m-word corpus of tweets

- Vector representations of word meaning
    - Type-level (e.g. word2vec) $\Rightarrow$ vocabulary-level overview
    - Token-level (e.g. BERT) $\Rightarrow$ in-depth analyses of target words

- Evaluation of implemented methods
    - Custom-built test set for binary classification
    - Face-to-face interviews with speakers from Montreal

e l l i s
European Laboratory for Learning and Intelligent Systems

Semantic shift detection

False positives

pour 'for'
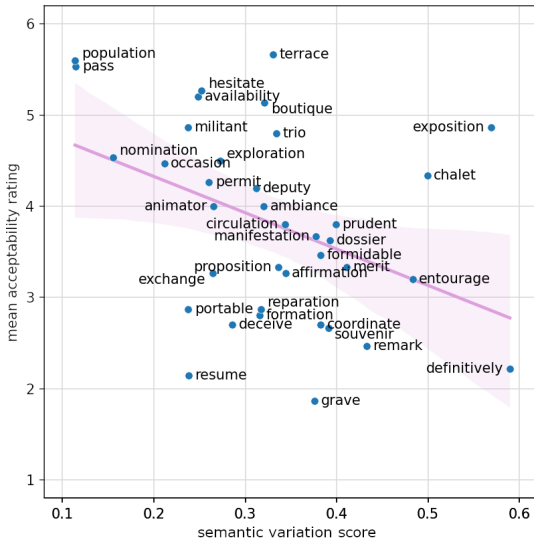French homographs in codeswitched tweets

plateau 'Plateau-Mont-Royal'
local proper noun denoting a Montreal neighborhood

True positive

exposition 'art exhibition'
previously described case, included in our test set

Comparing vector-based and human scores

- A range of data sources and corpus-based methods led to important descriptive contributions

- Validation based on input from local speakers

- Moving beyond isolated lexical items
  ⇒ SemChangeMWE project

- Incorporating a temporal dimension, further linguistic and extralinguistic features

- Interpretability and reliability of the implemented models

Enjoy the kick-off event!