Smell terms are not rara:
A semantic investigation of odor vocabulary in Thai

Abstract: Large lexica of basic smell terms are considered to be restricted to a handful of small languages of non-industrialized societies. Accordingly, they are thought to belong to the sphere of rara within lexical typology (Plank 2001). However, smell lexica might in fact be more common than previously suggested. In this article, we discuss the case of Thai – a language with a population of tens of millions of speakers – which defies this assumption. We show Thai has a sizeable lexicon of terms for olfactory qualities, and investigate their semantics using a multi-method approach. In particular, we demonstrate a novel use of exemplar listing where – in addition to giving insights into the terms’ extensions – exemplar data is used to reveal the structure of the lexicon. Additionally, we use corpus data to provide complementary information on meaning and usage, thereby showing the advantage of multi-method approaches. Overall, the findings suggest smell lexica are not rara, and their distribution in the world’s languages goes beyond the stereotypical cases of languages spoken by small-scale societies.

Keywords: smell term, linguistic rarum, semantics, exemplar listing, Thai

1 Introduction

Smell has a reputation of being a poorly lexicalized domain. Linguists, philosophers, and psychologists have all noted the apparent lack of dedicated smell terminology in language. For example, Weisgerber in his article “The sense of smell in our languages” (1928) spoke of a gap in our vocabularies, and equated our difficulty in assigning odors to classes with the difficulty faced by an amnesia patient when asked to do the same for colors. More recently, Sperber (1974) suggested there is no such thing as a semantic field of smells since there are no specific smell lexemes, unlike color. The view that smell is a “muted sense” (Ackerman 1990:6) is widespread (Olofsson and Gottfried 2015) across disciplinary boundaries. We are said to be “astonishingly bad” at naming smells (Yeshurun and Sobel 2010), so much so that olfactory function tests used in clinical settings typically do not rely on independent generation of labels by participants, but instead involve multiple-choice tests with ready-made answers (Doty, Shaman and Dann 1984).

Despite its fame as an elusive domain, the claims about smell do not generalize to all languages. There are accumulating examples of elaborate smell vocabularies across the world (Aschmann 1946; Hombert 1992; van Beek 1992; Demolin et al. 2016; Blench and Longtau 1995; Shepard Jr. 1999; Storch 2004; Storch and Vossen 2006; Lee 2010; Burenhult and Majid 2011; Tufvesson 2011; Wnuk and Majid 2014; Majid and Burenhult 2014; Beer 2014; Lee 2014; O’Meara and Majid 2016; Majid and Kruspe 2018; O’Meara et al. in press; Floyd, San Roque and Majid 2018). However, dedicated smell lexica seem to be missing from English and other large Indo-European languages (Plank and Plank 1995). The causes for this apparent gap are not clear. One account suggests that the absence of elaborate smell lexica could be linked to the deodorization of the environments and derogation of smell that took place in the Western

Corresponding author: Ewelina Wnuk, Department of Anthropology, University College London, London, WC1H OBW, United Kingdom & Max Planck Institute for Psycholinguistics. E-mail: e.wnuk@ucl.ac.uk
Rujivan Laophairoj, Faculty of Liberal Arts, Ubon Ratchathani University, Warinchamrap Ubon Ratchathani 34190, Thailand. E-mail: pormin_anne@hotmail.com
Corresponding author: Asifa Majid, Department of Psychology, University of York, Heslington, York, YO10 5DD, United Kingdom & Radboud University Nijmegen. E-mail: asifa.majid@york.ac.uk
world in the eighteenth and nineteenth centuries (Classen, Howes and Synnott 1994), with figures such as Broca and Freud playing instrumental roles (McGann 2017).

Irrespective of whether this account is correct, the fact remains that smell terminology is considered a rarity associated with a handful of small languages. Rather than being a systematically surveyed domain, its place remains within the sphere of linguistic oddities. In fact, “a sizeable inventory of basic smell terms, i.e., one with more than two or three items” is explicitly listed as a *linguistic rarum* in the online database *Das grammatisiche Raritätenkabinett* (Plank 2001). What further contributes to this conclusion is the scarcity of detailed descriptions of smell terminology. If we compare our knowledge about smell terms with color terms, for instance, for which large-scale cross-linguistic data exists (Berlin and Kay 1969; Kay et al. 2009), it is clear more descriptive work is needed in order to develop a typology of smell lexica.

