

Unsmurfed

How an LLM Interprets the Smurfs' Distributional Language



by
Peyo

Unsmurfed

“In which drawer is the **thing** again?”

“Come on, hand me the **thing**, so I can finally open this bottle of Merlot.”

“No, you need to turn the **thing** in the other direction!”



by
Peyo

Context is meaning

“You shall know a word by the company
it keeps.”

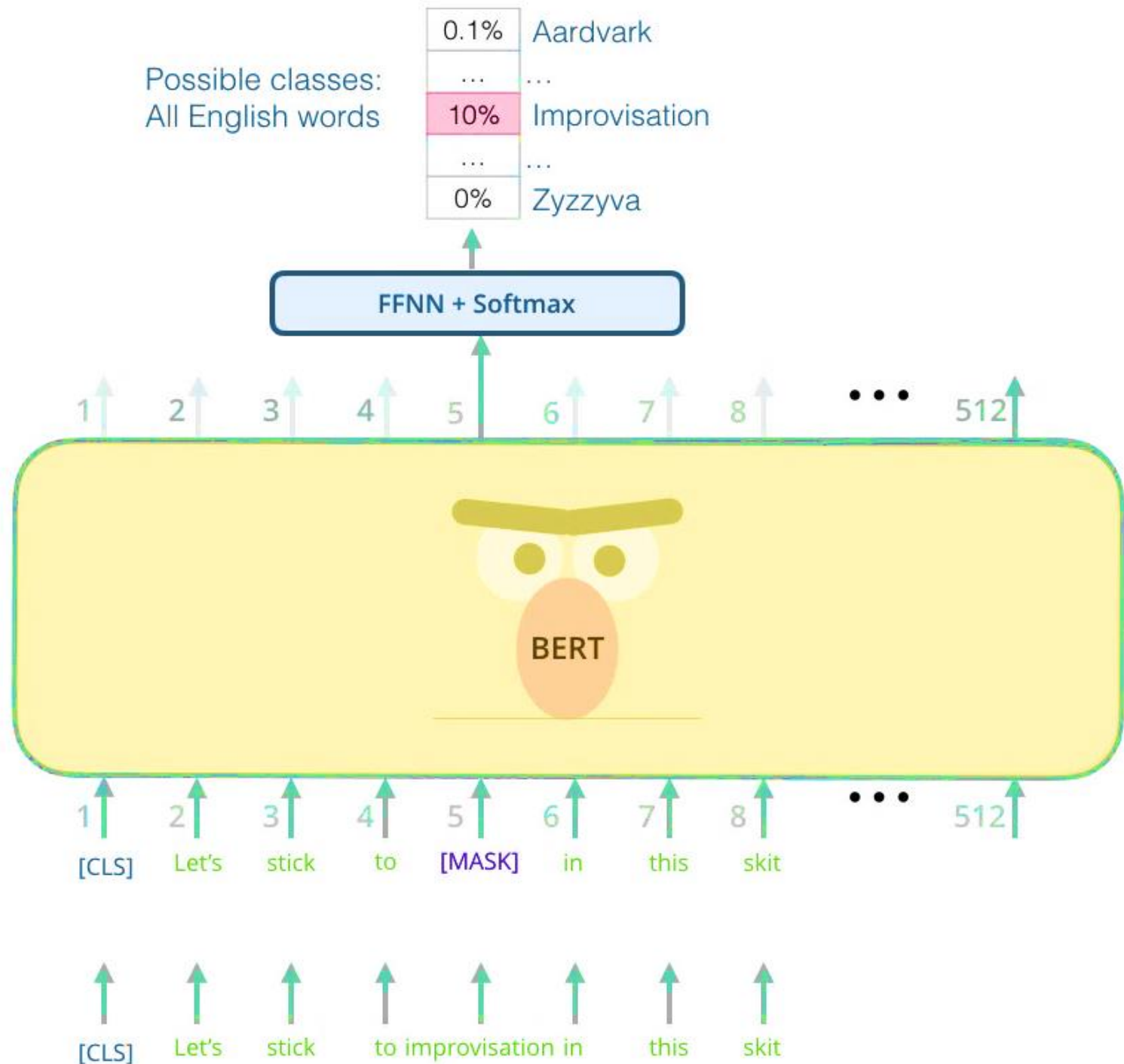
(John R. Firth, 1957)



