



## UvA-DARE (Digital Academic Repository)

### Towards Olfactory Information Extraction from Text: A Case Study on Detecting Smell Experiences in Novels

Brate, R.; Groth, P.; van Erp, M.

**Publication date**

2020

**Document Version**

Other version

**Published in**

The 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature

**License**

CC BY

[Link to publication](#)

**Citation for published version (APA):**

Brate, R., Groth, P., & van Erp, M. (2020). Towards Olfactory Information Extraction from Text: A Case Study on Detecting Smell Experiences in Novels. In S. DeGaetano, A. Kazantseva, N. Reiter, & S. Szpakowicz (Eds.), *The 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature: Co-located with the 28th International Conference on Computational Linguistics COLING'2020 : COLING 2020 : proceedings : December 12, 2020, Barcelona, Spain, (Online)* (pp. 147-155). International Committee on Computational Linguistics. <https://www.aclweb.org/anthology/2020.latechclfl-1.18>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*

# Towards Olfactory Information Extraction from Text: A Case Study on Detecting Smell Experiences in Novels

**Ryan Brate** and **Paul Groth**

University of Amsterdam  
Amsterdam, the Netherlands  
r.brates@gmail.com  
p.t.groth@uva.nl

**Marieke van Erp**

KNAW Humanities Cluster  
Digital Humanities Lab  
Amsterdam, the Netherlands  
marieke.van.erp@dh.huc.knaw.nl

## Abstract

Environmental factors determine the smells we perceive, but societal factors shape the importance, sentiment and biases we give to them. Descriptions of smells in text, or as we call them ‘smell experiences’, offer a window into these factors, but they must first be identified. To the best of our knowledge, no tool exists to extract references to smell experiences from text. In this paper, we present two variations on a semi-supervised approach to identify smell experiences in English literature. The combined set of patterns from both implementations offer significantly better performance than a keyword-based baseline.

## 1 Introduction

We rely on our senses: touch, taste, hearing, sight and smell; to complement one another in shaping our interpretation of our environment. There is shifting historical relevance placed on smell - its worthiness for attention, its association with social standing, lifestyle, emotion, science and superstitions, and other topical associations shifting with time (Vroon et al., 1997). English language vocabulary specific to the description of smell experiences is not expansive, and, to the best of our knowledge, language technology to identify references to smell in text even less so. A topic search in the Cambridge Dictionary online of words categorised as relating to smells and smelling<sup>1</sup> returns fewer than 30 words, that are predominantly concerned with *intensity* or *sentiment* such as *fetid* and *reek*. Other characteristics of smell are instead often described in terms of reference smell sources as similes such as *There is a strange unwholesome smell upon the room, like mildewed corduroys*.

In this paper, we present a dataset of annotated references to smell, which we call ‘smell experiences’ in literature, as well as a first approach and experiments to automatically recognise references to smells in texts.

The remainder of this paper is organised as follows. In Section 2, we discuss related work. In Section 3, we describe our corpus and its creation process. In Section 4, we present our extraction approach and experiments. In Section 5, we discuss the results. We conclude with Section 6 in which we present our conclusions and directions for future work. Our data is available at <http://doi.org/10.5281/zenodo.4199996> and the code to the experiments is available at <https://github.com/DHLab-nl/Detecting-Smell-Experiences-in-Novels>. This work is a preliminary result of the Odeuropa project which will commence formally in January 2021: <https://odeuropa.eu/>.

## 2 Related Work

The cultural significance of smells is a niche topic in the humanities domain, but one that has recently gained more interest with a translation of Muchembled’s 2017 *La Civilisation des odeurs (XVIe siècle-début XIXe siècle)* to English (Muchembled, 2020) and Barwich’s *Smellosophy: what the nose tells the*

<sup>1</sup><https://dictionary.cambridge.org/topics/senses-and-sounds/smells-and-smelling/> Accessed: 5 August 2020

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>.

*mind*(Barwich, 2020) being reviewed in mainstream media.<sup>2</sup> While historians interested in olfaction such as (Tullett, 2019) analyse textual accounts of experienced smells, (computational) linguistic analysis of smell experiences (at least for English) has received little attention.