Linguistic rara are often thought of as being characteristic of small languages (Wohlgemuth 2011; Nettle 1999), and in the case of smell vocabulary – aside from some brief reports (de Sousa 2011) – the published literature can indeed create an impression they are especially prevalent in small-scale non-industrialized societies. The tacit implication is they would not be found in larger languages of industrialized societies. However, recent work suggests Thai might in fact be such a language (de Valk et al. 2017; Wnuk et al. 2017).

There are also methodological challenges in studying smell vocabulary, which could be impeding progress in the field. Odor stimuli do not always elicit smell vocabulary (even if the language has it), since the selected odors might not be representative of the local smell terms. This is compounded by the fact that our understanding of olfactory psychophysical space is still nascent in comparison to, say, color space (cf. Keller et al. 2017). For instance, a commonly used stimulus set in odor identification studies – Sniffin’ sticks (Hummel et al. 1997) – includes mostly pleasant-smelling odors. However, smell vocabularies often lexicalize mostly unpleasant smell qualities, so these stimuli might not be tapping into the terms’ extensions (cf. Majid, Burenhult, et al. 2018).

In this article, we use a multi-method approach in order to explore the smell terminology of Thai. We investigate existent language resources (i.e., corpus, dictionaries), as well as collecting original data from native speakers. We highlight in particular exemplar listing, a method previously employed to investigate the extensional semantics of smell terms (e.g., Shepard Jr. 1999; Wnuk and Majid, 2014). In a novel application of this technique, however, we show that by comparing exemplars across terms, we can derive an implicit measure of similarity and reveal the structure of the domain as a whole, thus gaining access to aspects of intensional semantics.

2 Talking about smell in Thai

Thai is the national language of Thailand and the largest member of the Tai–Kadai family spoken by over a half of the total 80-90 million Tai-Kadai speakers (Diller 2008). It is an isolating language, with a basic SVO constituent order. Most of its native vocabulary follows a monosyllabic pattern. The language makes heavy use of compounding and reduplication in derivation (Haas 1964; Iwasaki and Ingkaphirom 2005).

Basic smell predicates in Thai, following Viberg’s (1984) verbs of perception paradigm, are the following: the controlled act of smelling by an agent is described with the verb *dom* ‘to smell, to sniff’, whereas the uncontrolled experience is expressed with the complex verb *dài glin* ‘to smell’ (get smell). There is no general phenomenon-based smell verb (as in *It smells good,* but instead this generic meaning is expressed with a construction involving the noun *glin*, as exemplified in (1). Specific odor qualities are expressed using a dedicated lexicon of smell terms, described in detail in Sections 3-5.
(1) ภูมิ ɔ̂ ลม แย่ ร้อง
smell 3SG be.strong
‘Its smell is strong.’

Looking at word frequency data, smell talk in Thai is common. The general noun ภูมิ ‘smell’ is among the thousand most frequent words in Thai – it comes as 942nd in the Top 5000 word list of TNC (Thai National Corpus; Arroonmanakun, Tansiri and Nittayanuparp 2009) (compare this to the English noun smell, which is ranks as 2853rd in the COCA frequency list; Davies 2008; Winter et al. 2018; Floyd et al. 2018).

The prevalence of smell-related terms in the Thai corpus reflects the general salience of smell in everyday life, especially in relation to food and traditional medicine. Millions of Thais use pocket-size herbal inhalers (ยากรง) on an everyday basis, and Thai cooking and medicinal recipes use a variety of aromatic ingredients valued for their smell: e.g., ginger (ขิง), lemongrass (ตะไคร้), holy basil (กะเพรา), galangal (กะ่า), kaffir lime (มะกรูด), tamarind (มะขาม), and many others (Van Esterik 1988; Bamber 1998).

Commenting on food’s smell is common in everyday talk and is incorporated in cooking instructions on occasion, as in Example (2) from the Thai National Corpus.

(2) ต้ม ɔ̂ ส้วก รัก ผัก-ชี, ภูมิทัย, ผัก ผั๊ก
must put.in root coriander garlic black.pepper stir-fry with
khāw-sān phɔ̀ ��ām gò sāy nāam diat (...) milled.rice when be.fragrant then put.in water boil
‘You must put in coriander root, garlic, black pepper, stir fry with rice until it’s fragrant, then put in boiling water (...)’
(BIO030)

Smell is also significant in ceremonial religious practices such as, for example, incense burning, and can be the basis for religious prescriptions. For example, the Mahayana-Buddhist Vegetarian Festival tradition (เทศกาญจน์ จันจั่ง) practiced by the Chinese minority in Southern Thailand involves a prohibition on consuming what is referred to as “the five strong-smelling vegetables” because of their potential to stimulate passions (Ungpho 2010). Smells are also central to some health-related beliefs. For instance, in Northern Thailand women in the postpartum period are believed to be vulnerable to “wind illness” which according to common belief can be caused by strong smells (Muecke 1979; Liamputtong 2004).

