Poppy Siahaan*

**Indonesian basic olfactory terms: more negative types but more positive tokens**

https://doi.org/10.1515/cog-2021-0092
Received August 6, 2021; accepted July 14, 2022; published online August 17, 2022

**Abstract:** The present study investigates the semantics of a dozen basic smell terms in Indonesian using data from a large corpus of written register. Examining how these smell terms lexicalize some odors but not others raises questions that are central to our understanding of the language of olfaction. How are smell terms structured? What does the structure of smell terms tell us about human behavior? By applying cluster analysis, the present study reveals that the Indonesian odor lexicon is structured based on one dimension correlating with pleasantness. The large dataset of a written corpus enables the present study to reveal the differences in lexicalization and frequency: Indonesian smell terms have more negative types but more positive tokens in texts. This novel approach to investigating smell terms allows us to take a step closer toward our goal of understanding olfactory vocabulary, as data on token frequency are difficult to obtain in studies of (unwritten) minority languages. This key finding supports the Pollyanna Hypothesis: people tend to use positive words more often than negative words, but the negative words convey more information.

**Keywords:** basic smell terms; codability; Indonesian; olfaction; Pollyanna Hypothesis

1 Introduction

Contrary to widely held beliefs (see references in Majid and Burenhult 2014; Wnuk and Majid 2014), human languages can encode differentiations among odors. Recent studies of several languages across the world have challenged the view that the olfactory experience is “ineffable” (cf. Levinson and Majid 2014), “universally encoded as objects” (Olofsson and Gottfried 2015a: 629), and viewed as difficult or impossible to express through language (Yeshurun and Sobel 2010) due to the lack of interaction between the odor processing and the language systems in the human

*Corresponding author: Poppy Siahaan, University of Cologne, Institute of Languages and Cultures of the Islamicate World, Albertus-Magnus-Platz, 50923 Cologne, Germany, E-mail: poppy.siahaan@uni-koeln.de
brain (Olofsson and Gottfried 2015b: 314). As observed by Wnuk et al. (2020), smell terms are often considered as a linguistic rarum, “a sizable inventory of basic smell terms, i.e., one with more than two or three items” (Plank 2001. https://typo.uni-konstanz.de/rara/ (accessed 14 February 2022)), which is characteristic of languages with a small number of speakers (Wohlgemuth 2010).

However, studies of languages in diverse communities worldwide have provided evidence that a sizable lexicon of dedicated abstract smell terms (as opposed to concrete object-based descriptions) are found not only in the less-described languages of small communities, such as Buli on Halmahera (Bubandt 1998), Semai (Tufvesson 2011), Jahai (Burenhult and Majid 2011), Maniq (Wnuk and Majid 2014) on the Malay Peninsula, Formosan languages (Lee 2010, 2015), Boholano in the Philippines (Beer 2014), Totonac (Aschmann 1946) and Seri (O’Meara and Majid 2020) in Mexico, or Cha’palaa in Ecuador (Floyd et al. 2018) and Luwo in South Sudan (Storch 2013). Dedicated abstract smell terms are also attested in major languages with millions of speakers in industrialized societies; for example, Javanese (Septiawati 2010), Cantonese in Hong Kong and Macau (de Sousa 2011), and Thai (Wnuk et al. 2020). A recent study found that many languages across the world have elaborate smell words (Majid 2021a).

Studies of smell terms of the smell lexicon have used different methods to elicit smell vocabulary and to understand their meaning and structure, such as linguistic elicitation (Wnuk and Majid 2014), elicitation by odor stimuli (Lee 2015; Majid and Burenhult 2014), exemplar listing (O’Meara and Majid 2020; Wnuk et al. 2020), similarity judgment (Wnuk and Majid 2014), rating tasks (Wnuk and Majid 2014), using dictionaries (Wnuk et al. 2020), and interviews (de Sousa 2011). Furthermore, corpus data have proven to be useful in studies of smell terms of the smell lexicon to estimate the salience of smell terms (de Sousa 2011), to gain additional information on their meaning and use (Wnuk et al. 2020), or to compare the frequencies of smell verbs and smell terms in three different languages (Floyd et al. 2018). However, thus far, no study has adopted a corpus linguistic approach based on a large written corpus to investigate the semantics of smell terms and to examine the quantitative distribution and internal structure. One of the challenges may be due to the lack of large corpora. Large corpora already exist for major languages such as English, German, or French, but data concerning token frequency for unwritten minority languages with elaborate smell terms are difficult to attain. For example, a study using a spoken corpus (approximately 500,000 words) of an endangered language, Cha’palaa, revealed that seven out of the 15 smell terms that were identified via elicitation were not found in the corpus (Floyd et al. 2018). The present paper demonstrates a novel corpus linguistic approach to smell terms, and makes use of an existing large corpus of Indonesian with 1,206,281,985 word tokens (Goldhahn et al. 2012). It applies the method of hierarchical agglomerative cluster analysis to visualize a tree diagram that is easy to
understand to reveal the structure of the olfactory domain (Gries 2013: 336–349; Divjak and Fieller 2014; Levshina 2015: 306–317).

The present study also attempts to answer one of the “Outstanding Questions” raised in a recent study (Majid 2021a: 10), namely whether smell terms are used metaphorically, as this is not the case in Jahai (Burenhult and Majid 2011) although it is the case in Seri (O’Meara and Majid 2020) and Thai (Wnuk et al. 2020). The data from a large corpus enables this paper to reveal the rich figurative use of smell terms in Indonesian. This finding contradicts the claim that there are few olfactory metaphors across languages, particularly when compared to metaphors from the visual domain (Ibarretxe-Antuñano 1999; Sweetser 1990). The paper reveals a novel metaphor **GOOD IS FRAGRANT**, which contradicts the assumption that olfaction has cross-linguistic negative connotations (cf. Ibarretxe-Antuñano 2021).

The structure of the paper is as follows. Section 2 provides the theoretical background. The data and method are presented in Section 3, while Indonesian basic smell terms are described in Section 4. Section 5 presents the results of the cluster analysis that was conducted to investigate the structure of the basic smell terms. Section 6 contains a discussion in which Indonesian smell terms are compared to smell terms from other languages attested in literature, and the Pollyanna Hypothesis of smell terms across languages is proposed. Section 7 concludes with the study’s findings.

## 2 Theoretical background

### 2.1 Smell language and perception literature

Indonesian speakers use dozens of basic olfactory terms in their written register. The notion of “basic smell terms” is derived from “basic color terms” (Berlin and Kay 1969). To the best of my knowledge, Classen was the first scholar to use the notion of “basic olfactory terms such as ‘stinking’ and ‘aromatic’” (1993: 54), but Burenhult & Majid were the first to define “abstract olfactory terms” as “mono-lexemic, psychologically salient, not restricted to a narrow class of objects, nor are they source descriptors” in their paper on basic smell terms in Jahai (2011: 24–25).

In her subsequent study, Majid elaborated on and applied the definition of basic smell terms in English, and explained that the words **petrichor ‘the smell of rain on dry earth’**, and **odorous** and **odoriferous** were not basic smell terms because the former was restricted to a narrow class of objects and was not commonly known, whereas the latter only indicated the presence of smell (Majid 2021a). The words **fruity** and **chocolatey** are also not basic smell terms because they each describe the source of an odor (Majid 2021a). Furthermore, words such as **nice** or **disgusting**
provided as responses in elicitation tasks are evaluative, and are therefore not considered to be abstract smell terms (Majid and Burenhult 2014). Majid (2021a) suggested stinky, fragrant, and musty as candidates for basic smell terms in English. Hence, as opposed to languages in the studies mentioned previously, English has a limited basic vocabulary for encoding odor qualities.

The descriptions of psychological salience vary depending on the number of speakers and whether the language has a written register. With regard to smaller communities, psychological salience is described as “common parlance, known and used by all” (Burenhult and Majid 2011: 24), “appearing in everyday conversation and child speech” (Majid et al. 2018a: 2), “produced earlier” in elicitation (O’Meara and Majid 2020: 370), and “known by everyone in a speech community” (Majid 2021a: 22). In studies of languages with millions of speakers and written registers, indices of psychological salience include, among others, “occurrence on the internet” (de Sousa 2011: 41) or in the “corpus” (Wnuk et al. 2020: 940). The 12 smell terms in this study may be considered psychologically salient because they are known and used by Indonesian speakers in online communication and have dictionary entries (Alwi et al. 2005; Stevens and Schmidgall-Tellings 2008).