This lack of attention may be due to the fact that Western languages such as English and Dutch do not contain rich vocabularies for describing odorants as opposed to some other languages. In (Majid and Burenhult, 2014) Jahai and English speakers and in (Majid et al., 2018) Jahai and Dutch speakers were contrasted in describing a range of odorants. The Jahai, a group of nomadic hunter-gatherers in Malaysia, have over a dozen terms to describe odours. In the experiment, the Jahai speakers were both more consistent and greatly more controlled in the terms they used than the English and Dutch speakers relying on reference smell sources.

There is evidence that consistency in the use of smell sources in English can be conditioned. Croijmans and Majid (2016) examined the accuracy and consistency of wine experts, coffee experts and people with no expertise in identifying smells. No group was better at naming smells outside of the domain of their expertise. However, it was apparent the domain experts had developed a toolkit of common smell sources they frequently drew upon. An experiment on predicting properties of wines from experts' wine reviews confirms this (Hendrickx et al., 2016).

Sensorial lexicons have been developed that include terms related to smell (cf. Tekiroğlu et al. (2014)). However, as the initial seed words to bootstrap the lexicon for smell are limited, the olfactory clusters in such lexicons are by extension also limited. A different approach is taken in (Kiela et al., 2015), where terms related to smells are connected to their chemical compounds. While this is useful for translating olfactory information from the chemistry domain to language, it does not aid us in recognising the wide variety of expressions used in texts to describe smell experiences.

### 3 Literary Smell Dataset

To begin to tackle the problem of recognising smell experiences, we created a unique dataset focused on such expressions in literary texts. Specifically, by selecting texts from Project Gutenberg<sup>3</sup> that had the highest rate of occurrence of keywords derived from Table 1, we assembled a set 139 English literary texts. Each sentence in this collection was tokenised using NLTK<sup>4</sup>, and POS-tagged and syntactically parsed using spaCy (Montani et al., 2020).<sup>5</sup> We split the set into three datasets: a harvesting dataset of 99 texts; a validation dataset of 20 texts; and an evaluation dataset of 20 texts.

From the evaluation dataset, a gold standard of manually labelled sentences was assembled consisting of seven documents, each of 100 randomly assigned sentences from multiple literary texts and annotated by a single annotator. To evaluate the inter-annotator agreement between the three annotators, one additional document consisting of 100 randomly selected sentences was annotated independently by the annotators.

Despite having chosen the harvesting, validation and extract sets for their high frequency of Table 1 derived keywords, on average only approximately 1 in 100 sentences contain a keyword. Thus, assuming that smell experiences typically contain a keyword, the evaluation set contains smell experiences in very low proportion. A gold standard set of extracts was sampled from the evaluation dataset to ensure a substantial number of smell extracts. The evaluation set was scanned for the high smell association keywords derived from Table 1. 80% of the sentences in the gold standard documents contain a Table 1 related word, the remaining 20% were randomly sampled. There is no overlap between documents, or redundancy within a document.

Annotators were asked to highlight and annotated spans according to the following criteria:<sup>6</sup>

---

<sup>2</sup>cf. <https://www.spectator.co.uk/article/where-are-the-scents-of-yesterday-entire-countries-have-lost-their-distinctive-smell> ; <https://slate.com/culture/2020/07/smells-history-book-review-france-plague-farts.html> ; <https://www.wsj.com/articles/smells-and-smellosophy-review-what-the-nose-knows-11594391739>

