

# FrameNet-like Annotation of Olfactory Information in Texts

Sara Tonelli, Stefano Menini

Fondazione Bruno Kessler

Via Sommarive 18, Trento (Italy)

[satonelli|menini]@fbk.eu

## Abstract

Although olfactory references play a crucial role in our cultural memory, only few works in NLP have tried to capture them from a computational perspective. Currently, the main challenge is not much the development of technological components for olfactory information extraction, given recent advances in semantic processing and natural language understanding, but rather the lack of a theoretical framework to capture this information from a linguistic point of view, as a preliminary step towards the development of automated systems. Therefore, in this work we present the annotation guidelines, developed with the help of history scholars and domain experts, aimed at capturing all the relevant elements involved in olfactory situations or events described in texts. These guidelines have been inspired by FrameNet annotation, but underwent some adaptations, which are detailed in this paper. Furthermore, we present a case study concerning the annotation of olfactory situations in English historical travel writings describing trips to Italy. An analysis of the most frequent role fillers show that olfactory descriptions pertain to some typical domains such as religion, food, nature, ancient past, poor sanitation, all supporting the creation of a stereotypical imagery related to Italy. On the other hand, positive feelings triggered by smells are prevalent, and contribute to framing travels to Italy as an exciting experience involving all senses.

## 1 Introduction

In the humanities and social sciences, the senses, which have traditionally received little attention from researchers, are now very high on the academic agenda (Smith, 2007). In fact, in the last two decades, scholarly attention has shifted away from the visual and textual to the embodied and multi-sensory, following a so-called ‘sensorial revolution’ (Howes, 2006; Classen, 1999). In the field of history, for example, recent works have focused

on describing the meaning of odours in particular places and times (Dugan, 2011), or on the role played by smells in signalling identity, community and otherness in the past – especially race (Smith, 2006; Tullett, 2016). This turn, however, has received little attention from the NLP community, probably due to a number of inherent challenges related to the intangible nature of scents and odours. The few existing works in the field have mainly focused on building resources aimed at capturing and modelling sensory vocabularies (Tekiroğlu et al., 2014b) and analysing how the different senses interfere from a lexical point of view (Winter, 2019). This lack of attention may be partly explained by the fact that Western languages, which are prevalent in NLP studies, do not contain rich vocabularies for describing odorants as opposed to other senses (Majid and Burenhult, 2014). The only exception in this direction is the work in (Brate et al., 2020), where the authors present two semi-supervised approaches to identify smell experiences in English literature after annotating a gold standard of 700 sentences with smell-related information.

In this work, we present a more comprehensive scheme for the annotation of olfactory information, which has been developed within the ODEUROPA H2020 Project<sup>1</sup>. The goal of this effort is to capture smell events and situations in texts and manually label the main participants in the scene. Our guidelines are inspired by frame semantics (Fillmore and Baker, 2001) and the FrameNet annotation project (Ruppenhofer et al., 2006), although we carry out some modifications both to the olfactory-related frames and to the annotation practice. This work will guide a subsequent annotation task, which will lead to the creation of a multi-domain multilingual benchmark of historical texts annotated with olfactory information. The guidelines and the benchmark represent also the backbone upon which an

<sup>1</sup><https://odeuropa.eu/>

automated system for olfactory information extraction will be built. As a preliminary study, we apply the annotation scheme to a set of English travel narratives describing trips to Italy, and present the results obtained by analysing the most frequent role fillers and grouping them by domains.

## 2 Related work

Only very little research within the NLP community has dealt with sensory information extraction. Most works have focused on the creation of structured resources to capture the sensory domain, automatically deriving them from WordNet (Tekiroğlu et al., 2014b,a). Other works have dealt with the automated analysis of texts related to specific domains like wine reviews, where olfaction plays a central role (Lefever et al., 2018). Other studies have focused on synaesthetic aspects of language, starting from a controlled lexicon of perception (Lievers and Huang, 2016). To our knowledge, only the work by (Brate et al., 2020) proposes both a simple annotation scheme to capture smelly experiences and two semi-supervised approaches to automatically replicate this annotation. Another line of research has addressed so-called urban smellscapes, i.e. how modern cities can be described from an olfactory point of view. More specifically, (Quercia et al., 2015, 2016) obtain descriptions of different urban areas by asking annotators to walk around cities and take note of the smell characterising different places. Such descriptions are then combined with social media posts about the same places, allowing the authors to build an olfactory representation of different cities and categorise urban smells into odour wheels.

## 3 FrameNet and Olfactory Information

Our annotation of olfactory information is inspired by frame semantics (Fillmore and Baker, 2001), a theory which has been implemented through the FrameNet annotation project (Ruppenhofer et al., 2006)<sup>2</sup>, whose goal is to capture situations and events present in texts. In FrameNet, events and situations are so-called *frames*, which from a cognitive point of view, are defined as components of the internal model of the world that a language-user has created by interpreting his/her environment (Fillmore, 1976). Frames are used as synonyms for schemata, semantic memory or scenarios, and represent the perceptual base of our knowledge that is

necessary to understand the meaning of words. For example, the predicate ‘marry’ refers to a scenario where two partners get involved in some kind of social relationship.