Furthermore, there is an on-going discussion regarding ineffability across perceptual domains; that is, “the difficulty or impossibility of putting certain experiences into words” (Levinson and Majid 2014: 408). In other words, the notion of ineffability concerns whether our visual or olfactory experiences, for example, may be expressed in language. One can distinguish different kinds of ineffability. Weak ineffability refers to a case in which a particular sensation is not codable in a certain language, although it may be codable in another language, whereas strong ineffability refers to a case in which no language at all codes a certain sensation (Levinson and Majid 2014). For example, in one language, the sense of vision appears to be ineffable. Yéli Dnye ‘Yéli language’, a language isolate spoken on Rossel Island by approximately 3,500 speakers, does not have a superordinate term for ‘color’; nor does it have basic color terms, although there are individual strategies for referring to hues (Levinson 2000). In another language, the sense of smell appears to be ineffable. Fon, a Kwa language spoken in Benin, does not have a verb indicating ‘to smell’; nor does it have any abstract smell terms, although strategies for talking about odors exist (Lambert-Brétière 2021). Thus, in the relationship between human language and visual/olfactory

1 In the color domain, Berlin and Kay (1969: 6) first precisely defined psychological salience as “(1) a tendency to occur at the beginning of elicited lists of color terms, (2) stability of reference across informants and occasions of use, and (3) occurrence in the idiolects of all informants”, but later amended this to the more general “relatively salient as evidenced in frequent and general use (unlike puce and mauve)”, Kay and McDaniel (1978: 612).
experiences, we can speak about weak ineffability. There are three notions of
codability; that is, how to measure the efficiency with which sensory experience
(for example) is coded in language: linguistic codability, psycholinguistic cod-
ability, and communication accuracy (Levinson and Majid 2014: 412). The fact that
the Indonesian language has a sizable inventory of smell terms means that
olfaction has a degree of linguistic codability in this language.

2.2 Pollyanna Hypothesis

The Pollyanna Hypothesis, first posited by Boucher and Osgood (1969), and, later
proposed to be upgraded as the Pollyanna Principle by Matlin, because it was “so
pervasive and strong” (2004: 255), refers to the human tendency toward a positivity
bias in communicating. Based on small-scale, cross-cultural studies, Boucher and
Osgood (1969) provided evidence that evaluatively positive words were used more
frequently and diversely, and were acquired earlier by children than were evalu-
atively negative words; evaluatively positive words take the negative affix (for
example, happy-unhappy) more frequently than do evaluatively negative words
(such as broken-unbroken). The Pollyanna Hypothesis has been confirmed in a
wide range of studies with different data sources, a variety of methodologies, or
subjects with different ages or backgrounds. For example, in business commu-
ication between corporations and their stockholders, positive words predominate
regardless of the corporations’ financial situation (Hildebrandt and Snyder 1981).
In a psychological study based on self-rating assessments, students “process
pleasant information more accurately and efficiently than less pleasant informa-
tion” (Matlin and Gawron 1979). Matlin (2016) outlined studies corroborating the
Pollyanna Hypothesis in perception, language, and memory. A study using online
written communication in three languages (English, German, and Spanish)
revealed that the frequency of words depended not only on the information con-
tent, but also on the emotional content, and that negative words conveyed more
information than did positive words because “the informativeness of a word
increases uniformly with its valence decrease” (Garcia et al. 2012: 1). Dodds et al.
(2015) studied 100,000 words in 24 corpora in 10 languages (English, Spanish,
French, German, Brazilian Portuguese, Korean, Simplified Chinese, Russian,
Indonesian, and Arabic). Native speakers of each language rated the words from

---

2 The Pollyanna Hypothesis is based on the title of the best-selling children’s book, Pollyanna,
which was written in 1913 by the American author Eleanor Porter. Pollyanna is also the name of the
child heroine of the book, who has an optimistic and positive attitude and sees the bright side of
life beyond measure, despite her misfortunes.
negative to positive on a nine-point scale. The computationally processed data thus represent five million individual human scores and confirm the Pollyanna Hypothesis: There is a universal positivity bias in human languages. The tendency toward the positivity bias is also attested in English and German books for children and young adults (Jacobs et al. 2020).

Examining how smell terms are used in communication to refer to pleasant and unpleasant odors, and how pervasive and strong the Pollyanna Hypothesis is, raises questions that are central to our understanding of the language of olfaction. What is the structure of olfactory domain as a whole? How do Indonesian smell terms compare to other smell terms attested in other languages? Do Indonesian speakers also have a tendency toward the positivity bias and use pleasant smell terms more frequently than unpleasant ones? Do Indonesian corpus data confirm or contradict the Pollyanna Hypothesis? This paper aims to address the research questions by applying cluster analysis, which is a method that is used in corpus linguistics. In doing so, the paper aims to contribute to the literature on smell in language and perception, as well as to the Pollyanna Hypothesis.

3 Data and method

A list of a dozen monolexemic abstract words lexicalizing smell was generated based on the author’s native-speaker knowledge (see Table 1 in Section 4). They satisfy the criteria for basic smell terms (Burenhult and Majid 2011; Majid 2021a), as they are abstract olfactory terms used to identify the properties of smells (Levinson and Majid 2014). Subsequently, Javanese smell terms (Septiawati 2010) have also been taken into consideration, as Javanese is one of the main donors of Indonesian loanwords (Tadmor 2009b). A bilingual dictionary (Stevens and Schmidgall-Tellings 2008), a monolingual dictionary (Alwi et al. 2005), and a thesaurus (Endarmoko 2007) were consulted, resulting in the addition of eight more words referring to smell properties.3 The Leipzig Corpora Collection (Goldhahn et al. 2012) was used to obtain information about the frequency of use of each word. Abstract

3 The following six smell terms were found in the dictionaries and the thesaurus: aring/haring ‘stinking, smelly; smelling of urine’, raksi ‘scent, fragrant’, gangsri ‘scent’, ganda ‘odor, fragrance, aroma’, pedar ‘rancid’, and perat ‘rancid’; see Alwi et al. (2005); Stevens and Schmidgall-Tellings (2008); Endarmoko (2007). Four are polysemous: aring/haring, ganda, pedar, perat. Nevertheless, as their frequency of occurrence in language use amounted to less than five tokens, they are not studied further in the present paper. Two further smell terms registered in the authoritative monolingual dictionary of Indonesian, see Alwi et al. (2005), hancing ‘putrescent smell like urine stench’ and hangit ‘smell like the smell of crust’ (the author’s translation) were identified as appearing only on Malay web pages, and were therefore not included in this study.
smell words with less than five tokens in the corpus have not been considered, as they do not appear to satisfy the criterion of psychological salience.

A database of the remaining 12 Indonesian basic smell terms was then created (see Table 1). A random sample of 1,000 tokens was created for each of the first three terms on the list with a frequency greater than 10,000 tokens; for the remaining nine smell terms with less than 10,000 tokens each, all the examples were included in the investigation.

Before investigating the smell terms, it is important to clarify the notion of “odor quality”. Basic smell terms encode abstract odor qualities that are independent of the object from which the smell emanates. As the philosopher Cavedon-Taylor (2018: 81) argued, “olfactory experience is about odors, not objects”. Visuocentric thinking about olfaction would lead one to falsely believe that olfactory experience was about objects, which is not true, as justified via veridical olfactory perception:

First, there is an event: molecules evaporate from an object. Second, those molecules form a cloud. Third, molecules from that cloud travel up the perceiver’s nose. Fourth, the molecules contact the olfactory bulb. Last, olfactory experience arises after subpersonal processing of the resulting stimulation (Cavedon-Taylor 2018: 83).

Odors can transfer from one object to another. For example, after cutting a fish, we can say my hands smell of fish in English, despite having washed our hands and the fish having been consumed, because “olfactory experiences of transferred odor are veridical” (Cavedon-Taylor 2018: 87). If the odor were an object, such as the fish, the olfactory experience (that our hands smell of fish) must be non-veridical, which is not right. When talking about their olfactory experiences, Indonesian speakers would say tangan saya bau amis (lit. ‘hands my smell amis’), in which amis is the abstract smell term denoting the odor quality emanating from an object, such as (but not exclusively) fish. The hands are not the object of the odor, but the object to which the odor has transferred. Hence, the word hands is the lexical item referred to by the dedicated smell term amis.

A concordance was created with each smell term as the keyword in context. All the citations in the concordance were then inspected manually to determine the lexical items referred to by the smell terms based on their syntactic relationship. For example, in a citation, the smell term amis ‘fishy smell’ is a predicate of the lexical item ikan ‘fish’, as in Ikan memiliki bau amis yang khas ‘Fishes have typical fishy smell’ (lit. ‘fish have odor fishy.smell which typical’). In another citation, the smell term anyir ‘bloody smell’ is an adjective appearing as a modifier of the lexical item darah ‘blood’ within the nominal phrase anyir darah ‘the smell of blood’ (lit. ‘bloody.smell blood’). In another citation, the lexical item daging kambing (lit. ‘meat goat’) was the object of a preposition in a prepositional phrase that modified
the smell term *prengus* ‘goaty smell’; *bau prengus pada daging kambing* ‘the smell of goat on mutton’ (lit. ‘odor goaty.smell on meat goat’). In cases with multiple lexical items, only the first lexical item mentioned was counted. For example, the lexical item of *amis* ‘fishy smell’ was *ayam* ‘chicken’ in the example *Kencur memberi aroma khas dan mengurangi aroma amis dari ayam, daging, atau ikan* ‘Rhizome (*Kaempferia galanga*) gives a distinctive aroma and reduces the fishy smell of chicken, meat, or fish’. Citations with typos were manually removed. Citations were also removed manually if the lexical item was not specified explicitly; for example, *Masak hingga mendidih dan harum* ‘Cook (it) until boiling and fragrant’. Figurative uses of smell terms were also counted to demonstrate the metaphorical potential of olfaction (Ibarretxe-Antuñano 1999; Kövecses 2019; Sweetser 1990).