BERT and Masked Language Modeling

Randomly mask
15% of tokens

Input



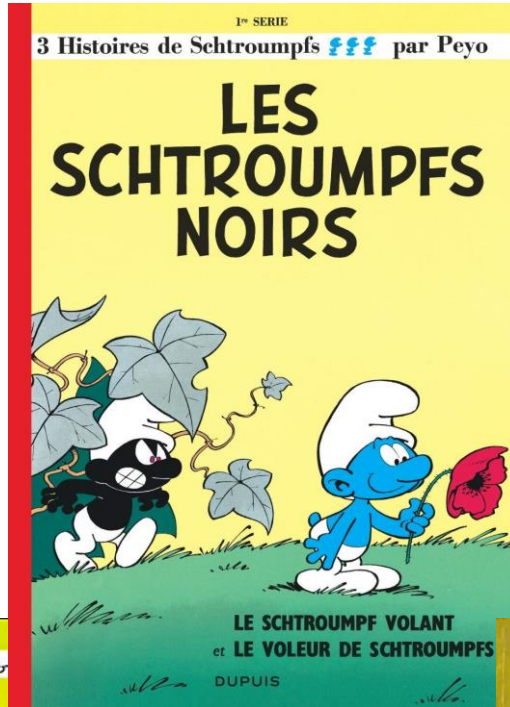
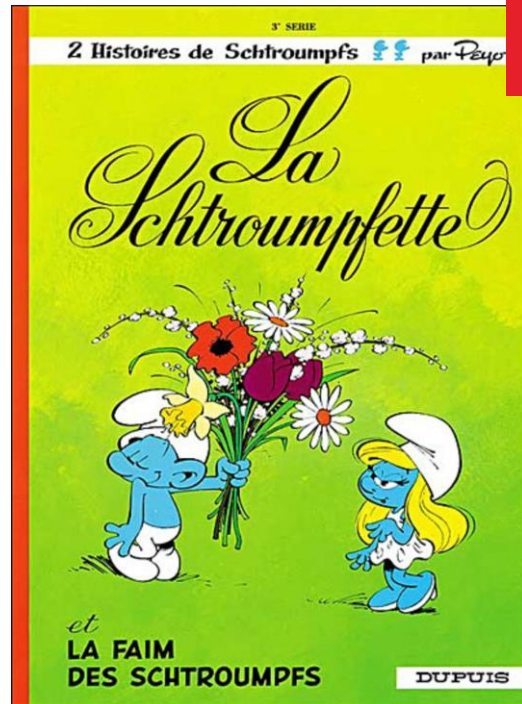
Many **LLMs** are trained through a Masked Language Modeling (MLM) task.

When **humans** read Smurfs comics, they achieve the same task with ease, naturally.



- Do LLMs perform this essential task in the same way as humans?
- Would an LLM be able to read and disambiguate Smurf language?
- What do the LLM completions of Smurf language tell us about these models?

