Semantic Frame Extraction in Multilingual Olfactory Events

Stefano Menini (Fondazione Bruno Kessler)

LREC-COLING 2024

Task

Supervised system for multilingual olfactory information extraction covering six languages English, French, Italian, Dutch, German and Slovene

FrameNet-like structure

Detect the parts of text involved in an olfactory event.

Our system identifies:

- lexical units triggering the smell event
- semantic roles associated to the olfactory event

Approach

To identify olfactory events, we fine-tuned a transformer-based models adopting a multi-task framework.

Data

To train the models we used a manually annotated multilingual olfactory benchmark (Menini et al. 2022).

The dataset contains annotations of olfactory events and situations in texts.

Domains	10 different domains (literature, medicine, travel ....)
Time span	1620 to 1920
Languages	6 languages: English, Italian, French, German, Dutch and Slovene.

Menini, Stefano, et al. "A multilingual benchmark to capture olfactory situations over time." *Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change*. 2022.

Data

The annotation follows a FrameNet-like structure focusing on:

- **Lexical Units:** words that evoke an olfactory frame (e.g. smell, odor, perfume). Also defined *'smell words'*
- **Frame Elements:** semantic roles related to the olfactory frames.

Smell Source
Odor Carrier
Perceiver
Quality
Evoked Odorant
Location
Time
Circumstances
Effect

Data

Distribution of the Frame Elements is **not balanced**.

Smell Source and **Quality** are the most frequent.

	NL	EN	FR	DE	IT	SL
Smell Word	1,788	1,530	845	2,659	1,254	1,973
Smell Source	1,922	1,313	710	2,297	952	1,638
Quality	1,071	1,084	450	1,730	707	936
Perceiver	336	362	140	399	153	266
Circumstances	399	248	88	274	202	228
Odor Carrier	351	310	106	170	195	408
Effect	243	187	53	425	104	214
Evoked Odorant	228	91	103	258	74	285
Place	255	302	172	200	158	394
Time	127	126	49	131	119	75

On average, each lexical unit is associated with between 2.1 and 2.7 frame elements.

Olfactory Information Extraction

To extract olfactory information from text we compare two different approaches.

Single-task:

Lexical unit detection and frame element classification are considered as part of the same multiclass token classification task.

Multi-task:

The classification of lexical units and of each frame element is considered as a separate task.

Olfactory Information Extraction

Both approaches: transformer-based models fine-tuned to perform a token classification task.

We tested **monolingual** and **multilingual** variants of either BERT or RoBERTa

Configurations: Both single task and multi-task approaches are tested in two configurations:

- fine-tuning of **monolingual models on monolingual data** for each language
- fine-tuning is done on **all the six languages together** using a multilingual model that is then **tested on each language** separately.

Olfactory Information Extraction

Evaluation: **10-fold** configuration

- 80% of the lexical units and related frame elements used as training data,
- 10% for validation and 10% as test.

IOB labeling data format:

tokens in a span are marked with **Inside–Outside–Beginning** of Olfactory frame element labels.

Single-Task Classification

Token classification task:

Assign to each token in the text one out of 21 labels:

- 20 labels for being either **‘begin’** or **‘inside’** of each lexical unit / frame element (e.g. B-Quality , I-Quality)
- 1 label for **‘outside’**

Single task multiclass classification problem.

Multi-task Classification

Multi-task learning configuration:

we train the network **to learn different tasks in parallel** while using a **shared representation**, aiming at a more robust representation with less over-fitting.

In this configuration, **each task** corresponds to the classification of **a single olfactory element** (e.g. ‘Smell Source’ or ‘Quality’)

Hypothesis

simpler tasks (i.e. detecting the lexical units) can **act as auxiliary task** and share information for the classification of more difficult and ambiguous frame elements.

Results (F1)