Following the procedure described in Wnuk et al. (2020), synonyms or highly similar lexical items were merged; for example, *lokia* ‘leucorrhoea’, *keputihan* ‘whitish’, *lendir kemaluan* ‘genital slime’ (lit. ‘slime abashment’); *ikan mati* ‘dead fish’ (lit. ‘fish dead’), *ikan yang mati* ‘fish which is dead’ (lit. ‘fish which dead’), and *bangkai ikan* ‘fish carcass’ (lit. ‘carcass fish’); *baju* ‘clothes’ and *pakaian* ‘clothes’. However, lexical items in different subcategories were not merged; for example, *daging* ‘meat’ was not merged with *daging sapi* ‘beef’ (lit. ‘meat cow’), and *ikan bandeng* ‘*Chanos chanos*’ was not merged with *ikan gabus* ‘*Channa striata*’. A complete list of lexical items for every smell term is provided in the Supplementary material. Numbers in brackets refer to the token’s frequency in the corpus.

### 4 Basic olfactory terms: semantic analysis and their frequency in the corpus

Indonesian is an Austronesian language that is used as the sole national language of Indonesia by over 250 million speakers (around 40 million first-language speakers and approximately 200 million second-language speakers), in addition to their ancestral home languages (Evans 2009: 113; Tadmor 2009b, 2009c), comprising 704 indigenous languages. Developed from an indigenous Malay language, the national language was chosen judiciously among Dutch and Javanese for the new nation that was created from the Dutch East Indies, encompassing present-day Indonesia. The Malay language was given a new name, *Bahasa Indonesia*, ‘language Indonesian’ during the Second Congress of Indonesian Youth that was held in 1928 (Paauw 2009). The regional languages mainly influenced Indonesian on the lexical level. Words have also been borrowed, with Javanese,
Sundanese, and Balinese being the main donors. However, the frequent use of even borrowed smell terms in Indonesian corpus strengthens their integration into Indonesian. In his book about endangered minority languages, Evans (2009: 1) stated that Indonesian was one of a few world languages, apart from English, Chinese, Spanish, Hindi, Arabic, Portuguese, French, Russian, and Swahili, that spread with an “accelerating tempo”.

The word order in Indonesian is SVO. Following the perception verbs’ classification (Viberg 1983), the Indonesian olfactory verb *cium* ‘to smell, to sniff’ conflates activity controlled by a human agent and experience referring to a state that is not controlled. *Cium* also means ‘to (sniff) kiss’ the Malay tradition of greeting or showing affection, which is also attested in South and Southeast Asia (Hopkins 1907; Schapper 2019; Wnuk et al. 2020). Indonesian does not have a phenomenon-based verb of olfaction, such as *It smells good*. To express the generic meaning of smell, Indonesian speakers use the word *bau* ‘smell’ or ‘odor’, see Example 1. Like the English word ‘smell’, the Indonesian word *bau* also has a bias toward an unpleasant smell (cf. Krifka 2010). The bare root *bau* can be categorized as a noun, a verb, or an adjective with the meanings of ‘smell’, ‘to emit a smell’, or ‘(to be) smelly’, respectively.5

(1) Mulutnya sudah **bau** alcoh. mouth=3S already **smell** alcohol
   ‘Her/his mouth already smelled of alcohol.’

Indonesian speakers use dedicated smell terms to talk about the specific qualities of smells. For practical reasons, this paper only considers the bare roots. The grammatical category of bare roots is unclear and can be flexible in Indonesian (Ewing 2005); thus, the word order plays an important role. The three generic smell terms *aroma*, *wangi*, and *harum* can be categorized as nouns; for example, as the head of a noun phrase *wangi bunga* ‘the fragrance of a flower’ (lit. ‘fragrance flower’). All smell terms except *aroma* can be categorized as adjectives; for example, *bunga wangi* ‘fragrant flower’ (lit. ‘flower fragrance’).

The paper examines 12 monolexemic smell terms in Indonesian: *aroma* ‘aroma’, *wangi* ‘fragrant’, *harum* ‘fragrant’, *amis* ‘fishy smell’, *apek* ‘musty’, *tengik* ‘rancid smell’, *anyir* ‘bloody smell’, *sangit* ‘burnt smell’, *pesing* ‘urine stench’, *langu* ‘foul smell’, *perengus* ‘goaty smell’, and *bacin* ‘smell of saliva’. Note that, for practical reasons, the glosses in this paper have been simplified; some are based on their prototypical objects. Translations from dictionaries are presented in Table 1. Note also that *amis* does not mean ‘fish’ but ‘fishy smell’; nor does it derive from *ikan* ‘fish’; *anyir* does not mean ‘blood’ but ‘bloody smell’, and is not derived from

---

5 I Wayan Arka, pc.
"darah" ‘blood’. Moreover, "pesing" does not mean ‘urine’ but ‘urine stench’, and is not derived from "kencing" ‘urine’ and so forth. The meaning of each smell term refers to the olfactory quality, not to the object. They are abstract terms in the sense of not being restricted to a narrow class of objects. That is, "amis" is not restricted to the smell of fish, "pesing" is not restricted to the smell of urine, "prengus" is not restricted to the smell of goat, and "bacín" is not restricted to the smell of saliva. In fact, each smell term can refer to the olfactory quality of numerous objects (see the Supplementary material). These smell terms indicate that olfaction is codable in Indonesian, as there are words that only code the “descriptive, abstract property” of odor; that is, words that “identify the precise properties of smells” and not other

### Table 1: Indonesian smell terms.

<table>
<thead>
<tr>
<th>Smell terms</th>
<th>Dictionary definition</th>
<th>Prototypical lexical items</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>aroma</td>
<td>fragrance</td>
<td>perfume, body, coffee, food, tea</td>
<td>27,478</td>
</tr>
<tr>
<td>wangi</td>
<td>fragrant, scent, perfume, scented</td>
<td>perfume, <em>Chrysopogon zizanioides</em> grass, flower, body, car interior</td>
<td>19,055</td>
</tr>
<tr>
<td>harum</td>
<td>fragrance, odor, smell, fragrant, aromatic</td>
<td>flower, spice, body, perfume, fruit</td>
<td>13,773</td>
</tr>
<tr>
<td>amis, hamis</td>
<td>nasty smell, stinking, reeking, fetid, fishy smell</td>
<td>fish, blood, egg, duck meat, meat</td>
<td>1,968</td>
</tr>
<tr>
<td>apak, apek, hapak</td>
<td>having a sweaty smell, musty,usty, stuffy, frowsy (of food)</td>
<td>clothes, rice grain, room, subsidized rice grain for poor people, car oil, food, fat, bran, margarine</td>
<td>935</td>
</tr>
<tr>
<td>tengik</td>
<td>rancid, stinking</td>
<td>blood, fish, egg, vaginal discharge, duck meat</td>
<td>577</td>
</tr>
<tr>
<td>anyir, hanyir</td>
<td>rancid, tasting/smelling of train oil, stinking (of uncooked fish/meat)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sangit, angit</td>
<td>(it has a) burnt (smell, smells burnt scorched)</td>
<td>rice bug, smoke, car, brake, tempeh</td>
<td>461</td>
</tr>
<tr>
<td>pesing</td>
<td>stinking/smelling (e.g., the smell of urine)</td>
<td>toilet, urine, room, bathroom, pork</td>
<td>419</td>
</tr>
<tr>
<td>langu</td>
<td>unpleasant/rotten odor</td>
<td>soybeans, soy milk, vegetable, carrot, enzyme lipoxygenase</td>
<td>215</td>
</tr>
<tr>
<td>prengus, perengus</td>
<td>the smell like the smell of a male goat/sheep</td>
<td>mutton, goat milk, meat, goat, lamb meat</td>
<td>134</td>
</tr>
<tr>
<td>bacín</td>
<td>stinking (as the smell of saliva/tainted fish)</td>
<td>mouth, city, salty fish, drainage ditch, sweat</td>
<td>43</td>
</tr>
</tbody>
</table>

6 If not stated otherwise, all the translations in this table have been taken from Stevens and Schmidgall-Tellings (2008).
7 For comprehensive data about the lexical items, see the Supplementary material.
8 Prengus or perengus is not listed in Stevens and Schmidgall-Tellings (2008).
9 Alwi et al. (2005); the author’s translation.
properties such as “affect” or objects of smells (Levinson and Majid 2014: 411). Their frequent use in the corpus confirms their psychological salience. These smell terms qualify as basic terms in the sense of Berlin and Kay (1969) (cf. Burenhult and Majid 2011).