Data



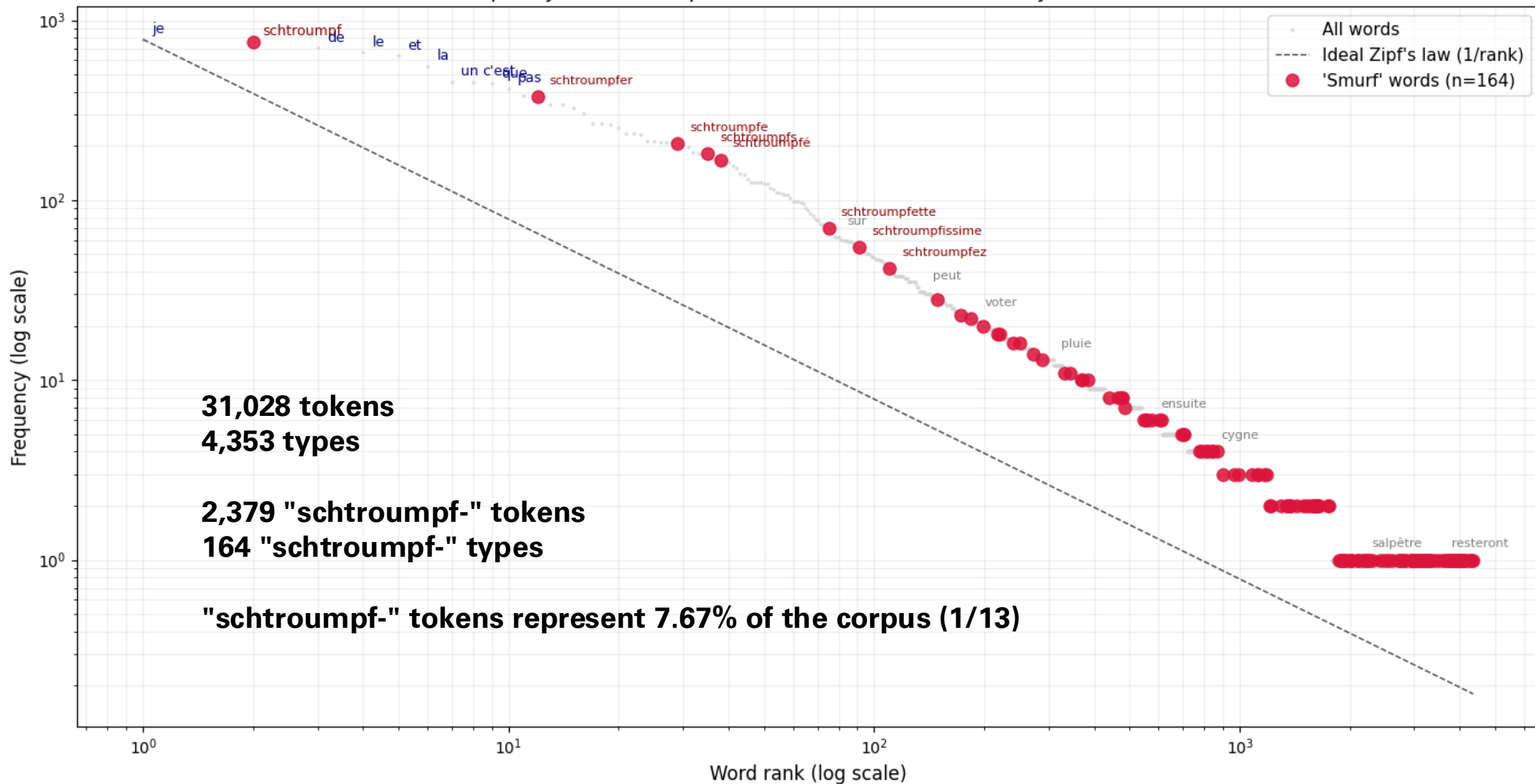
5 comic books
by Belgian
author Peyo
(Dupuis
Editions)