According to frame semantics, a frame includes two main components: *lexical units* (LUs) and *frame elements* (FEs). The former are words, multi-words or idiomatic expressions that evoke a specific frame, while the latter are frame-specific semantic roles that, in case of verbal LUs, are usually realized by the syntactic dependents of the verb. For example, the *Commerce pay* frame includes as lexical units ‘pay’, ‘payment’, ‘disburse’, ‘disbursement’, ‘shell out’, and has the following frame elements: Buyer, Goods, Money, Rate, Seller.

FrameNet aims at being a general-purpose resource, capturing all possible situations and events that may happen in real life. Therefore, in the last FrameNet version, more than 1,200 frames are listed. Since in our study we are only interested in olfactory situations, we narrow the scope of the annotation and we simplify the frame repository, while adopting the same structure as the original FrameNet based on lexical units and related frame elements.

For example, in FrameNet the ‘smell’ lexical unit is present as a noun or verb in five frames: *Sensation*, *Perception active*, *Give impression*, *Perception experience* and *Chemical sense description*. Their distinctions are rather fine-grained, for example *Perception active* and *Perception experience* both capture perception events but in the former the perceptual experience is intentional, while in the second it can be unintentional. On the other hand, *Chemical sense description* captures mainly descriptions of tastes and smells, while *Give impression* includes characterisations and appraisals of smells and *Sensation* includes nouns that refer to sensations in different modalities. All these distinctions are not needed in our framework, whose final goal is to build a methodology and techniques for olfactory information extraction. Therefore, we only define a single frame, the *Olfactory event*, where all relevant roles of the above frames are captured. We also create domain-specific semantic roles that are the outcome of discussions with experts in olfactory heritage and history. For example, we introduced the roles of *Smell source*, *Evoked odorant* and *Odour carrier*, while we borrowed from FrameNet some generic roles such as *Perceiver*, *Time*, *Location* and *Circumstances*. For

<sup>2</sup><https://framenet.icsi.berkeley.edu>

details, see Section 4.2. Also the list of lexical units (LUs) is defined with the help of domain experts, choosing smell-related lexical units that evoke olfactory situations and events. Our goal is to develop an annotation scheme that may be applied to different languages, therefore the LU lists are being created in seven languages, namely English, Dutch, Italian, French, German, Slovenian, Latin. They include both translations and synonyms of words such as ‘to smell’, ‘odour’ ‘odorous’, ‘smelly’, ‘perfume’ and language-specific terms. For example, German includes a list of compound nouns created with the roots ‘-gestank’, ‘-geruch’, ‘-duft’ (e.g. Regengeruch, Viehgestank, Lavendelduft). We report in Table 1 the lexical units defined so far for English and Italian, while for the other languages the definition is still in progress.

With this annotation, we pursue several objectives: we want our guidelines to be generic enough to accommodate all possible smell-related events which may be mentioned in a text, and also to cover all languages, without the need to provide language-specific adaptations from a semantic point of view. We also want our framework to be flexible enough to deal with meaning change of terms over time, enabling annotators to add new smell-related terms to our initial list during the annotation process. Finally, we define semantic roles using labels that are self-explanatory and not ambiguous (e.g. *Evoked odorant*, *Smell source*), so to facilitate the role selection ensuring a good agreement among annotators.

## 4 Annotation of Olfactory Events

### 4.1 Lexical unit annotation

Following FrameNet guidelines, the annotation of olfactory situations and events should be carried out in two steps: given a sentence, first frame-evoking words or expressions should be marked as lexical units, and then their related semantic roles should be identified, labeled and connected to the LU.

The *Olfactory event* frame is typically evoked by ‘smell’ words, i.e. terms (different for each language) that unambiguously evoke or describe an odor-related situation or event. Using the same terminology of FrameNet, smell words are the lexical units. These include nouns, verbs, adverbs and adjectives. To create an initial set of smell words, we asked domain experts to prepare a list of possible lexical units for each language, which is reported in Table 1 and can be further expanded during an-

notation. The **Other** category covers (ambiguous) smell words, that should be annotated only if they refer to a smell experience or an olfactory situation. We report few example sentences below, underlying the smell-related lexical units:

- (1) The guardsmen drenched their beards in scent.
- (2) Hungry dogs scent a meal in the refuse heaps of the gutters.
- (3) Beside the road the tall asphodel is smelling of cats.

Each olfactory situation can be evoked only by one smell word. If more smell-related terms are present, they have to be annotated as different frame instances. If a smell word can correspond also to a frame element, it has to be labeled with both tags, following the same convention as in FrameNet. Consider for example the following sentence:

- (4) The air in the room was mephitic.