In Table 1, we can see that each of the first three smell terms, aroma, wangi, harum, referring to a pleasant smell has a frequency greater than 10,000 tokens, while there are nine types of unpleasant smell terms: amis, apek, tengik, anyir, sangit, pesing, langu, prengus, and bacin. Each unpleasant smell term accounts for only about 10% tokens or less compared to each pleasant smell term. The pleasant smell terms are shown in gray, whereas the unpleasant smell terms are indicated in dark gray. Hence, this novel approach to smell terms, which is based on a large written corpus, helps us to understand their structure and to reveal a difference in lexicalization (more negative types) and frequency (more positive tokens). Figure 1 displays the token frequencies of smell terms as a bar plot. The gray bars show the frequency of pleasant smell terms with more than 10,000 tokens, whereas the dark gray bars show frequency of unpleasant smell terms with less than 1,000 tokens except for amis.

Figure 1: Token frequencies of Indonesian basic smell terms.
4.1 Pleasant smell terms

4.1.1 Aroma

Despite its status as a loanword from Dutch or English (Stevens and Schmidgall-Tellings 2008), aroma ‘fragrance’ is widely used as a smell term in Indonesian and has a dictionary entry. Overall, aroma has the relatively general meaning of pleasant smells. The most common lexical items of aroma are perfume and the human body, followed by a range of edible items and beverages such as coffee, food, tea, and fruit. However, aroma can also be used to describe unpleasant smells. Thus, the expression aroma tubuh, literally ‘aroma body’, can refer to the good, bad, or neutral odor of a human body depending on the context. Nevertheless, the reference to the good smell of a human body in aroma tubuh ‘aroma body’ predominates. To refer to ‘bad body odor’, Indonesian speakers use the word bau, which has a bias toward an unpleasant smell as in bau badan (lit. ‘smell body’).

(2) Betul, dari arah meja tercium aroma daging ayam.
Right from direction table be.smelled fragrance meat chicken ‘That’s right, (you could) smell the fragrance of chicken (coming) from the table.’

Similar to what has been reported in other languages that have elaborate olfactory lexicon (O’Meara and Majid 2020; Wnuk et al. 2020), Indonesian smell terms are also used figuratively. In other words, the use of smell terms is motivated by “conventional extensions of word meanings”10 or “a metaphoric or metonymic relationship” (Dancygier and Sweetser 2014: 4), and not by a literal meaning.11 In a random sample of 1,000, 131 citations of aroma were used figuratively. For example, the literal meaning of an utterance such as mencium aroma korupsi is ‘to smell the aroma of corruption’, which is, of course, not to be taken at face value. In its metaphorical use, aroma refers more frequently to the unpleasant domains of corruption, politics, and competition. The linguistic expression mencium aroma korupsi exemplifies the metaphor BAD IS SMELLY (Ibarretxe-Antuñano 1999; Kövecses 2019; Sweetser 1990).

---

10 Before the paradigm shift created by Lakoff and Johnson (1980), a metaphor was traditionally regarded as a rhetorical device that was not in conventional or everyday usage. Lakoff and Johnson (1980) viewed metaphor as a conceptual mapping from a relatively more concrete source domain onto a relatively more abstract target domain, whereas metonymy refers to the use of an entity to refer to another that is related to it.

11 For the procedure of metaphor identification, see Steen et al. (2010).
4.1.2 Wangi

According to the World Loanword Database (WOLD), the status of *wangi* is “clearly borrowed” from Javanese, Sundanese, or Balinese (Tadmor 2009a). Nevertheless, its integration into Indonesian is marked by its frequent occurrence in the corpus. *Wangi* refers to pleasant smells in general, and its meaning is similar to that of *harum*. However, their contexts of use are quite different. Two common idioms adopt the smell term *wangi* to lexicalize perfume and the plant name, namely *minyak wangi* ‘oil fragrant’ (139 citations) referring to perfume in general, (see Example 3), and *akar wangi* ‘root fragrant’ (43 citations) referring to the root of the perennial bunchgrass (*Chrysopogon zizanioides*). This distinct pattern of idiomatic expressions involving smell terms to name natural objects is also attested in Seri (O’Meara and Majid 2020). As idioms, they “act like single words” and are not separable (Sneddon et al. 2012: 27). Thus, separated by the relativizer *yang*, the linguistic expression *minyak yang wangi* ‘oil which is fragrant’ does not refer unambiguously to perfume, but has the general meaning of oil with a pleasant smell and could refer to any kind of oil, and *akar yang wangi* ‘root which is fragrant’ does not refer to *Chrysopogon zizanioides*, the extracted oil of which is used for cosmetics or aromatherapy, but may refer to any fragrant roots. *Wangi* also refers to the smells of flowers, bodies, car interiors, fruit, rice grains, or pandanus leaves.

(3) **Minyak wangi yang dipakai memang mengasyikkan.**

‘The perfume (lit. oil *fragrant*) (which was) used was really fun.’

In contrast to Seri, which lacks a smell verb dedicated to pleasant odors (O’Meara and Majid 2020), hence also lacks a metaphorical mapping of pleasant or positive properties onto the target domain, *wangi* is used beyond its literal meaning. In a random sample of 1,000, only 27 citations of *wangi* were used metaphorically; for example, *wangi surga* depicting the ‘fragrance of paradise’, which is the counterpart of the more common phrase *bau surga* ‘the scent of paradise’. Those who can smell *wangi surga* ‘fragrance of paradise’ or *bau surga* ‘smell of paradise’ are believed to be people who will go to heaven because they are pious. The concept of the fragrant scent from the garden of paradise is probably derived from Islam, which arrived in Indonesia in the twelfth century.12 Furthermore, the expression

12 “The association of good fragrance with paradise has a long history[...] In early Christian literature the idea of paradise as a garden is abundantly attested [...] In Islamic literature the image is greatly expanded and elaborated by numerous references to specific aromatics in place of the more generic good scents and incense of Syriac literature [...] Aromatics, so long associated with purity and immortality”, King (2017: 354–356).
sastra wangi “fragrant literature” has become a fixed label; in the late 1990s, it was attributed to a group of female Indonesian writers who wrote about the controversial issue of female sexuality (Lee 2007; Marching 2007), thus providing a slight hint about a negative connotation of the pleasant smell term wangi.

4.1.3 Harum

Harum is a pleasant smell term that is mainly associated with the fragrance of vegetation, particularly from flowers such as roses, and fruit. The smell term harum is also commonly used in cooking recipes as a reference system for the standard preparation methods of spice, onion, or garlic, as shown in the example below. In contrast to the Indonesian reference system that is based on the olfactory sense, (cf. also Example 2 in Thai “stir fry with rice until it’s fragrant” in Wnuk et al. 2020: 940), the visual sense is used as a reference system in Western culture, as shown in English recipes; for example, to fry onions until they are translucent, golden, brown, yellow, or opaque. Harum also refers to the odors of the human body, including the mouth, and the fragrance of perfume.

(4) Tumis bawang putih, bawang bombay sampai harum.

‘Stir-fry garlic and onion until fragrant.’

Both wangi and harum refer to pleasant smells, but their distinct usage, particularly figuratively, suggests their different meanings. Unlike its counterpart wangi, the figurative use of harum is more frequent, with 198 citations in a random sample of 1,000 (almost 20%), none of which suggests a negative connotation. In 138 citations, harum is linked to the lexical item nama ‘name’ in 60 citations, it is linked to other lexical items from different domains, such as reputasi ‘reputation’, pekerjaan ‘job’, cerita ‘story’, akhlak ‘courtesy’, and the like. The linguistic expression nama harum ‘fragrant name’ refers to ‘fame or good reputation’, while membawa harum nama Indonesia (lit. ‘carry fragrance name Indonesia’) means ‘raising Indonesian’s prestige (on the international stage)’. These linguistic expressions exemplify the understanding of a pleasant smell as a personal or national identity, which is the positive counterpart of the metaphor BAD IS SMELLY that has been reported previously in other languages (Kövecses 2019). It is interesting to compare the Indonesian harum to the generic term for a pleasant smell in Thai, หอม, which has a similar metaphorical meaning “in demand, popular” (Wnuk et al. 2020: 953). To the best of
my knowledge, the conceptualization of a good reputation, or more generally
the novel metaphor GOOD IS FRAGRANT, has only been attested in Thai (Wnuk et al.
2020) and in Indonesian thus far.