In their
original
French
versions

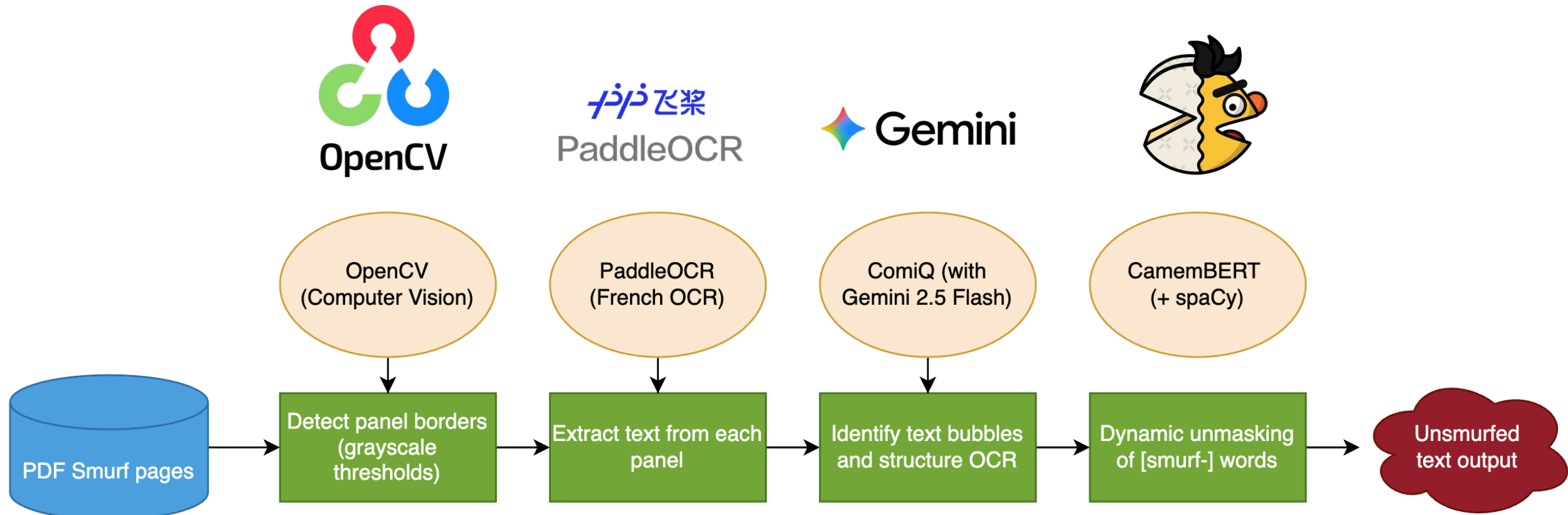
283
pages

~2400
panels

Frequency of 'schtroumpf-' words within the full vocabulary (5 albums)



Pipeline



An AI-powered Python library for comic book OCR

(Kanishq, 2025)

<https://github.com/StoneSteel27/ComiQ>







CamemBERT

French-trained BERT model
(Martin et al., 2020)

Unmasking Smurf language

1. Detect *smurf*- words and their grammatical nuances (*smurfs*, *smurfing*, *smurfed*, *smurfily*...).
2. Use CamemBERT to replace all *smurf*- words with **[MASK]** tokens.
3. Have the model predict the **most likely word** for each **[MASK]**, based on the (co)-text in its panel and in the n previous panel(s), with enforced grammatical information.



I won't be back before a few **smurfs**!
Be very **smurf** while I'm gone.



I won't be back before a few **[MASK]**!
Be very **[MASK]** while I'm gone.



I won't be back before a few **smurfs**!
Be very **smurf** while I'm gone.



I won't be back before a few **days**!
Be very **good** while I'm gone.

→ Top-1 prediction



I won't be back before a few **smurfs**!
Be very **smurf** while I'm gone.



I won't be back before a few **months**!
Be very **careful** while I'm gone.

→ Top-2 prediction

Examples



« It's important to **smurf** my popularity a little! »

Top 5: « It's important to **revive** my popularity a little! »

Top 4: « It's important to **maintain** my popularity a little! »

Top 3: « It's important to **know** my popularity a little! »

Top 2: « It's important to **augment** my popularity a little! »

Top 1: « It's important to **boost** my popularity a little! »

Top 5: « Il faut bien **relancer** un peu ma popularité! »

Top 4: « Il faut bien **entretenir** un peu ma popularité! »

Top 3: « Il faut bien **connaître** un peu ma popularité! »

Top 2: « Il faut bien **augmenter** un peu ma popularité! »

Top 1: « Il faut bien **booster** un peu ma popularité! »

Examples

« ... And by the way, I always tell the Great **Smurf** that! I say to him, "Great **Smurf**, aren't I the smurfiest of all the **Smurfs**?" And the Great **Smurf** always says yes. That's why you have to vote for me! »



« ... And by the way, I always tell the Great **God** that! I say to him, "Great **Jesus**, aren't I the smartest of all the **animals**?" And the Great **Master** always says yes. That's why you have to vote for me! »



Examples



- « What's more **smurf** than the **smurf** of my **smurf**? »
- « The **smurf** of my **smurf**! »
- « All good, come in! »
- « What's more **important** than the **content** of my **blog**? »
- « The **birth** of my **mother**! »



Conclusions & Perspectives



“Unsmurfed” is a playful take on how the integration of large language models (LLMs) into narrative art **strips both authors and readers of their creativity.**

Art-research project of **publishing** a 1:1 AI-unsmurfed version of a *Smurfs* album.