The term ‘mephitic’ would be a quality, but since no other smell-word is present in the sentence, it is clearly this term that evokes the smell event, and it should therefore be marked as lexical unit (this is also part of the seed words listed in Table 1).

If a text clearly describes an olfactory situation but smell words from the pre-defined list cannot be found, other terms can be considered as lexical units if they are used as near-synonyms of ‘smell’, even if out of context they may have another meaning. See for example ‘composition’, ‘essence’, ‘perceive’, etc.

### 4.2 Annotation of Frame Elements (FEs)

Each odor-related event or situation is evoked by a smell-word but can involve one or more participants in the event, each having a specific role. These olfactory frame elements (i.e. semantic roles) have to be identified and annotated as well. The list of participants or semantic roles pertaining to an odour-related situation is reported below.

**Smell source:** the person, object or place that has a specific smell. It can also refer to (non)human / object that produces an odour (e.g. plant, animal, perfume, human). This FE is the entity or phenomenon that the perceiver experiences through his or her senses.

**Odour carrier:** this FE corresponds to the carrier of an odor, which can be either an object (e.g.

<p><b>English</b></p> <p><b>Nouns:</b> stink, scent, scents, smell, smells, odour, odor, odours, odors, stench, reek, aroma, aromas, aromatic, whiff, foetor, fetor, fragrance, musk, rankness, redolence, pong, pungency, niff, deodorant, olfaction</p> <p><b>Verbs:</b> smelling, smelled, reeked, sniff, sniffed, sniffing, whiffed, fragrance, deodorized, deodorizing, snuffing, snuffed</p> <p><b>Adjectives:</b> stinking, stank, stunk, scented, odourless, odoriferous, odorous, malodorous, reeking, aromatic, whiffy, fetid, foetid, fragrant, fragranced, redolent, frowzy, frowsy, pungent, funky, musty, niffy, unscented, scentless, deodorized, noisome, smelly, mephitic, olfactory</p> <p><b>Adverbs:</b> musky, pungently</p> <p><b>Other:</b> atmosphere, essence, putrid</p>
<p><b>Italian</b></p> <p><b>Nouns:</b> lezzo, morbo, putidore, fiatore, puzzo, puzza, fetore, miasma, putrefazione, effluvio, esalazione, estratto, odore, aroma, olezzo, fragranza, profumo, aulimento, odoramento, afrore, tanfo, tanfata, zaffata</p> <p><b>Verbs:</b> odorare, puzzare, profumare, deodorare, odorizzare, aromatizzare, fiutare, annusare, nasare, olezzare, ammorbare, appestare, impestare, impuzzare, impuzzire, impuzzolentire, impuzzolire, intanfare</p> <p><b>Adjectives:</b> puzzolente, fetente, fetido, deodorizzato, putrefatto, odorato, odoroso, odorifero, aromatizzato, profumante, profumato, suave, soave, olfattivo, olfattorio, maleodorante, aromatico, pestilenziale, puzzoso, fragrante</p> <p><b>Adverbs:</b> profumatamente, odorosamente</p> <p><b>Other:</b> essenza, atmosfera, sentire</p>

Table 1: Initial list of possible lexical units for English and Italian

pomander, bottle of perfume, handkerchief) or atmospheric elements like wind and air. Note that the *Odor carrier* has a different role from the *Source*, since the *Source* produces an odour, while the *Carrier* carries a smell that is produced by something else (possibly unknown). A *Carrier* should be annotated only when there is a clear distinction w.r.t. the *Source*. When this distinction cannot be understood or inferred from the text, a *Source* label should be selected. This means that when an odour is described as coming generically from an object or entity and it is not specified or clear from the context whether the object or entity actually produced the odour it should be annotated as *Source*.

**Quality:** This is a quality associated with a smell and used to describe it. For example ‘rancid’, ‘fresh’, etc. This is typically expressed by qualitative adjectives. It is often preceded by an intensifier such as ‘very, really’. The intensifier has to be annotated with the related adjective in the same span. Qualities include intensity (‘not perceptible’, ‘weak’, ‘distinct’, ‘strong’), volume / reach (‘far reaching’), duration (‘lasting’, ‘permanent’), state (‘old’, ‘deteriorated’), character (‘humid’, ‘dry’, ‘garlicky’, ‘fruity’, ‘woody’), hedonic characteristics (‘malodorous’, ‘aromatic’, ‘healthy’).

**Perceiver:** The being that perceives an odour, who has a perceptual experience, not necessarily on purpose. The perceiver is mostly a person or an animate entity. The perceiver can also be expressed by mentioning the perceptive organ (e.g. nose, nostrils, nerves) used in the olfactory experience

**Evoked Odorant:** This frame element describes the object, place or similar that is evoked by the odour, even if it is not visible in the scene. In English, this is often part of a comparison or similarity using the verb ‘to smell’ and introduced by ‘like’. *Evoked Odorants* also include situations, recollections or abstract concepts that are evoked in the *Perceiver’s* mind by smelling an odour.