3 Basic structural properties of the smell lexicon in Thai

Thai has a set of at least seven monolexemic smell terms, which include the stative verbs ฉัน, หอม, ชิน, กลิ่น, หนึ่ง, อภ, the noun ลม. In addition, there are several compounds, most of which are headed by one of the two most commonly used terms ฉัน ‘to be stinky’ and หอม ‘to be fragrant’, plus either a modifying term which comes from the set of the monolexemic terms listed above, e.g., อภ ‘to be stuffy/musty’ in ฉัน อภ (be.stinky be.stuffy/musty), or a non-smell term, e.g. ขี้ว ‘to be green’ in ฉัน ขี้ว (be.stinky be.green). These compounds are lexicalized and most are listed as entries in dictionaries. For the full list, see Section 4. Aside from the basic inventory, a number of regional terms exist, e.g., ขี้ว in Isan, and ชอง in Southern Thailand, exclusive to the varieties spoken in these regions.

Most of the monolexemic smell terms belong to the class of property-encoding words, analyzed as stative verbs or adjectives (Haas 1964; Iwasaki and Ingkaphirom 2005). They form independent predicates and frequently occur as modifiers of the general noun ภูมิ ‘smell’. The noun ลม typically occurs on its own, as a modifier in the compound ฉัน ลม, or as an
element of an associative construction, with *glin* ‘smell’ as head, or itself as head (where the juxtaposed NP is the smell object, cf. (7)). The restricted modifier *tú-(tú)* most frequently occurs in reduplicated form, and is a bound constituent typically combined with the verb *mën* ‘to be stinky’ in *mën tú-(tú)* or with the noun *glin* ‘smell’ in *glin tú-(tú)*.

Examples (3) – (8) below illustrate typical responses employing this vocabulary elicited in an odor naming task (de Valk et al. 2017), in which native speakers were asked to smell and describe various odor stimuli. Note that for practical reasons the glosses in this section are simplified and are largely based on dictionary definitions. More detailed meaning descriptions follow in Section 4.1.4. Intensification of smell terms can be achieved by lexical means as in (5) and (8), or by reduplication with emphatic tone, with the first element surfacing as a high tone, as in (9). A simple reduplication with no tonal change, as in (6), results in the meaning of ‘fairly, sort of’, similar to the meaning of the English suffix –*ish* (Smyth 2002).

(3) *áp*  
be.stuffy/musty  
‘(it’s) stuffy/musty’

(4) *glin*  
*tú-tú*  
smell  be.foul-be.foul  
‘foul smell’

(5) *chún*  
máak  
smell.pungent  very  
‘(It) smells very pungent.’

(6) *hūnn*  
hūnn  
be.foul/rancid  be.foul/rancid  
‘(It’s) sort of foul/rancid.’

(7) *sàap*  
khon  
rank.smell  person  
‘a rank smell of a person’

(8) *hɔm*  
reeng  máak  gəən  
be.fragrant  be.strong  much  too  
‘fragrant (which is) too strong’

(9) *mën*  
mën  
be.stinky  be.stinky  
‘(It’s) really stinky!’

4 Extensional semantics of Thai smell vocabulary

In order to understand the extensional semantics of the smells terms in greater depth, we carried out an exemplar listing task in which we asked speakers to list the best examples of each smell term. Previous investigations have shown this method can yield rich data (Shepard Jr. 1999; Wnuk and Majid 2014; O’Meara and Majid 2016). In order to see the resulting patterns as clearly as possible and identify the most frequently recurring exemplars, we tested a large sample of speakers. The goal was thus, first, to establish the exemplar range for each term and, second, to identify the prototypical exemplars among them.

4.1 Method

4.1.1 Participants

Participants were 301 native speakers of Thai (236 female, 56 male, 9 unknown) between 18 and 24 years of age (*M* = 20), recruited at the Ubon Ratchathani University and Chiang Mai
University. Participation was voluntary. Instead of providing small payment to each individual participant, we paid a total of 7000 THB as scholarship to a small group of students who were in need of financial support. This decision met with consensus among a representative group of participants consulted before carrying out the study and fit the institutional ethical protocol.