<sup>3</sup><https://www.gutenberg.org/>

<sup>4</sup><https://nltk.org>

<sup>5</sup><https://spacy.io/>

<sup>6</sup>The full annotation guidelines can be found on our Github page

<b>smell-only (in all contexts)</b>	<b>smell-only (in sensory contexts)</b>	<b>smell and taste only</b>
<i>odour(N), odorous(A)</i> <i>malodorous(A)</i> <b><i>fetid(A), foetid(A)</i></b> <i>whiffy(A)</i> smell(N,V), <i>scent(N)</i> <i>smelly(A)</i> <i>scented(A)</i> <i>perfume(N)</i> <i>aroma(N), aromatic(N)</i> <i>fraganced(A)</i> <b><i>petrichor(N)</i></b> <b><i>musty(A), musky(A)</i></b>	<i>fragrance(N)</i> <b><i>musk(N)</i></b> <b><i>fusty(A), frowsty(A)</i></b> <i>ripe(A), ripeness(N)</i> <i>reek(N,V), stink(N,V)</i> <i>stench(N), niff(N)</i> sniff(V), <b><i>piney(A)</i></b> waft(N,V), <i>stinky(A)</i> <i>whiff(N)</i> ,	<i>pungent(A)</i> <i>pungency(N)</i> <i>pungently(ADV)</i> <i>savour(N,V)</i> <b><i>acrid(A)</i></b>

Note 1: A,N,V denotes adjectives, nouns and verbs, respectively

Note 2: underlined: words with smell strength connotations

Note 3: *italicised*: words with sentiment associations

Note 4: **bold face**: describes characteristics beyond strength or sentiment

Table 1: Results from Cambridge Dictionary ‘smells and smelling’ SMART Thesaurus search

Annotation tag	Number of corresponding text spans
‘d’	533
‘o’	129
‘v’	186
‘s’	34
‘a’	37
‘n’	75

Table 2: Number of text spans by annotation tag

- ‘d’. A smell description; e.g., ‘An odd fragrance, a smell of damp plaster, wafted from the new house to his senses’.  
The inherent subjectivity of when precisely a smell experience becomes a description is left to the perception of the annotator.
- ‘o’. A smell alluded to without expansion of its characteristics; e.g., ‘A fragrance wafted from the new house to his senses’.
- ‘v’. Any verb in the sentence which is associated with smell generally or with a specific smell experience within the extract; e.g., ‘An odd fragrance wafted from the new house to his senses’;
- ‘s’. Sense of smell alluded to directly; e.g., ‘An odd fragrance, wafted from the new house to his senses’.

Additionally, two documents (of the aforementioned group of 7) were annotated with an additional set of tags:

- ‘a’. An adjective being applied to the smell alluded to; e.g., ‘An odd fragrance, a smell of damp plaster, wafted from the new house to his senses’.
- ‘n’. The noun group referred to as a smell source; E.g., ‘An odd fragrance, a smell of damp plaster, wafted from the new house to his senses’.

The inter-annotator agreement is measured using Cohen’s Kappa (Cohen, 1960) per single gold standard document in a pairwise fashion. Specifically:

- The level of agreement with respect to those sentences which were tagged as *a smell experience*, by one or more annotators, i.e., those sentences with a text span annotated with either ‘d’ or ‘o’.
- The level of agreement with respect to those sentences which were tagged as *a smell description*, by one or more annotators, i.e., those sentences with a text span annotated with ‘d’.

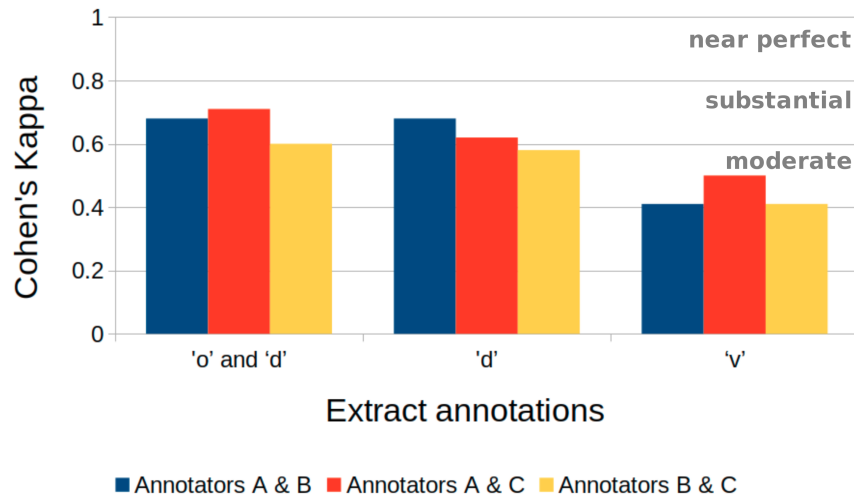


Figure 1: Cohen's Kappa scores of pairwise annotator agreement

- The level of agreement of those verbs within sentences what were tagged with 'v', i.e., verbs the annotator perceives as being related to smell generally, or in the context of the sentence.