In some cases, there is no linguistic evidence that a frame element is an *Evoked Odorant* and not a *Smell source*. The former is usually evoked in the mind of the perceiver, often as a recollection of past experiences, while the latter is the actual entity emitting an odour. In these cases, annotators need to interpret the whole sentence and infer if the smell source is real or only evoked, as in the following example where ‘nothing but painted paper and tinsel’ would be *Smell source*:

- (4) The horrid stench [of the leek]<sub>*Evoked\_odorant*</sub> was composed of [nothing but painted paper

Frame Element	Example Sentence
Smell Source	The waiter <u>smelled</u> [something] foul in the kitchen. He carefully <u>smelt</u> [the soup] for any trace of poison.
Odour Carrier	[An Egyptian gale] came rushing upon me, impregnated with <u>pestilential</u> vapours. [The pomander] emitted a <u>smell</u> of musk.
Quality	The coffee had a [pungent] <u>smell</u> . The waiter <u>smelled</u> something [foul] in the kitchen.
Perceiver	The <u>olfactory</u> [nerves of women of quality] are amazingly tender. They have the old <u>smell</u> which [to me] would bring back Knole.
Evoked Odorant	He <u>smells</u> [like flowers]. I have no desire to <u>reek</u> [like the floor of a florist’s stall].
Location	[In Venice] the canals have an offensive <u>smell</u> . The peaceful <u>odour</u> of Mrs. Dillon was prevalent [in the hall of the house].
Time	[In Summer] the city was inundated with a pungent fish <u>smell</u> . [By day] they have little or no <u>smell</u> except in rainy weather, but [in the evening] they are delightfully fragrant.
Circumstances	[The alteration] it made in him would sometimes fill the room with a musty <u>scent</u> . [At high heat] the <u>smell</u> of mud was pungent.
Effect	The <u>odour</u> was so strong [that I fainted].

Table 2: Frame Elements (FEs) related to Olfactory situations and events with corresponding examples. Lexical units are underlined and the FE of interest is put between squared brackets.

and tinsel]<sub>Smell\_source</sub>.

**Location:** This frame element describes the location where the smell event takes place. *Locations* can include both named places (for example names of cities) and common nouns describing locations such as garden, street, kitchen, cliffs, promenade, neighborhoods, etc. Similar to the annotation of *Odour Carrier*, *Locations* are to be marked only when they are different from the *Smell source*. Otherwise, if it is not possible to distinguish whether a place produces a smell or is just impregnated by it, the more generic *Smell source* label is preferred.

Example of annotation of a place as *Smell source*:

- (5) They kneel on the wet flags of this foetid [cave]<sub>Smell\_source</sub>.

**Time:** an expression describing when the smelling event occurred. It includes expressions of duration, frequency and point(s) or period(s) of time.

**Circumstances:** This frame element describes the state of the world under which the smell event takes place. Note that this does not include places and temporal expressions, which should be annotated as *Location* and *Time* respectively. The role

can describe causal implications that lead to or influence the smell event. *Circumstances* may also describe bigger events that include the smell event, for example historical (named) events. Annotators should first try to assign to the FE a *Time* or *Location* role and, only if it does not apply to the specific case, resort to *Circumstances*.

**Effect:** This frame element describes an effect or reaction caused by the smell. This can include entire sentences or clauses describing another event, that is not necessarily a smell event. This can include also the description of emotions triggered in the *Perceiver* by the smell event.

**Creator:** This frame element describes the person that creates an (usually pleasant) smell. This role is frequent in documents in the perfumery domain, while it is almost never found in other types of texts.

Each semantic role may be present or not, or be expressed through multiple instances, if they appear separately in the text. For example, in the following sentence both ‘the excrements of animals’ and ‘their sweat’ are to be labeled as *Smell source*.

- (6) Usually, [the excrements of animals]<sub>Smell\_source</sub>, and in particular [their sweat]<sub>Smell\_source</sub>, are faetid.

For an olfactory event to be present, however, at least a smell word should be annotated.

### 4.3 Annotation Conventions

Following FrameNet annotation practice, we annotate whole constituents that realize frame elements relative to smell words, rather than just tagging the head words of these constituents. That is, we work with a phrase structure grammar, rather than a dependency grammar. A consequence of this is that many frame element labels cover words that have no direct relation of their own to the target, but only to the head of their constituent. For instance, when a frame element is expressed by a noun which takes adjectival, prepositional or clausal complements or which is modified by such elements, these complements and modifiers are included in the frame element tag. This means also that, when a frame element is expressed by a noun phrase, also articles at the beginning of the phrase are labeled. The same holds with prepositions at the beginning of a prepositional phrase (Example 7).

- (7) The peaceful odour of Mrs. Dillon was prevalent [in the hall of the house]*Location*.

In some cases, the same FE label appears multiple times relative to a given target. This can apply to two cases: *i*) multiple separate instances of the same frame element, for example when several Smell sources are mentioned for the same smell event; *ii*) a single instance of a frame element, which is realized in two discontinuous pieces rather than as a single constituent. The cases of discontinuous FEs are particularly frequent in languages that foresee separable terms, for example German split verbs. Both segments have to be annotated and tagged with the same label, and a relation has to be specified going from the peripheral elements to the head of the constituent (see Section 4.4).