4.1.2 Materials

The task was administered as a written questionnaire in Thai. A list of 22 items used in the task was generated in separate interviews with two native speakers of Thai. We included words and compounds that people considered to refer to smell. When selecting the target compounds, we singled out those which appeared to have a high degree of lexicalization and had an idiomatic meaning denoting a specific odor quality, e.g., mên khaaw ‘be stinky (e.g., of some green vegetables)’ (be.stinky be.green), and excluded those which were instances of regular modification, e.g., hōm ṭon ḍon ‘be mildly fragrant’ (be.fragrant be.soft be.soft). We also decided to include mên-headed compounds for all of the unpleasant terms in our list, i.e., mên khaaw, mên ṭūn, mên ṭō, mên ṭō chúan, mên tū-tū. At first sight, these terms and the corresponding compounds do not appear to differ in meaning, e.g., khaaw and mên khaaw both refer to a fishy/bloody smell. However, the presence of mên makes the negative valence explicit and could be linked to differences in salient exemplars, so both variants were included. Table 1 below lists all the terms used in the task together with glosses for modifiers. No dictionary definitions were available for a few terms (hōm wāan, ṭō chúan, mên ṭō chúan), suggesting a lesser degree of lexicalization.

**Table 1: Smell terms used in the task.**

<table>
<thead>
<tr>
<th>Monolexemic terms</th>
<th>Term</th>
<th>Dictionary definition</th>
<th>Modifier / Second element</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
<td>Sources: Haas (1964); NECTEC’S LEXITRON project</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mên</td>
<td>to smell bad, stink, be foul-smelling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hōm</td>
<td>to be fragrant, odoriferous, sweet-smelling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chúan</td>
<td>to be strong (of odors), pungent (as the odor of strong tobacco)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>khaaw</td>
<td>to be fishy (in smell)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ṭūn</td>
<td>rancid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ṭō</td>
<td>smelly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sāap</td>
<td>an unpleasant, rank smell (as of body odors)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tū-tū</td>
<td>stinky, smelly; odorous; reeking; foul</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Compounds**

<table>
<thead>
<tr>
<th>mên + smell term</th>
<th>mên khaaw</th>
<th>mên ṭūn</th>
<th>mên ṭō</th>
<th>mên ṭō chúan</th>
<th>mên sāap</th>
<th>mên tū-(tū)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mên + other term</td>
<td>mên biat</td>
<td>mên priaw</td>
<td>mên khaaw</td>
<td>to stink, smell foully of raw meat, blood, fish</td>
<td>to be spoiled’</td>
<td>biat ‘to be spoiled’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rancid, foul, strong-smelling, stale</td>
<td>to be sour’</td>
<td>priaw ‘to be sour’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>to smell musty, stuffy</td>
<td>to be green’</td>
<td>khaaw ‘to be green’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>to smell, stink (as of someone who needs a bath)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>to have a slightly unpleasant odor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hōm + other term</td>
<td>hōm grīn</td>
<td>hōm chūy</td>
<td>hōm ṭūn</td>
<td>to smell very bad (of spoiled food)</td>
<td>to be smoldering”</td>
<td>grīn ‘to be smoldering”</td>
</tr>
<tr>
<td></td>
<td>hōm wāan</td>
<td></td>
<td></td>
<td>to smell unpleasantly sour</td>
<td>to be left</td>
<td>chūy ‘to be left</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>to smell bad; according to some speakers, to have an odor of crushed green leaves</td>
<td></td>
<td>hūn ‘to go back’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td>wāan ‘to be sweet’</td>
</tr>
<tr>
<td>áp + other term</td>
<td>áp chúan</td>
<td></td>
<td></td>
<td>scented, sweet-smelling, fragrant</td>
<td>to be damp, moist, humid’</td>
<td>chúan ‘to be damp, moist, humid’</td>
</tr>
</tbody>
</table>
4.1.3 Procedure
Participants completed the task in a classroom. Each person received one of two versions of the questionnaire with a different semi-random order, such that monolexemic terms were listed directly next to corresponding compounds (e.g., khaaw and mën khaaw). The entire procedure was carried out using Thai, including the language and script of the questionnaire. The questionnaire started with a brief intro text stating the purpose of the study, a note on the voluntary character of participation, and the possibility of the participant to freely withdraw from the study at any time. Questions regarding basic demographic information (age, gender, province of birth) were also included in the questionnaire.