4.2 Unpleasant smell terms

4.2.1 Amis or Hamis

Amis\textsuperscript{13} derives from Javanese (Stevens and Schmidgall-Tellings 2008). However, it
is commonly used in Indonesian and has a dictionary entry. Amis is probably best
glossed as a ‘fishy smell’. Fish is the most frequent lexical item associated with this
smell term in the corpus. Fish occurs almost twice as often as the second lexical
item, blood. Further frequent lexical items associated with amis are egg, duck
meat, meat, vaginal discharge, food, chicken meat, and seafood. In the corpus,
numerous sentences discuss ways of removing the unpleasant fishy smell when
preparing a dish containing fish, egg, duck meat, meat, or seafood, using local
herbs that have a more pleasant smell and delicious taste: jeruk nipis ‘lime’,
rempah ‘spice’, jahe ‘ginger’, kapur sirih ‘lime betel’, air kunyit ‘turmeric juice’,
daun kari ‘bay leaves’, daun ketumbar ‘coriander leaves’, asam jawa ‘tamarind’,
daun bawang ‘spring onion’, cengkeh ‘clove’, kecombrang ‘etlingera elatior’,
kapulaga ‘cardamom’, pandan ‘Pandanus leaves’, and the like. Pregnant women
are advised not to eat food that smells amis, but to drink jamu ‘traditional herbal
medicine’ instead so that their babies will not smell amis when they are born, as
shown in the example below.

\begin{align*}
(5) \quad \text{Ia minum jamu dua kali se-hari dan} & \\
\quad \text{NEG eat REL fishy.smell like fish and egg} & \\
\quad \text{S/he drank herbs twice a day and did not eat (food) which had a fishy} & \\
\quad \text{smell, such as fish and eggs.}' & \\
\end{align*}

Amis also has metaphorical senses, albeit rare. Of 1968 citations in the corpus, only
32 citations of amis are used figuratively, such as bau amis korupsi (lit. ‘odor
fishy.smell corruption’) politik berbau amis (lit. ‘politics have.odor fishy.smell’),
and bau amis perebutan kursi kekuasaan (lit. ‘odor fishy.smell seizing chair
power’). When associated with abstract concepts such as corruption, politics, or

\textsuperscript{13} Hamis is a variant orthography of amis.
power, the unpleasant smell term amis refers metaphorically to repudiating undesirable human behaviors, thus exemplifying the metaphor BAD IS SMELLY.

4.2.2 Apek or apak or hapak

Apek\textsuperscript{14} means ‘musty’ and refers to smells in different domains, such as rice grains, damp clothes, closed areas, or unventilated rooms. In its literal meaning, apak refers solely to olfactory perception. As the most important staple food in Indonesia, rice\textsuperscript{15} is one of the most frequently discussed topics associated with apak. Apart from the rice grains available on the market, there is also raskin,\textsuperscript{16} the subsidized rice grain for poor families. In example (8), the bad quality of raskin is described based on the five senses of sight ‘yellow’, touch ‘not good and broken’, taste ‘not delicious’, tactile ‘not soft and smooth’, and smell ‘musty’. The corpus contains various smell terms denoting rice grains depending on their olfactory qualities, such as apek/apak (81), wangi (12), tengik (10), aroma (10), bau (2), harum (2), and sangit (1).

(6) “Dari segi warna, raskin biasanya berwarna kuning, tekstur=nya tidak bagus dan pecah-pecah, rasa=nya kurang enak dan tidak pulen, bahkan beraroma apak,” kata=nya.

‘Raskin (rice for poor families) has a yellow color, its texture is not good and it is broken, its taste is not delicious and it is not smooth, it is even musty (lit. has a musty smell), s/he said.’

In its metaphorical use (10 citations from 935 occurrences), apak can refer to different abstract domains, such as dusta yang terlalu apak (lit. ‘lie which too musty’), which refers to the condemnation, contempt or criticism of a lie that has

\textsuperscript{14} Apak and hapak are variant orthographies of apak.

\textsuperscript{15} There are three Indonesian words for rice, namely padi ‘rice plant’ (rice in the husk), beras ‘rice grain’ (rice with husk removed, sometimes also translated as ‘uncooked rice’), and nasi ‘cooked rice’.

\textsuperscript{16} Raskin is an acronym for beras untuk keluarga miskin, ‘rice for poor families’ (KBBI). Raskin’s price is below the market price, and is subsidized by the Indonesian government (https://id.wikipedia.org/wiki/Raskin, accessed 28 April 2019).
long been kept hidden and causes bad effects, *bau apek penjajahan* (lit. ‘smell musty colonialism’), which refers to the colonial supremacy, which has been feared and rejected, and *politik uang yang membuat pemerintah jadi bau apek* (lit. ‘money politics that made government become smell musty’), which refers to the attitude toward a government that is involved in money politics, which is disapproved of.

### 4.2.3 Tengik

*Tengik* refers to the unpleasant smells of rancid oil and rotten or foul food. *Tengik* is also associated with the smell of fat in general, such as margarine, the rice bran used for animal feed, and other animal feed, particularly when it becomes rancid and produces foul smells. Advice regarding how to keep oil or food from smelling *tengik* is found, as in the example below. *Tengik* is the olfactory reference used to warn about the danger of rancid oil, which can cause cancer, followed by the advice to avoid using the same oil for deep frying repeatedly. In an Indonesian talk show, not in the corpus, the doctor advised the audience about signs indicating that home-made tempeh had expired, and used the smell terms *amis* or *tengik*. Home-made tempeh is commonly not marked with an expiry date.

(7) *Simpan minyak goreng di wadah yang tertutup rapat, dingin, dan terhindar dari sinar matahari agar tidak terjadi oksidasi dan tidak mudah tengik.*

‘(You should) store cooking oil in a tightly closed container, (in a) cool place, and protected from the sun so that it doesn’t oxidize and doesn’t easily go rancid.’

Compared to other smell terms in Indonesian, *tengik* is the smell term that occurs most frequently with a metaphorical extension. Almost one-third of all the instances of *tengik* in the corpus are metaphorical (162 of 595 citations). Most of the metaphorical uses of *tengik* are meant and understood as expletives or swear

---

17 The talk show “Dokter Oz Indonesia” adopts the format of the American pseudo-scientific talk show with a similar name “The Dr. Oz Show” and discusses health and nutrition topics. The sentence was extracted from https://youtu.be/uMtdZdNGCuI, begin and end times 00:11:45.200–00:11:49.880, accessed 28 May 2021.
words with vulgar connotations, such as bajingan tengik ‘crook’ (lit. ‘bastard rancid.smell’) (30),\(^{18}\) or bocah tengik ‘juvenile delinquent’ (lit. ‘child rancid.smell’) (16), or penjahat tengik ‘criminal’ (lit. ‘criminal rancid.smell’) (10).

### 4.2.4 Anyir or hanyir

Anyir\(^{19}\) is a loanword from Javanese (Stevens and Schmidgall-Tellings 2008), but has a dictionary entry and is commonly used in Indonesian. Anyir is similar to amis; however, the first and second most common lexical items are reversed. The most common lexical item associated with anyir is blood, which occurs over twice as frequently as the second most frequent lexical item, fish. Further lexical items associated with anyir are egg, vaginal discharge, duck meat, meat, corpse, water, sperm, and the human body.

(8) Bau anyir darah sangat tajam menusuk.
    smell smell.of.blood blood very sharp stab
    ‘The rancid smell of blood was very pungent.’

Anyir also has metaphorical extensions (34 citations in 577 occurrences) in the abstract domains of corruption, words, oppression, rejection, war, and so on. Examples of metaphorical expressions are bau anyir korupsi ‘the rancid smell of corruption’ and kata-kata anyir ‘rancid words’, meaning defamation.

### 4.2.5 Sangit

Sangit is borrowed from Javanese (Stevens and Schmidgall-Tellings 2008); it is related to the rice bug (Leptocorisa oratorius) and to the unpleasant burnt or scorched odor of several other objects. The Indonesian and Javanese name for the rice bug is walang sangit; as walang means ‘grasshopper’, the literal meaning of walang sangit ‘sangit-smelling grasshopper’. In fact, the corpus search for the term sangit returned over 82\% of hits for walang sangit and 17\% for the remainder. Further lexical items associated with sangit are smoke, car, brake, tempeh, human body, burnt meat, burnt wire, charcoal, and clutch. According to an Indonesian online dictionary, sangit refers to the smell of scorched (rice) crust.\(^{20}\) The reference to rice crust emphasizes the importance of rice in the Indonesian culture. Also note that four other Austronesian languages, namely Kavalan, Thao, Paiwan, and Amis, have their own smell terms denoting the smell of burnt rice (Lee 2010, 2015).

---

\(^{18}\) The numbers in brackets refer to the frequency of a word’s occurrence in the corpus.

\(^{19}\) Hanyir is a variant orthography of anyir.

Bau sangit yang tajam dari bawah kap

smell burnt smell REL sharp from below hood

mesin memaksa kami bermalam di Pamenang.
machine force 1.PL.INCL stay overnight in Pamenang

‘The pungent burnt smell from under the machine’s hood forced us to stay overnight in Pamenang.’

Sangit can also be used metaphorically (14 citations in 461 occurrences). The most frequent metaphorical expression, mati sangit (lit. ‘dead burnt smell’), means ‘to die ridiculously in vain’, and is a type of pun showing condemnation to counter the rhyming expression of mati syahid (lit. ‘dead martyr’) ‘to die as a martyr’ believed to be “enjoined by Islam” (Hassan 2007: 1041). The second most frequent metaphorical mapping is in the abstract domain of corruption, as in bau sangit korupsi (lit. ‘smell burnt smell corruption’), thus showing disapproval.