Concerning negated events or texts describing the lack of smell events, we annotate them as if they were standard smell events. This is because we are interested in understanding how smells (or a lack thereof) were described at scale and in extracting odor-related terminology, not so much in distinguishing whether a specific event description refers to the presence or absence of a smell. Therefore, the following example would be annotated ignoring the presence of negation:

- (8) The aroma cannot be described [as a floral emanation]*Evoked\_odorant*.

Since we are also interested in metaphorical use of smell-related expressions, we include them in our annotation. This decision is in line with past annotation efforts related to emotions in Dutch texts from 1600 to 1800: no distinction was made between the references to body parts or bodily processes in a literal and a metaphorical sense because such a distinction is often quite difficult to make in early modern texts, where expressions which we now consider to be metaphorical often had a quite material basis in humoral theories of the passions (Leemans et al., 2017).

### 4.4 Deviations from FrameNet annotation

In FrameNet only one relation type is foreseen, that is the relation connecting a lexical unit and the different frame elements. We consider this relation type to hold also for our annotation as the standard relation, which should go from each annotated frame element to the related smell word. However, we introduce two additional relation types. One is used to mark discontinuous frame elements, roles that are expressed by two or more non contiguous strings. While in FrameNet discontinuous FEs are annotated, no specific relation is set to connect them. On the contrary, we introduce this relation to explicitly link strings of text belonging to the same FE. We create a **Discontinuous** relation oriented towards the head or governor of the FE. For example, in the case of a separable verb, the relation will be directed from the prefix to the verb root.

The second relation type is **Anaphoric**, which is not present in FrameNet annotation because it is outside the scope of the project, as stated in the official guidelines (Ruppenhofer et al., 2006). Indeed, FrameNet was originally designed for sentence-level annotation, so no ways to capture relations across sentence boundaries are foreseen. In ODEUROPA, instead, we plan to perform annotation at document level, therefore we introduce anaphoric relations for cases where a frame element is a pronoun and its antecedent is mentioned in discourse (usually preceding the pronoun). In this case, we would annotate the pronoun as bearing a FE label (for example, *Smell source* or *Perceiver*) but we would also manually mark the string explicitly stating whom the pronoun refers to.

This kind of annotation is reported below with *Perceiver1* as the antecedent and *Perceiver2* as the pronoun. In the annotation tool, an arrow has to be set connecting the first to the second element, and

the label *Anaphoric* should be selected.

- (9) There was no persuading [the lady in question]<sub>Perceiver1</sub> that [she]<sub>Perceiver2</sub> had the [most powerful]<sub>Quality</sub> scent.
- (10) He detected [persons of impure life]<sub>Smell\_source1</sub> by [their]<sub>Smell\_source2</sub> smell.

Concerning the annotation tool, we selected IN-CEpTION (Klie et al., 2018), a highly flexible web-based system that supports multiple languages and is extremely user-friendly. The platform foresees three different levels of control i.e. administrator, curator and annotator, so that multiple annotators can perform the task in parallel, the curator can check the annotation quality and compute the agreement, and finally the administrator can close the task, export the data and make further assignments. A screenshot of the annotation window is displayed in Figure 1.

Two expert linguists were trained to annotate English texts containing olfactory situations using the tool. They first annotated the same three historical documents (from 1630, 1833 and 1921) in two training sessions, after which cases of disagreement were discussed and adjudicated. Then, a fourth document was annotated in parallel and used to compute IAA, which was 0.68 Krippendorff’s Alpha on FEs extension and labels.

## 5 Case Study: Olfactory Annotation in Travel Writings

In order to validate the proposed annotation guidelines, we test them on a set of smell-related sentences extracted from 51 books on travel narratives presented in (Sprugnoli et al., 2017) and (Sprugnoli et al., 2018). The books include non-fictional narratives (reports, diaries, letters) and guidebooks about Italy written by English native authors and published between 1860 and 1930<sup>3</sup>. This period was chosen because it followed Italy unification and was characterised by the new phenomenon of leisure-oriented tourists’ travels. Indeed, after unification, new routes to Southern Italy and the islands were opened, so that travelers could reach less visited sites, beyond the classic destinations such as Venice, Florence and Rome.

We first extract from the corpus all sentences containing an English smell word, taken from the

<sup>3</sup><https://sites.google.com/view/travelwritingsonitaly/>

LU list for English. We try to exclude highly ambiguous words such as "essence" or "atmosphere". Overall, we obtain 650 sentences, corresponding to olfactory events or situations. Then, the sentences are annotated following the guidelines described in the previous sections.