Cohen's Kappa is a metric used to measure pairwise inter-annotator agreement. A Cohen's Kappa of 0, denotes an even probability of agreement. Landis and Koch (1977) denote a Cohen's Kappa score of 0.41 to 0.60, and .61 - 0.80 as representing *moderate* and *substantial* strength of agreement, respectively. A score of 0.81 to 1.0 can be considered as near perfect agreement.

Figure 1 shows the pairwise annotator agreement with regards the single gold standard document of 100 extracts, annotated by multiple annotators. All annotators are in *substantial* agreement in identifying all and any extracts that allude to smell, i.e., all spans labelled 'o' or 'd'. Annotators are generally in substantial agreement in identifying extracts which *describe* smell experiences, i.e., extracts with a spans labelled 'd'. Although, one pair of annotators are at the very upper end of *moderate* agreement only. Finally, in identifying verbs either highly associated with smell, or associated with smell in the context, there was only *moderate* annotator agreement.

It is reasonable to assume that human error, i.e., misreading, miscomprehending or simply skipping an extract, played some role in the observed imperfect inter-annotator agreement scores. Instances of likely human error are apparent on inspection of the gold standard, in those instances where there is arguably little room for personal subjectivity, for example, one of the three annotators did not attribute either a 'd' or 'o' tag to the span: *There was a smell of decaying leaves and of dog.*

However, a number of extracts clearly demonstrate the potential for subjectivity in smell experience interpretation, as a source for annotator disagreement, for instance, in the following extract, each of the three annotators attributed 'd', 'o' and no tag to it, respectively: *Seated beside her aromatic rest, In silence musing on her loveliness, Her knight and troubadour.*

In the following extract, one of three annotators tagged it as 'o', the other two presumably thought it sufficiently descriptive to be tagged 'd': *Between each pair of columns an elegant table of cedar bore on its platform a bronze cup filled with scented oil, from which the cotton wicks drew an odoriferous light.*

## 4 Extraction Approaches

As there is no known corpus annotated with smell expressions available to train a supervised smell language recogniser, we investigate pattern-based approaches to recognising such expressions. We base our approach on the concepts put forward in approaches such as the detection of hypernym-hyponym pairs (Hearst, 2000), e.g., *car is a type of vehicle*, and iterative bootstrapping (Brin, 1998) to detect author-book title pairs.

We start with seed features that are matched against the harvesting dataset. The resulting matches are then manually evaluated to identify new linguistic patterns, which are in-turn used to identify new features for the next bootstrapping round. The process can be seeded by introducing known features into the lexicon, or known patterns into the pattern set. In both (Hearst, 2000) and (Brin, 1998), this process was used to assemble a lexicon, the pattern set being effectively a by-product of the process. However, for our purposes, it is the set of patterns that is of interest, and their potential use to identify smell experiences.

The textual features targeted in prior work represented distinct real world concepts linked through a conceptual relationships. The expression of smell (in English) does not conform to a natural set of paired entities reflecting a relationship in the same way as authors and book titles do. However, surrounding adjectives, and verb and noun groups help characterise the smell experience and can therefore be targeted as complements.

A difficulty of identifying smell references over hypernym-hyponym or author-book title pairs is that there is no inherent relationship defining the number of coincident complements necessary. Instead it is a question of how restrictive we wish to make the criteria for matching sentences in the harvesting dataset. The greater the number of complements, the fewer extracts we can expect to retrieve. Thus, a too restrictive choice may result in stalling bootstrapping process. Conversely, if the choice is not restrictive enough and too many extracts unrelated to smell are returned, the process becomes uninformative. In evaluating smell related vocabulary, (Iatropoulos et al., 2018) concluded that the most commonplace words used in smell descriptions are those that could apply to a wide range of sensory contexts. This makes intuitive sense given the heavy reliance on reference smell sources to define smell characteristics. Hence, single complements are not targeted, as being too relevant outside of smell contexts. In this pilot project, we therefore focused on at least two complements in a pattern.