4.2.6 Pesing

Pesing derives from Javanese (Robson and Wibisono 2002) and is mainly associated with the stench of urine. The 10 most frequent lexical items associated with pesing are toilet, urine, room, bathroom, pork, body, clothes, place, terminal, and bed. It is important to note that pesing does not mean the ‘smell of urine’. Olfaction is not related to objects, but to odor (Cavedon-Taylor 2018). Thus, smell terms do not refer to objects such as urine, but to odors. Therefore, when a slice of pork is described as being pesing in Indonesian, it does not mean that the pork smells of urine, nor is this expression metaphorical. The pork is claimed to actually emit an odor referred to by the smell term pesing. It is interesting to note that three other Austronesian languages, Paiwan, Seediq (Truku), and Amis, also have their own terms for denoting the smell of urine (Lee 2010, 2015).

Konsumen daging babi sering mengeluhkan bau pesing pada daging babi.

consumer meat pig often complain odor on meat pig

‘Pork consumers often complain about the urine stench in pork.’

Pesing is also used in a figurative sense, albeit less frequently (five citations in 419 occurrences). Two figurative meanings are attested. First, pesing is used metaphorically to express criticism, as in parpol baunya pesing (lit. ‘political party its smell urine.stench’), bau pesing kekuasaan (lit. ‘smell urine.stench power’), and situasi berbau pesing (lit. ‘situation smell urine.stench’), which are expressions of criticism toward a political party, toward authority, or toward a situation, respectively. The second figurative meaning is a matter of metonymy. Pesing refers derogatively to the childhood of a person considered as still having a urine stench.
to refer to a point of time that is far in the past and sometimes to the person’s inexperience or immaturity. *Kakek buyutmu masih bau pesing* (lit. ‘grandfather great-great-grandparents. you still smell urine.stench’) means that something has happened long before your great grandfather was still a child; that is, still smelled of urine.

### 4.2.7 Langu

*Langu* derives from Javanese. According to Alwi et al. (2005: 636), *langu* means “unpleasant odor or taste (regarding the smell or taste of undried tobacco leaves, the taste of raw cassava).”²¹ Our perceptions of smell and taste are closely related when consuming food or drink. In Indonesian, the smell term *langu* is also used to refer to taste; almost a quarter of hits for *langu* refer only to taste, not smell, whereas less than 1% refers both to smell and taste. Lexical items in the corpus show that *langu* refers to the unpleasant smell of vegetables or plants. The most frequent lexical items associated with *langu* are soybeans, soy milk, vegetables, carrots, the enzyme lipoxygenase, and fruit.

(11) \[
\text{Di samping itu, adanya proses pemanasan juga dapat menghilangkan bau *langu* kedelai.}
\]

‘In addition, the heating process could also eliminate the foul smell of soybeans.’

### 4.2.8 Perengus or prengus

*Perengus*²² derives from Javanese and is defined in the monolingual dictionary as “smell like the smell of a male goat (sheep)” (Alwi et al. 2005: 857).²³ In fact, the most frequent lexical items associated with *perengus* are goat meat or mutton, followed by goat milk, meat, goat, lamb meat, and mutton satay. Furthermore, *perengus* is also used to refer to the smell of a dish, milk, cigarette smoke, or duck meat. *Perengus* is an unpleasant smell, as indicated in an example in the corpus, ‘*Perengus* smell can spoil the flavor of the dish’. The corpus contains numerous sentences referring to different strategies to remove the unpleasant goat-like smell; for example, how to slaughter the goat or cook the goat meat and, in particular,

---

²¹ The author’s translation.

²² *Prengus* is the Javanese orthography of *perengus*.

²³ The author’s translation.
what kind of traditional strong spices should be added to the dish, such as tongseng seasoning (curry-based stew seasoning for mutton or beef), turmeric, garam masala, clove seeds, bay leaves, lemongrass, coconut water, cucumber juice, and so forth.

(12)  Bau **prengus** pada kambing terletak pada tanduk=nya. smell **goat.smell** on goat lie on horns=DEF

‘The **goaty smell** of goats lies in their horns.’

Perengus only occurs once in the corpus with its figurative meaning as a synesthetic metaphor; hence, it is presented with its gloss below. Synesthetic metaphors are mapped from one domain of perception (usually lower) onto another domain of perception (usually higher) (see also Strik-Lievers 2017). In this example, the lower domain of smell, lexicalized as prengus, is mapped onto the higher domain of sight, lexicalized as looked and dirty. The expression ‘at the beginning of the 20th century, the face of Islam in Europe still looked **prengus** and dirty’ conforms to the metaphorical concept of the image of Islam at that time, evoking an unpleasant smell and the dirty appearance that one wants to avoid.

(13)  Maklum-lah, awal abad ke- 20, wajah Islam di know-IMP early century ORD-20 face Islam in Eropa masih ter-lihat **prengus** dan kotor. Europe still DEAG-see **prengus** and dirty

‘It is commonly known that, at the beginning of the 20th century, the face of Islam in Europe still looked **prengus** and dirty.’

4.2.9 Bacin

Bacin derives from Javanese and refers to a foul or rotten smell, but has a dictionary entry and is used by Indonesian speakers. The most frequent lexical items associated with bacin are mouth, city, salty fish, drainage ditch, sweat, water, saliva, dog’s saliva, jungle, and rice field mud. In the example below, bacin refers to the foul and unpleasant smell of a fasting person, which nevertheless is considered as harum ‘fragrant’ in the eye of Allah.

(14)  Niscayalah bau **bacin** mulut orang yang undoubtedly smell **foul.smell** mouth person REL berpuasa itu lebih **harum** di sisi fast that more **fragrant** in side Allah daripada bau minyak kasturi. Allah than smell oil musk

‘Surely the **foul smell** of the mouth of a fasting person is more **fragrant** in the sight of Allah than the smell of musk oil.’
5 The structure of basic olfactory terms

In order to understand how the smell terms are related to one another and how they are structured, the present study applied a hierarchical agglomerative cluster analysis using the computer package R. Cluster analysis is one of a large number of corpus linguistic statistical methods that are “hypothesis-generating in nature” (Gries 2010: 285). Cluster analysis is “an exploratory data analysis technique, encompassing a number of different algorithms and methods for sorting different objects into groups” (Divjak and Fieller 2014: 405). In a cluster analysis, the similarity of members of the same group is high and the similarity of members of different groups is low. Cluster analyses are good tools for detecting patterns in large data sets, and have often been used in studies in different linguistics fields (Gries 2013, Ch. 5.6); among others, in a semantic study of Indonesian denominal verbs (Rajeg et al. 2019) and in a semantic study of Thai smell terms (Wnuk et al. 2020). A hierarchical agglomerative cluster analysis was used in this study to divide the smell terms into groups in such a way that the members of a group had similar lexical items while being dissimilar to members of other groups. In other words, smell terms with shared lexical items are grouped together as the branches of a tree called a dendrogram (Figure 2) which, unlike normal trees, – grows from the branches to the root (cf. Levshina 2015: 309–317).

Figure 2: Cluster analysis of Indonesian smell terms.
The present paper follows the procedure in Gries (2013: 336–349). A data frame in which the numbers of certain words were counted as the lexical items for smell terms was created (see the Supplementary material). Each cell records the frequency of occurrence of the co-occurrences of the two words; that is, the smell term and its lexical item. A cluster analysis is “applied to data sets consisting of \( n \) objects that are characterized by \( x \) characteristics” with the purpose of identifying structures in big data and in complex data sets that are “characterized by a large within-cluster similarity and a small between-cluster similarity” (Gries 2010: 285). The \( n \) objects clustered were the 12 smell terms, and the \( x \) characteristics were the 1011 lexical items (see the Supplementary material). A cluster structure was computed and represented in a dendrogram (Figure 2).

The results revealed by the analysis resemble a reversed family tree. The most closely related siblings are joined at the very low level on the right-hand side, and are then joined by cousins at the next level, and so on, with the root at the far left-hand side. In this dendrogram, vertical lines show joined clusters. The closer to the right the vertical lines are, the more similar are the members of those clusters. The long horizontal lines indicate more autonomous subclusters. There appear to be two main clusters that are joined at the far left by a vertical line, which indicates the two categories of smell terms: pleasant and unpleasant. Note that the cluster of unpleasant smells consists of branches with far more smell terms, meaning that there are more types of unpleasant smell terms than pleasant ones. In addition, the cluster analysis used in this study revealed that pleasantness was the underlying dimension of the smell term structure, which is in line with previous studies of smell lexicons (Burenhult and Majid 2011; Wnuk et al. 2020). The top five lexical items associated with each smell term are shown in descending order on each of the horizontal lines. For example, the five most frequent lexical items associated with the pleasant smell term harum are flower, spice, body, perfume, and fruit. To understand the relationship of the clusters in Figure 2, reference to the Supplementary material containing a complete list of the lexical items associated with each smell term is needed.