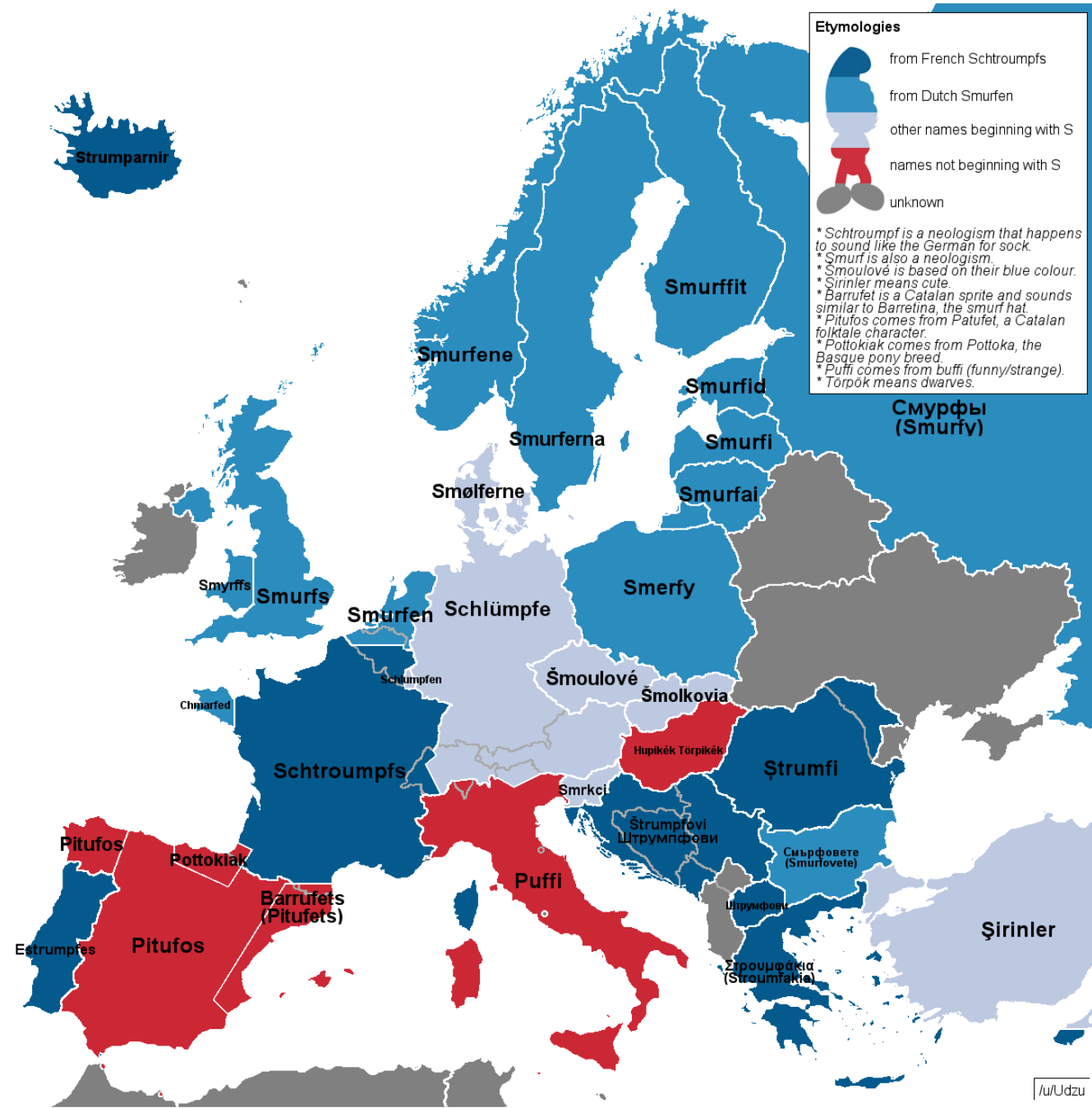
- Add **image captioning** into the pipeline.
- Extend to **other languages**.
- Ask **human annotators** to select the model’s most relevant prediction (top-1 to top-10). Is top-1 always the best?
- Have **human participants** perform the same task of “unsmurfing” and compare with LLM results.

Thank you !

escouf@mit.edu
louis.escouflaire@unamur.be

'THE SMURFS' IN DIFFERENT LANGUAGES

translations and etymologies of Peyo's little blue creatures



https://www.reddit.com/r/etymologymaps/comments/7g56ox/the_smurfs_in_european_languages_xpost_from/