A basic assumption in our approaches is that found complements are indicative of the presence of a smell expression. We therefore aim to detect the following types of smell expression complements:

1. **Adjectives** modifying the smell experience perception, and the coincident **noun group** acting as the reference smell (one or more nouns modified by adjectives); E.g., ‘An *odd* fragrance, a smell of *damp plaster*, wafted from the new house to his senses’, where ‘odd’ and ‘damp plaster’ are the adjective and noun group, respectively. Thus, it is assumed that ‘odd’ and ‘damp plaster’, being coincident in defining this smell experience, are indicative of a smell experience when both present in other extracts.
2. A **noun group** acting as the reference smell, and the coincident **verb group** (adverbs, verbs, associated prepositions) describing how the smell moves; E.g., ‘An odd fragrance, a smell of *damp plaster*, *wafted from* the new house to his senses’, where ‘damp plaster’ and ‘wafted from’ are the noun group and verb group, respectively. Thus, it is assumed that ‘damp plaster’ and ‘wafted from’, being coincident in defining this smell experience, are indicative of a smell experience when both present in other extracts.

### 4.1 Approach 1: Targeting Adjective and Noun Groups

To capture and enable pattern matching with parts of speech, synonyms and flexible groupings of these in text, we start with words derived from Table 1 assembled in synonym groups. For example, *<smell\_noun>* is defined to match against ‘aroma’, ‘odour’, ‘scent’, ‘perfume’ etc. The patterns further contain part of speech chunks that can match various tokens. These chunks were defined, and updated in

process, based on observed language patterns from harvesting set extracts. Listing 1 is an example of a high-level pattern representation.

```
[<adj>] <smell_noun> _ , -* _of_ <pronoun>* [<noun> { _of_ <noun> }*]
```

Listing 1: Example identified pattern

It will match the boldfaced adjective and noun groups in for example *the **warm** aroma of **multitudinous** exotics* and *the **ammoniacal** smell of **the** horses*

## 4.2 Approach 2: Targeting Verb and Noun Groups

As in the Adjectives & Nouns approach, we start with the same high-smell association seed lexicon entry, *\_aroma\_NOUN*, as it results in a very manageable number of bootstrapped extracts. Based on the resulting extracts, new patterns are hypothesised: e.g., from the extract ‘*the aroma of the **newly-sawn timber and saw dust mingled** in the air*’, we may hypothesise the pattern *<smell\_noun> \_of|like\_ \_DET\* <pronoun>\* [<noun> { \_of\_ <noun> }\*]*.

This approach also includes a validation loop. As patterns with low precision risk introducing a large volume of vocabulary into the lexicon which is unrelated to smell. The validation loop is an attempt to limit this, by estimating pattern precision and setting a minimum acceptance threshold by taking 10 extracts per pattern and manually tagging them as *true positive*, *false positive*, or *unknown*. Patterns that pass a validation threshold of 0.7 estimated precision are accepted. If no example of a pattern is present in the validation set, it is removed.

Two variations of the patterns are retained, *identification patterns* and *extraction patterns*. *Extraction patterns* target the previously discussed feature pairs, such to introduce new vocabulary into the lexicon. Thus, extraction patterns are used to drive each iterative cycle. For example, based on the preceding hypothesised pattern, e.g., *<smell\_noun> \_of|like\_ \_DET\* <pronoun>\* [<noun> { \_of\_ <noun> }\*]*. *Identification patterns* are a superset of the extraction pattern set, and are concerned with matching any and all smell experiences, not just matching feature pairs. Several identification patterns may be derived from a single hypothesised extraction pattern, expressing the potential variation in matching smell experiences. The identification pattern set is our desired output from the iterative bootstrapping process. In the final step of the approach, the extraction patterns are applied to the harvesting set, and targeted complements are collected and added to the lexicon.