The “pleasant” cluster consists of three smell terms, harum, aroma, and wangi, which join each other on the vertical line close to the right-hand side of the tree and show more similarities than in the “unpleasant” cluster consisting of nine smell terms, which only join each other at the very far left of the tree, thus indicating greater differentiation. Aroma and wangi join each other much earlier, and are therefore more similar to each other than to harum, which they join later. All the members of the “pleasant” cluster share a wide range of similar lexical items with more than five tokens, including the body and other natural objects such as flower, fruit, spice, and tea, and artifacts such as perfume and room (see the Supplementary material). Wangi and harum also share three lexical items with more than
five tokens, namely clothes, hair, and leaves. *Aroma* and *harum* have three lexical items in common with more than five tokens: coffee, garlic, and mouth. The wide variety of shared lexical items in the “pleasant” cluster suggests that there does not seem to be a distinct category of pleasant smells; for example, a food smell being differentiated from perfume or a room’s smell. Instead, the “pleasant” cluster appears to have a general meaning, and therefore provides less specific information about the quality of a smell. The relatively high similarity among the members of the “pleasant” cluster is also found in other languages (Lee 2021; Wnuk et al. 2020). Furthermore, the members of the “pleasant” cluster have more tokens than do the members of the “unpleasant” cluster (see Table 1). The corpus study of smell terms in this paper is in line with the claim that extremely common words provide less information than do less common ones (Garcia et al. 2012).

The main cluster of basic smell terms for unpleasant smells, with considerably more branches, is divided into two subclusters. The first unpleasant subcluster, consisting of *amis* and *anyir*, has a vertical line that is very close to the right-hand side of the tree, meaning that the two members are extremely similar to each other but are very different from the members of the second “unpleasant” subcluster, which they join only shortly before they join the “pleasant” cluster. The first “unpleasant” subcluster is associated strongly with the smell of blood and fish. *Amis* and *anyir* also share similar four of the most frequent lexical items with more than five tokens: egg, duck meat, meat, and vaginal discharge (see the Supplementary material).

The second subcluster consists of the remaining basic smell terms for unpleasant smells, with vertical lines that are quite far from the right-hand side of the tree, meaning that they are extremely different from each other. The second subcluster also branches into two subsclusters; the first subscluster consists of two subsubsclusters, each of which branches into a small group. Compared to the other groups in the second subcluster, the subsubscluster of *tengik* and *apek* has a vertical line that is furthest to the right-hand side of the tree, suggesting that they are more similar to each other than are the members of the other groups. However, their long horizontal lines also indicate their dissimilarity: They do not share the top five lexical items (see Figure 2). Nevertheless, *tengik* ‘rancid smell’ and *apek* ‘musty’ do share lexical items with more than five tokens: body, bran, food, food rations, and rice grain (see the Supplementary material).

The second subsubscluster consists of *pesing* referring to ‘urine stench’ and *bacin* referring to the ‘smell of saliva’. Based on their long horizontal lines, *pesing* and *bacin* are noticeably dissimilar to each other and only share six lexical items with less than five tokens: city, corpse, drainage ditch, mouth, toilet, and water (see the Supplementary material). The second subsubscluster consists of two subsubsclusters, in which *sangit* ‘burnt smell’ is separated from a small group of
two smell terms: *perengus* ‘goaty smell’ and *langu* ‘foul smell’. *Perengus* ‘goaty smell’ and *langu* ‘foul smell’ are very dissimilar to each other and only share two lexical items with less than five tokens: soy milk and dish/cuisine (see the Supplementary material). *Langu* ‘foul smell’ and *sangit* ‘burnt smell’ share three lexical items with less than five tokens: soybeans, tofu, and tempeh. *Perengus* ‘goaty smell’ and *sangit* only share one lexical item with one token: cigarette smoke. In summary, all the members of the second subcluster are highly dissimilar to each other and rarely share similar lexical items. Unlike basic smell terms for pleasant smells, basic smell terms for unpleasant smells are not used to describe the olfactory quality of a wide variety of objects. This suggests that the meanings of unpleasant smell terms are very specific. In addition, basic smell terms for unpleasant smells occurred less often than did basic smell terms for pleasant smells (see Table 1). This finding supports the claim that unpleasant words carry more information because of their rareness (Garcia et al. 2012).

6 Discussion

The present paper examined how Indonesian smell terms lexicalize some odors but not others, and provided answers to questions about the olfactory classification system and what it tells us about human behavior. Applying a cluster analysis using data from a large corpus of written registers enabled the present study to reveal the singular dimensionality of olfaction correlating with pleasantness and the striking difference in lexicalization and frequency. There are more negative types in Indonesian, but more positive tokens.

This key finding supports the so-called Pollyanna Hypothesis concerning a universal tendency toward a positivity bias in human language (Boucher and Osgood 1969), according to which positive words are potentially more prevalent and are used more diversely (Dodds et al. 2015), while negative words are more dissimilar and convey more information than do positive words (Garcia et al. 2012). It would appear that talking about pleasant smells is more vital for Indonesian speakers. Embarrassment and taboos have also been suggested as explanations for why fewer bad smells were enumerated in a free listing task in Boholano, in which individuals were asked to provide a list of possible smells that would then be collected by the researcher (Beer 2014). However, this may have stemmed from the methodological difference. Texts with vulgar or pornographic content are found in abundance in the corpus used for the present study, as people can write anonymously online. A more reasonable motivation lies in the properties of the smell terms and “the range of exemplars which are being encapsulated with a term” (Levinson and Majid 2014: 411). Pleasant smell terms do not differentiate among
distinct characteristics of smells, and can therefore be used more frequently to refer to the generally pleasant smells of a wide range of exemplars. On the other hand, terms for unpleasant smells “identify the precise properties of smells” (Levinson and Majid 2014: 411) and have a more restricted scope of use; therefore, they occur somewhat rarely.

There is a striking cross-linguistic similarity regarding the difference in olfactory lexicalization. In Jahai (Burenhult and Majid 2011), Luwo (Storch 2013), and Thai (Wnuk et al. 2020), there are fewer terms for pleasant smells than there are for unpleasant smells. Similarly, there is only one term for a pleasant smell and several terms for unpleasant smells in Kavalan, Paiwan, and Thao (Lee 2010), in Amis (Lee 2015), and in Cha’palaa (Floyd et al. 2018). However, not all languages with a sizable inventory of smell terms have a specific word dedicated to pleasant smell; for example, Buli (Bubandt 1998), Seediq (Lee 2010), and Seri (O’Meara and Majid 2020). Remarkably, they have a similar strategy in constructing a compound to refer to pleasant smell, which is by adding the root meaning ‘good’. In Buli, paù mafia means ‘pleasant smell’; paù is the term for neutral smell, and mafia means ‘good’ (Bubandt 1998: 74). In Seediq, suyan kanux refers to a “pleasant smell” (Lee 2010: 108); kanux is a term for neutral or unpleasant smells, and suyan means ‘good’. In Seri, ihasii quiipe refers to a ‘pleasant smell’; ihasii is the verb for neutral smell, and quiipe is the root verb that means ‘be good’ (O’Meara and Majid 2020). Oceanic languages of northern Vanuatu, a branch of the Austronesian languages have yet another striking strategy for constructing terms for pleasant smells, which is by the reduplication of the root of smell term that has the pejorative meaning ‘to smell bad’ (François 2021). In Lo-Toga, qune is glossed as ‘stink’ and qunequne as ‘fragrant’, whereas in Mwotlap, qôn means ‘smell’ and qôngôn means ‘fragrant’ (François 2021).

Based on the discussion above, I propose the following Pollyanna Hypothesis for basic smell terms across languages:

i. If a language contains smell terms, then it also contains more smell terms denoting unpleasant smells than pleasant smells (more negative types).

ii. If a language contains smell terms, then there are more positive tokens in language use.

More languages with large corpora need to be investigated to confirm the Pollyanna Hypothesis with regard to smell terms.

Further striking cross-linguistic similarities can also be found in the phenomenon of smell terms that are used for the names of plants and animals. In Thai, målêeng sàap ‘cockroach’ literally means sàap-smelling bug, whereas sàap is the smell term denoting unpleasant smells associated with certain animals (Wnuk et al. 2020: 953). In Indonesian, walang sangit is the name of a rice bug, whereas
sangit is the smell term denoting the unpleasant burnt or scorched smells of several objects (see the Supplementary material). This phenomenon is also found in an unrelated language: In Seri, which is spoken in Mexico, hant csii literally means ‘the land that (one) smells’ and is the word for ‘cow’, in which csii is the transitive verb ‘smell (something)’ (O’Meara and Majid 2020). Remarkably, målèng sàap in Thai and walang sangit in Indonesian are the most frequent objects that are associated with the respective smell terms.