### 5.1 Annotation statistics

We report an overview of FE occurrences in our dataset in Table 3. The most frequent FE is Smell Source, which appears in 84% of the cases. If it is not present, Evoked Odorant is usually mentioned. Since we deal with travel writings, Locations are also frequently mentioned, while Time, Effect and Circumstances do not seem to be equally relevant. Usually Time is specified in text or can be inferred by the reader without being part of the Olfactory event. For example, in reports and diaries it often appears at the beginning of the document. Such long-range relations, however, are not captured by our guidelines.

Frame Element	Frequency
Smell Source	546
Quality	272
Evoked Odorant	70
Time	42
Location	159
Odour Carrier	84
Perceiver	79
Effect	26
Circumstances	7
Creator	0

Table 3: Number of annotated FEs over 650 olfactory events found in text

We further extract the list of FE fillers and rank them by frequency. An overview of the output is presented in Fig. 2. We do not include the FEs that have less than 50 occurrences and *Perceiver*, since it is often expressed through the 1st person pronoun ‘I’ or through generic nouns such as ‘the traveler’, ‘the observer’ or ‘the nostrils’.

The list of FE fillers reveals the most striking aspects of Italy to the travelers’ eyes (and nose): the vegetation, typical of the Mediterranean landscape, including ‘violets’ but also ‘rosemary’, ‘oranges and lemons’, ‘primroses’, ‘tall asphodels’, ‘white roses’, ‘lentisk’ and ‘cistus’. These are frequently associated with positive qualities such as ‘delicious’, ‘sweet’ and ‘delicate’, but also ‘strange’

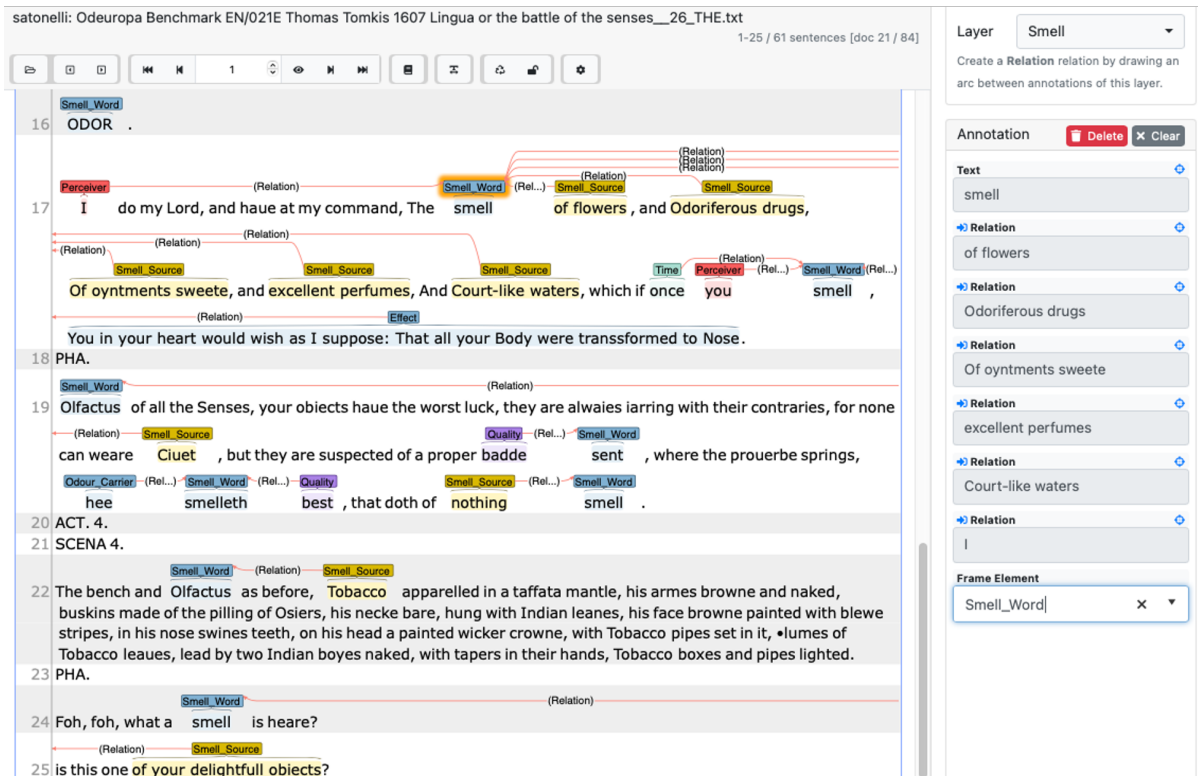


Figure 1: Annotation interface in INCEpTION

and ‘peculiar’. Food is also mentioned in relation with odours, mainly typical ingredients of Italian cuisine such as garlic, spaghettis, frying oil, fish and cheese. Another aspect that is often mentioned in the context of olfactory scenes is religion, not only because travelers mention often the ‘smell of sanctity’, i.e. the belief in the Catholic church that saints were characterised by a sweet smell. They were also struck by the pungent odour of churches, smelling of incense after religious rites, or by the stale odour of catacombs. Finally, travelers are negatively struck by the lack of sanitation of Italian cities, mentioning the ‘intolerable’, ‘pestilential’ and ‘villanous’ odours emanating from ‘filth, mud and garbage’, ‘the vile streets’ and ‘the litter of horses and cattle’. These aspects were noted by travelers coming from Northern Europe where, already in the 18th Century, processes of urban modernisation and sanitation had been initiated (Wrigley, 2012), while Italy, including Rome, was lagging behind the rest of Europe on this issue.