## 5 Evaluation and Discussion

Four complete cycles of the Adjectives & Nouns approach and three complete cycles of the Verbs & Nouns approach were performed.

The Adjectives & Nouns approach identified 48 new identification patterns as seen in Table 3. The majority of these patterns involve Table 1 derived words such as *<adj>\* compound\_ <smell\_noun>* matching *a delightful forest **aroma***, and *<adj> \_with\_ \_DET\* <pronoun>\* <smell\_noun> \_of\_ <pronoun>\* <verb> <noun> { \_of\_ <noun> }\** matching *heavy with the **smell** of freshly turned soil*.

Additionally, a small number of patterns identified do not involve the Table 1 vocabulary, such as *<adj>\* \_breath|breaths\_ \_of\_ <pronoun>\* <noun> { \_of\_ <noun> }\** matching ‘*...and inhale the sweet breath of autumn, which was borne upon gentle gales*’ and ‘*\_air\_\* \_ , \_sweet\_ \_with\_ <pronoun>\* <noun> { \_of\_ <noun> }\**’ matching ‘*the mild air, sweet with fading leaves and bracken*’.

The Verbs & Nouns approach identified 31 new identification patterns as show in Table 4. Again a majority involve Table 1 derived words, for example ‘*<smell\_noun> \_of—like\_ \_DET\* <pronoun>\* <noun> { \_of\_ <noun> }\* <verb> prep\_*’ matching phrases such as *the aroma of **new-sawn timber and sawdust mingled with...***. The single example of lexico-syntactic pattern not involving Table 1 derived words, introduced the ‘*incense*’ as synonymous with smell: ‘*\_fumes\_ \_of\_ \_incense\_ { \_of\_ <noun> }\* \_DET\* <verb> prep\_*’ matching for example ‘*the heavy fumes of incense rose up*’

Figure 2 shows the relative precision-recall performance of the group predictions, in respect of the gold standard, with regards the pattern sets of: approach 1; approach 2; and both combined.

Cycle	Lexicon entries	New (unseen) extracts	Hypothesised (new) patterns	New id. patterns/ New ex. patterns
0	1**	<b>91</b>	15	15 / 13
1	519	<b>1,509</b>	28	26 / 22
2***	874	<b>4,216</b>	14	13 / 8
3	463	<b>464</b>	4 ****	4 / 4

\*\*Seed word: *\_aroma\_NOUN*

\*\*\* sifted with Table 1 word search due to high volume

\*\*\*\* not subject to validation as cycles stopped

Note 1: Each lexicon (pair) entry is unique, and each extract is unique

Table 3: Record of iterative cycles outcomes for Approach 1: targeting coincident adjectives modifying the smell, and noun group reference smells

Cycle	Lexicon entries	New (unseen) extracts	Hypothesised (new) patterns	New id. patterns/ New ex. patterns
0	1**	<b>91</b>	11	10 / 9
1***	530	<b>2,968</b>	12	10 / 9
2	565	<b>1,030</b>	11 ****	11 / 8

\*\*Seed word: *\_aroma\_NOUN*

\*\*\* sifted with Table 1 word search due to high volume

\*\*\*\* not subject to validation as cycles stopped

Note: Each lexicon (pair) entry is unique, and each extract is unique

Table 4: Record of iterative cycles outcomes for Approach 2: targeting coincident verb groups associated with the smell experience and noun group reference smells

It is apparent that the Adjectives & Nouns approach (red) consistently outperformed the pattern set of Verbs & Nouns approach (black); and that they target complementary smell expressions (combined approach, green). The relative performance of the combined approach with that of the Verbs & Nouns approach. The combined approach is significantly better with a 5% significance level where patterns have a precision greater than 0.75. The combined pattern set *significantly* has superior recall performance at corresponding precision cut-offs. Thus, we can conclude that targeting different feature pairs did result in patterns that target different smell extracts in the gold standard set.

## 6 Conclusions and Future Work

In this paper, we set out to identify smell experiences in English literary texts. We created a gold standard dataset and our experiments demonstrated that iterative bootstrapping techniques can be used to identify smell experiences in text.