Lan.	Approach	Traiing Data	Smell Word	Smell Source	Quality	Odour Carrier	Evoked Oforant	Loc.	Perc.	Time	Circ.	Effect
EN	Multi-task	mono	0.871	0.571	0.758	0.482	0.572	0.542	0.510	0.434	0.461	0.405
		multi	0.865	0.574	0.759	0.462	0.517	0.546	0.488	0.528	0.480	0.339
	Single-task	mono	0.867	0.525	0.703	0.392	0.293	0.368	0.410	0.304	0.266	0.140
		multi	0.881	0.530	0.698	0.392	0.359	0.410	0.390	0.309	0.261	0.138
IT	Multi-task	mono	0.871	0.559	0.800	0.343	0.564	0.439	0.241	0.613	0.259	0.246
		multi	0.887	0.575	0.801	0.382	0.625	0.304	0.240	0.642	0.309	0.201
	Single-task	mono	0.854	0.387	0.739	0.249	0.383	0.193	0.254	0.461	0.148	0.141
		multi	0.880	0.407	0.755	0.269	0.299	0.186	0.231	0.398	0.201	0.149
FR	Multi-task	mono	0.839	0.459	0.567	0.440	0.536	0.373	0.380	0.481	0.279	0.235
		multi	0.838	0.472	0.591	0.379	0.505	0.322	0.258	0.561	0.306	0.273
	Single-task	mono	0.734	0.314	0.336	0.327	0.414	0.302	0.291	0.384	0.118	0.142
		multi	0.820	0.417	0.488	0.352	0.367	0.251	0.268	0.336	0.111	0.150
NL	Multi-task	mono	0.788	0.376	0.632	0.191	0.444	0.238	0.303	0.313	0.133	0.149
		multi	0.789	0.407	0.638	0.214	0.468	0.236	0.308	0.342	0.154	0.192
	Single-task	mono	0.725	0.225	0.545	0.041	0.091	0.068	0.104	0.096	0.053	0.045
		multi	0.765	0.235	0.556	0.072	0.063	0.082	0.064	0.071	0.030	0.039
DE	Multi-task	mono	0.812	0.454	0.668	0.157	0.454	0.308	0.358	0.184	0.300	0.241
		multi	0.814	0.470	0.677	0.215	0.490	0.293	0.351	0.255	0.273	0.253
	Single-task	mono	0.778	0.273	0.479	0.186	0.150	0.092	0.164	0.162	0.040	0.036
		multi	0.797	0.268	0.443	0.141	0.086	0.092	0.133	0.095	0.031	0.030
SL	Multi-task	mono	0.707	0.501	0.525	0.320	0.506	0.401	0.355	0.280	0.153	0.151
		multi	0.695	0.442	0.491	0.273	0.445	0.368	0.245	0.214	0.103	0.132
	Single-task	mono	0.675	0.406	0.451	0.119	0.277	0.236	0.170	0.155	0.068	0.074
		multi	0.655	0.358	0.448	0.186	0.263	0.212	0.195	0.137	0.051	0.086

Results

The **multi-task** classifier **is more effective** than a single task classifier (in all the languages).

The most represented elements (smell words, smell sources and qualities) shows better results.

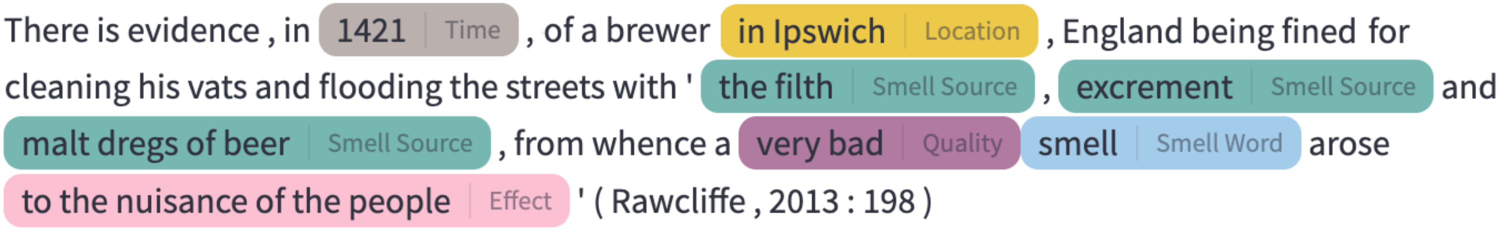
Multilingual fine-tuning helpful only on Italian, German and Dutch.

Errors not always related to wrong frame element classification.

Errors in the **boundaries** of the FE spans, e.g. predicting as smell source **‘flowers’** rather than **‘of flowers’.**

This type of errors has a larger **impact on longer FE** such as 'circumstances'

Results



There is evidence , in 1421 , of a brewer in Ipswich , England being fined for cleaning his vats and flooding the streets with ' the filth , excrement and malt dregs of beer , from whence a very bad smell arose to the nuisance of the people ' (Rawcliffe , 2013 : 198)

Conclusions

In this work we presented a system for **multilingual olfactory information** extraction.

Novel **multitask approach** to semantic frame extraction.

These **models** have been **used** by the Odeuropa project **to perform large-scale studies on olfactory language** on books from the 17th to begin of the 20th century.

Reconstruct the smells that are part of our olfactory heritage across different centuries and countries.

The **models are available** at https://zenodo.org/records/10598306

Thanks!