Overall, travelers describe their experiences as full of excitement for being on classic soil and involving all senses. Indeed, smell is often mentioned together with other senses, for example sweet odours are often associated with music, while bad smells characterise noisy places.

## 6 Conclusions

In this paper, we first present annotation guidelines to capture olfactory information in texts, and then describe a first use case aimed at analysing travel writings through the lens of smell descriptions. Although this analysis is carried out on English texts, our guidelines have been designed to be language-independent, and the preparation of LU lists covering different European languages is ongoing. As a next step, we plan to extend the annotation with emotion information, so to capture the effect or reactions triggered by smells. Once a multilingual benchmark with manually annotated olfactory information is available, we will also investigate different approaches to perform the task automatically, exploring cross-language frame information transfer (Tonelli and Pianta, 2008) as well as semi-supervised and few-shot classification.

More broadly, analysing how olfactory experiences are described in texts can offer fresh perspectives for humanities research in this field. Through the connection of different, multilingual sources, especially historical ones, we can promote scholarship in the field of sensory history, and engage audiences with more (nuanced) stories about their olfactory past. Once information extraction sys-



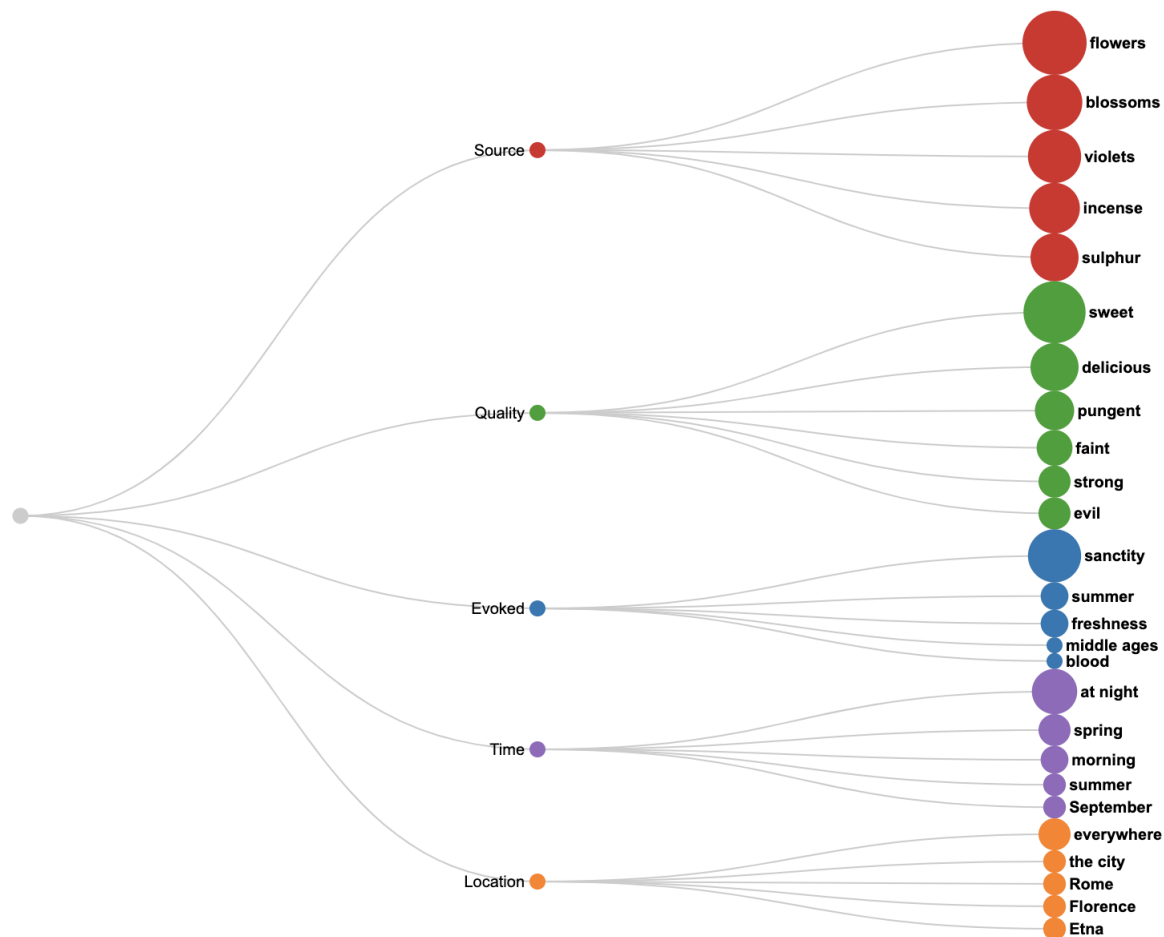


Figure 2: Most frequent FE fillers

tems trained on our annotated data are developed, we will be able to analyse extensive digital archives and trace how scent and smelling were instrumental in shaping communities in the past and how these bound and separated groups and nations.