Whilst the overwhelming majority of identified patterns involved keywords and phrases with a high smell association, the implementations revealed a number of new phrases used in smell contexts. Furthermore, we showed that at the very highest levels of precision, pattern group identification of smell experiences offers significantly better recall rates than a keyword search.

The focus of application of iterative bootstrapping implementations was centred on single sentence extracts for English literary texts. It would be interesting to explore the applicability of this semi-supervised method to other textual contexts, and longer-distance relationships spanning multiple sentences. Additionally, it would be informative to explore the influence of tweaking the implementation's parameters and approaches, including the pre-processing steps such as instead of using a shallow parser and experimenting with dependency grammars or semantic role labelling. With regards the precision-recall performance of the resulting pattern sets, such as exploring the influence of different seed words on the process outcomes, investigate the impact of a higher validation precision threshold for the number, and quality, of extracts returned each cycle and the corresponding identified patterns.

Our experiments demonstrated the potential use of patterns for identifying textual smell experiences in text. However, the number of extracts, and the quality of extracts in terms of smell experience density was identified as source of inefficiency which would benefit from being addressed further. The statistics presented in Table 3, for example, show the explosion in the number extracts for manual examination, in



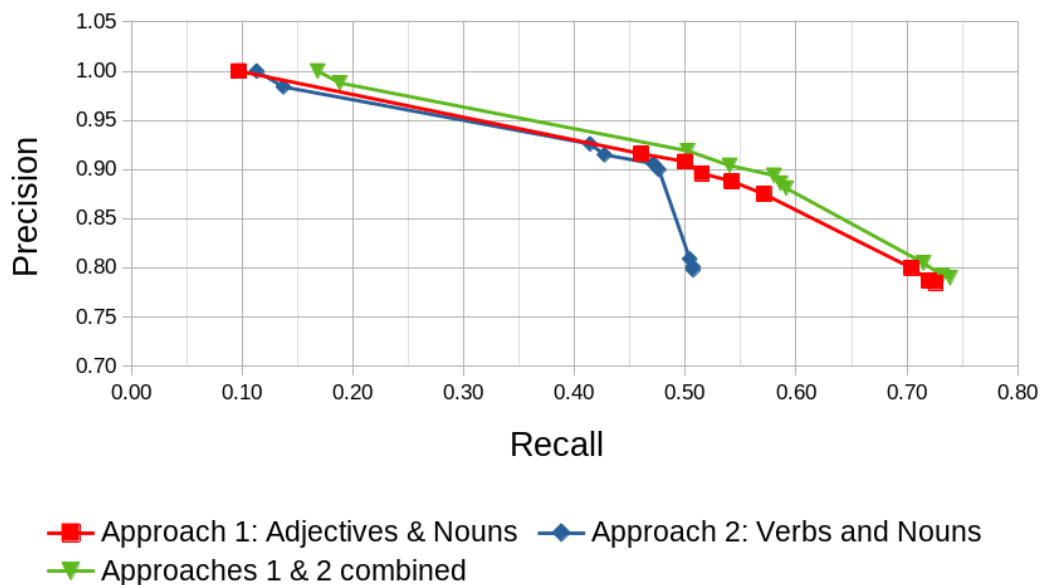


Figure 2: Group prediction precision-recall performance of the pattern sets of: approach 1; approach 2; and approach 1 and 2 combined.

certain cycles, which correspond to only comparatively few new patterns being identified.

The observed high volume of low smell experience extracts may be an inherent challenge of smell experiences relying on vocabulary which is equally, or more so, applicable in other sensory contexts. I.e., the targeting and adding to the lexicon of words with a low smell association, resulting in poorer quality extracts. However, there are a number of clear, possible avenues to explore such as increasing the validation set size, ensuring more accurate precision estimates thus improving the level of smell association of lexicon entries on average and exploring the effects of using more coincident features simultaneously, i.e., pairs were selected on the basis that if one feature alone was weakly associated with smell, two together may improve the association. More coincident features may further improve the likelihood of an extract relating to smell.