Smell terms are also used as plant names. For example, in Siri, hehe ccotxta is a plant name, whereas ccotxta ‘stink’ is a smell predicate (O’Meara and Majid 2020: 371). In Indonesian minyak wangi ‘oil fragrant’ is the general term for ‘perfume’ and akar wangi ‘root fragrant’ is a plant name, with wangi ‘fragrant’ as a smell term. Of note, minyak wangi ‘oil fragrant’ and akar wangi ‘root fragrant’ are prototypes; that is, the most frequent lexical items associated with the smell term in the corpus.

Wnuk et al. (2020) noted remarkable similarities between Thai and Kapsiki (van Beek 1992), two unrelated languages, in the “musty” cluster, which is also found in Indonesian. A further striking cross-cultural similarity is also found in the cluster “blood/fish”. Apart from Indonesian, other languages around the world have grouped the smells of blood and fish – sometimes including the smell of meat – as one cluster; for example, Buli (Bubandt 1998), Javanese (Septiawati 2010), Boholano (Beer 2014), Kavalan (Lee 2010), Amis (Lee 2015), Jahai (Burenhult and Majid 2011), Maniq (Wnuk and Majid 2014), Cha’palaa (Floyd et al. 2018), Kapsiki (van Beek 1992), and Thai (Wnuk et al. 2020). However, the “blood/fish” cluster is absent in the following languages: Paiwan, Seediq, Thao (Lee 2010), and Seri (O’Meara and Majid 2020).

There is also a noticeable cross-linguistic similarity in the “urine” cluster. Apart from Indonesian, other languages also have a dedicated smell term for the stench of urine: Javanese (Septiawati 2010), Buli (Bubandt 1998), Boholano (Beer 2014), Cha’palaa (Floyd et al. 2018), Kavalan (Lee 2010), Amis (Lee 2015), Kapsiki (van Beek 1992), Luwo (Storch 2013), and Jahai (Burenhult and Majid 2011). However, there are also languages that do not have a dedicated smell term for the stench of urine, Kavalan, Thao (Lee 2010), Seri (O’Meara and Majid 2020), and Thai (Wnuk et al. 2020). It is interesting to note that Indonesian exhibits a number of patterns that are also found in many other languages.

Furthermore, as noted previously by Wnuk and Majid (2014: 134), smell “does not remain constant over time”. An object can smell differently in different states. For example, hog badgers can smell different depending on their bodies and the seasons. During the dry season, they are fat and have an aromatic odor, denoted as caŋə, which is a term for a pleasant smell in Maniq that is also used to refer to the odor of various foodstuffs. This pleasant odor leads to them to be hunted by the Maniq-speaking hunter-gatherers. However, during the wet season, wet hog badgers smell
bad and are not hunted. In a similar vein, seven Indonesian smell terms, including the pleasant ones *aroma* (4), *wangi* (4), and *harum* (10), and the unpleasant ones *amis* (30), *tengik* (2), *anyir* (7), and *pesing* (4), for example, have attested uses for the odor of “vagina” in this study. Nonetheless, the fact that these smell terms are used to describe the odor of the vagina makes them by no means synonymous. First, olfactory experience is related to odors, not to objects (Cavedon-Taylor 2018). Second, an object can emit different odors depending on its condition and can therefore be referred to using different smell terms (Wnuk and Majid 2014). Furthermore, smell terms may also code affect (Levinson and Majid 2014).

As it is mainly used as a second language by its speakers, Indonesian is somewhat different from the other languages of the small speech communities mentioned above. Eight of the dozen basic smell terms in Indonesian are loanwords from Javanese; that is, *wangi* (also from Sundanese and Balinese), *amis*, *anyir*, *sangit*, *pesing*, *langu*, *perengus*, and *bacin*, whereas one smell term, *aroma*, is a loanword from Dutch or English. This is not surprising when we consider the long history of contact between speakers of Malay-Indonesian and other cultures and the role of Indonesian as a regional lingua franca. In his study on loanwords in Indonesian, Tadmor (2009b) identified approximately 34% of the words as being loanwords in his database; 8.9% of the total number of words in the database are loanwords from the greater Java area (Java, Bali, and Madura). Loanwords from these regional languages entered Indonesian through second-language speakers who were “imposing words from their indigenous vocabulary on their second language” (Tadmor 2009b: 706). This suggests that smell vocabulary can also be found in the donor languages; for example, in Javanese (Septiawati 2010). It can be assumed that the influence of the Javanese culture is widespread throughout Indonesia. Javanese is not only the largest but is also the most ubiquitous ethnic group in Indonesia, meaning that Javanese populations can be found in almost all the provinces (Ananta et al. 2015: Ch. 4). However, the borrowed smell terms fare well in terms of psychological salience as they are manifested in dictionaries (Alwi et al. 2005; Stevens and Schmidgall-Tellings 2008) and are attested in online written communication (Goldhahn et al. 2012).

The findings in this study reveal that Indonesian has a rich smell vocabulary; more generally, smells are linguistically codable in Indonesian. However, having sizable smell vocabularies does not necessarily equate to psycholinguistic codability (Levinson and Majid 2014); that is, the ability to talk about specific smell qualities when forced (Majid and Kruspe 2018). Following a cross-linguistic study, we learn that, in Malay, a language that is closely related to Indonesian, “shape is the most codable of the senses on average and smell is the least codable” (Majid et al. 2018b). The significant factor in the psycholinguistic codability of olfaction lies in the subsistence mode, as has been demonstrated in a comparative study.
between hunter-gatherers and swidden-horticultural societies of closely related languages in the Malay Peninsula (Majid and Kruspe 2018). Both languages have smell lexicons, but only the hunter-gathering Semaq Beri can use the smell lexicons consistently. The swidden-horticulturalists Semelai struggled to name odors and performed similarly to American speakers in a controlled odor- and color-naming experiment (cf. Majid and Burenhult 2014). I leave it to future research to determine whether Indonesian speakers would perform similarly to Malay speakers in a study of coding perception, as their lifestyles and subsistence are not particularly different.

7 Conclusion

The data studied in this paper reveal that Indonesian basic smell terms are structured in the hedonic dimension related to pleasantness, and that the pleasant smell terms have more tokens whereas the unpleasant smell terms have more types. The cluster analysis conducted in this study reveals that unpleasant smell terms are more distinct from each other than are pleasant smell terms, as members of the former group are less similar to each other while members of the latter group are used to refer to a wider variety of lexical items. This finding is in line with the previous claim that languages with elaborate odor terms “make more distinctions between types of unpleasant odors than types of pleasant odors” (Majid 2021b: 421); hence, there are more negative types. Pleasant smell terms, on the other hand, are more similar to each other, have a relatively general meaning of pleasant smells, and may also code “affect” (Levinson and Majid 2014: 411) instead of simply the odor quality; hence, there are more positive tokens. The hypothesis for smell terms proposed in this paper offers some partial support to the Pollyanna Hypothesis that has been developed over the years (Boucher and Osgood 1969; Dodds et al. 2015; Garcia et al. 2012).

Furthermore, the present paper provides a new approach to the study of smell terms, which enhances our understanding of olfactory language. In particular, the study reveals that the well-established corpus linguistic method is a powerful tool for exploring the meaning and usage of words, and can lead to a refinement of word descriptions and add definitions of words beyond a dictionary’s coverage, not only in the literal but also in the figurative meaning of the smell terms. Moreover, the cluster analysis allows us to determine patterns that cannot be identified by consulting a dictionary. This method can definitely be used in the future to shed more light on our knowledge of the language of olfaction. The results of the corpus analysis in this study contribute empirically to the on-going research on perception in general and on smell language in particular. A theoretical contribution of this study is the confirmation of the Pollyanna Hypothesis.
Abbreviations

| 1 | first person |
| 3 | third person |
| COP | copula |
| DEAG | deagentive |
| DEF | definiteness |
| EXCL | exclusive |
| IMP | imperative |
| INCL | inclusive |
| NEG | negation |
| PL | plural |
| RED | reduplication |
| REL | relativizer |

Data availability

The full results from this study can be found online at: https://doi.org/10.17605/OSF.IO/W8KS9.

Acknowledgments: I thank Karlina Denistia and Gede Primahadi Wijaya Rajeg for sharing their knowledge of R with me. I am very grateful to the anonymous reviewers for their constructive feedback that significantly improved the paper. I would also like to thank I Wayan Arka, Nils Ole Bubandt, Niclas Burenhult, Simon Devylder, Alice Gaby, Birgit Hellwig, Nicholas J. Long, Asifa Majid, Bradley McDonnell, Thomas Poulton, and Ewelina Wnuk for their helpful comments and critiques in relation to previous drafts. For their feedback, I also wish to thank the participants of the online Sharing Session on Linguistics 2021 at Udayana University.

References


Majid, Asifa & Niclas Burenhult. 2014. Odors are expressible in language, as long as you speak the right language. *Cognition* 130(2). 266–270.


**Supplementary Material:** The online version of this article offers supplementary material (https://doi.org/10.1515/cog-2021-0092).