## Acknowledgements

This research has been supported by the European Union’s Horizon 2020 program project ODEUROPA under grant agreement number 101004469. We thank in particular Inger Leemans, William Tullet, Caro Verbeek and Cecilia Bembibre for their suggestions on how to define and model olfactory events.

## References

Ryan Brate, Paul Groth, and Marieke van Erp. 2020. [Towards olfactory information extraction from text: A case study on detecting smell experiences in novels](#). In *Proceedings of the The 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Litera-*

*ture*, pages 147–155, Online. International Committee on Computational Linguistics.

Constance Classen. 1999. [Other Ways to Wisdom: Learning Through the Senses Across Cultures](#). *International Review of Education*, 45(3-4):269–280.

Holly Dugan. 2011. *The Ephemeral History of Perfume: Scent and Sense in Early Modern England*. Baltimore: Johns Hopkins University Press.

C. Fillmore. 1976. Frame semantics and the nature of language \*. *Annals of the New York Academy of Sciences*, 280.

Charles J. Fillmore and Collin F. Baker. 2001. Frame semantics for text understanding. In *In Proceedings of WordNet and Other Lexical Resources Workshop*.

David Howes. 2006. [Charting the sensorial revolution](#). *The Senses and Society*, 1(1):113–128.

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. [The inception platform: Machine-assisted and knowledge-oriented interactive annotation](#). In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*,

- pages 5–9. Association for Computational Linguistics.
- I. Leemans, J. V. D. Zwaan, I. Maks, E. Kuijpers, and K. Steenbergh. 2017. Mining embodied emotions: A comparative analysis of sentiment and emotion in dutch texts, 1600-1800. *Digital Humanities Quarterly*, 11.
- Els Lefever, Iris Hendrickx, Ilja Croijmans, Antal van den Bosch, and Asifa Majid. 2018. [Discovering the language of wine reviews: A text mining account](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Francesca Strik Lievers and Chu-Ren Huang. 2016. [A lexicon of perception for the identification of synaesthetic metaphors in corpora](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2270–2277, Portorož, Slovenia. European Language Resources Association (ELRA).
- Asifa Majid and Niclas Burenhult. 2014. Odors are expressible in language, as long as you speak the right language. *Cognition*, 130(2):266–270.
- D. Quercia, R. Schifanella, L. Aiello, and K. McLean. 2015. Smelly maps: The digital life of urban smellscapes. In *Proceedings of ICWSM*.
- Daniele Quercia, Luca Maria Aiello, and Rossano Schifanella. 2016. [The emotional and chromatic layers of urban smells](#). *Proceedings of the International AAAI Conference on Web and Social Media*, 10(1).
- Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, C. R. Johnson, and J. Scheffczyk. 2006. *Framenet ii: Extended theory and practice*.
- M. M. Smith. 2006. *How race is made: Slavery, segregation, and the senses*. Chapel Hill : The University of North Carolina Press.
- Mark Smith. 2007. *Sensing the past: seeing, hearing, smelling, tasting, and touching in history*. Berkeley: University of California Press.
- Rachele Sprugnoli, Sara Tonelli, Giovanni Moretti, and Stefano Menini. 2017. A little bit of bella pianura: Detecting code-mixing in historical english travel writing. In *Proceedings of the Fourth Italian Conference on Computational Linguistics (CLiC-it 2017)*, pages 304–309.
- Rachele Sprugnoli, Sara Tonelli, Giovanni Moretti, and Stefano Menini. 2018. Two days we have passed with the ancients...: a digital resource of historical travel writings on italy. In *In the Book of Abstracts of AIUCD 2018*.
- Serra Sinem Tekiroğlu, Gözde Özbal, and Carlo Strapparava. 2014a. [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. Association for Computational Linguistics and Dublin City University.
- Serra Sinem Tekiroğlu, Gözde Özbal, and Carlo Strapparava. 2014b. [Sensicon: An automatically constructed sensorial lexicon](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1511–1521, Doha, Qatar. Association for Computational Linguistics.
- Sara Tonelli and Emanuele Pianta. 2008. [Frame information transfer from English to Italian](#). In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).
- William Tullett. 2016. [Grease and sweat: Race and smell in Eighteenth-Century English culture](#). *Cultural and Social History*.
- Bodo Winter. 2019. *Sensory linguistics: Language, perception and metaphor*, volume 20. John Benjamins Publishing Company.
- Richard Wrigley. 2012. Making sense of rome. *Journal of Eighteenth-Century Studies*, 35(4).