Finally, consideration as to degree of agreement between people in their interpretation, suggested a less than perfect agreement not only of the subtly nuanced aspects of smell experiences, but even at recognition of smell experiences as a broad classification. On inspection of annotations, however, it is unclear how many of these were genuine discrepancies in terms of subjective perceptions. More annotators, supported by a more comprehensive approach to tagging, e.g., requiring the annotators to explicitly note their reasoning and deliberations would help. This would offer a window into the mind of the annotators, reinforcing any conclusions that may be drawn.

Whilst there is extensive scope for further study, the experiments in this paper have shown that smell experiences can be identified, and smell experience features can be extracted from text providing a useful foundation for understanding smell usage within literature.

## Acknowledgements

The authors would like to thank the annotators for taking the time to think about smelly language and tag multiple sets of sentences for us.

## References

- Ann-Sophie Barwich. 2020. *Smellosophy: What the Nose tells the Mind*. Harvard University Press.
- Sergey Brin. 1998. Extracting patterns and relations from the world wide web. In *The World Wide Web and Databases (WebDB 1998)*, 06.
- Jacob Cohen. 1960. Kappa: Coefficient of concordance. *Educational and Psychological Measurement*, 20(37).
- Ilja Croijmans and Asifa Majid. 2016. Not all flavor expertise is equal: The language of wine and coffee experts. *PLoS ONE*, 11(6):e0155845.
- Marti Hearst. 2000. Automatic acquisition of hyponyms from large text corpora. *Proceedings of the 14th Conference on Computational Linguistics (CoLing)*, 05.
- Iris Hendrickx, Els Lefever, Ilja Croijmans, Asifa Majid, and Antal van den Bosch. 2016. Very quaffable and great fun: Applying NLP to wine reviews. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 306–312, Berlin, Germany, August. Association for Computational Linguistics.
- Georgios Iatropoulos, Pawel Herman, Anders Lansner, Jussi Karlgren, Maria Larsson, and Jonas K. Olofsson. 2018. The language of smell: Connecting linguistic and psychophysical properties of odor descriptors. *Cognition*, 178:37 – 49.
- Douwe Kiela, Luana Bulat, and Stephen Clark. 2015. Grounding semantics in olfactory perception. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 231–236, Beijing, China, July. Association for Computational Linguistics.
- J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.
- Asifa Majid and Niclas Burenhult. 2014. Odors are expressible in language, as long as you speak the right language. *Cognition*, 130(2):266–270.
- Asifa Majid, Niclas Burenhult, Marcus Stensmyr, Josje De Valk, and Bill S Hansson. 2018. Olfactory language and abstraction across cultures. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1752):20170139.
- Ines Montani, Matthew Honnibal, Matthew Honnibal, Sofie Van Landeghem, Henning Peters, Adriane Boyd, Maxim Samsonov, Jim Geovedi, Jim Regan, György Orosz, Paul O’Leary McCann, Søren Lind Kristiansen, Duygu Altınok, Roman, Leander Fiedler, Grégory Howard, Explosion Bot, Sam Bozek, Wannaphong Phatthiyaphaibun, Mark Amery, Björn Böing, Pradeep Kumar Tippa, Yohei Tamura, Leif Uwe Vogelsang, Ramanan Balakrishnan, Vadim Mazaev, GregDubbin, jeannefukumar, Jens Dahl Møllerhøj, and Avadh Patel. 2020. explosion/spaCy: v2.3.2: Improved Korean tokenizer speed, experimental character-based pretraining and bug fixes, July.
- Robert Muchembled. 2020. *Smells: A Cultural History of Odours in Early Modern Times*. Polity.
- Serra Sinem Tekiroğlu, Gözde Özbal, and Carlo Strapparava. 2014. A computational approach to generate a sensorial lexicon. In *Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon (CogALex)*, pages 114–125, Dublin, Ireland, August. Association for Computational Linguistics and Dublin City University.
- William Tullett. 2019. *Smell in Eighteenth-century England: A Social Sense*. Oxford University Press.
- Piet Vroon, Anton Amerongen, Hans, and Vries. 1997. *Smell: The secret seducer*. Farrar, Straus and Giroux.