The instructions for the exemplar listing were as follows: *Glin X khun aray? Yök tua yâang sing không thì mii glin taam kham riak glin không lâang tâi pay mii (yâang nòoy tāo maa 1 tua yâang). Khun sàamâat chây tua yâang nai gaan âthiabaeu glin màak gwàa 1 glin, hâak sing nân sàamâat pen tua yâang thì dìi không glin.* (‘What is (of) smell X? Provide examples of things that have a smell described by the following terms for smells (give at least one example). You can use the same example for more than a single smell, as long as that thing is a good example of the smell’). Before beginning, the experimenter first read out the introductory note aloud and made sure its content was fully understood by all participants. Following common practice in similar free listing studies (Battig and Montague 1969; Storms 2001), we set a time limit to the task, in this case 15 minutes. The smell terms were listed in separate boxes, each containing numbered rows for up to 5 exemplars per term.

4.1.4 Results
For all smell terms in the questionnaire, we generated a summary list of exemplars with response frequencies (see Table 2). Some highly similar responses were merged, e.g., nîu tâai ‘dead mouse’ (mouse be.dead), sâak nîu ‘carcass of a mouse’ (carcass mouse), and sâak nîu tâai ‘carcass of a dead mouse’ (carcass mouse be.dead); sîa-phâa mây sâk ‘unwashed clothes’ (clothes NEG to.wash), sîa-phâa thîi mây sâk ‘unwashed clothes’ (clothes which NEG to.wash), and the aspectually-marked sîa-phâa mây dâi sâk ‘unwashed clothes’ (clothes NEG ASP to.wash). However, if there was a difference in the level of specificity or another lexical difference which could be related to a potentially meaningful semantic difference, we did not merge such responses; e.g., we counted separately rôrong thâaw phâa bây ‘trainers’ and rôrong thâaw ‘shoes’; sîa-phâa mây hêeng ‘undried clothes’ (clothes NEG be.dry) and sîa-phâa piâk ‘wet clothes’ (clothes be.wet). We also disregarded explicit mentions of smell terms within the exemplar listing if these smell terms were identical to the target smell terms, e.g., sîa-phâa âp ‘musty clothes’ (clothes be.musty) was considered the same as sîa-phâa ‘clothes’ when listed for the term âp ‘to be musty’.

Table 2 below provides a list of the most frequently listed exemplars (i.e., listed by at least 5 participants) for all terms tested. Numbers in brackets indicate how many participants provided that response.

**Table 2**: Most common exemplars of examined smell terms. Numbers in brackets indicate how many participants provided the response. The exemplars are arranged in descending order beginning with the most commonly named ones. For dictionary definitions of the terms, see Table 1.
<table>
<thead>
<tr>
<th>Term</th>
<th>Exemplars</th>
</tr>
</thead>
<tbody>
<tr>
<td>chûn</td>
<td>perfume (84), garlic (50), onion (31), pepper (19), holy basil stir fry (15), flower of suicide tree (<em>Cerbera odollam</em>) (14), strong perfume (13), ammonia (11), suicide tree (11), bathroom cleaner (10), celery (8), mosquito repellent spray (8), spices (7), insecticide (7), holy basil (7), vinegar (6), shallot (6), wasabi (5), coriander (5), chili powder (5), chili (5), oil (5)</td>
</tr>
<tr>
<td>mên</td>
<td>excrement (84), garbage (83), fart (55), sewage (37), feet (18), shoes (17), bin (16), body (14), fermented fish (13), socks (12), rotten garbage (11), restroom (11), mouth (11), wet garbage (9), spoiled food (9), dead animal (8), sewer (8), dead mouse (7), car smoke (7), toilet (7), shrimp paste (6), polluted water (6), manure (6), durian (6), canal water (6), vomit (5), cigarette (5), canal (5), waste (5)</td>
</tr>
<tr>
<td>mên biút</td>
<td>spoiled food (88), spoiled milk (48), rotten food (40), spoiled rice (37), food (22), expired food (18), milk (17), garbage (8), expired milk (8), vomit (8), rotten rice (8), food leftovers (7), rotten milk (7), food left overnight (6), yogurt (6), spoiled curry (6), rotten thing (6), rice (5)</td>
</tr>
<tr>
<td>mên khlaw</td>
<td>vegetable (36), body (22), leaf (19), Chinese kale (19), grass (19), green vegetable (11), sewage (11), lettuce (10), long beans (10), coriander (8), sweat (6), spinach (6), armpit (5), Gotu Kola (<em>Centella asiatica</em>) (5), spoiled food (5)</td>
</tr>
<tr>
<td>mên prław</td>
<td>body (130), armpit (50), sweat (32), spoiled milk (14), spoiled food (12), person who didn’t shower (11), shoes (6), garbage (6), feet (6), socks (5), pickled vegetable (5)</td>
</tr>
<tr>
<td>khaaw</td>
<td>fish (131), blood (55), seafood (19), fresh fish (15), octopus (15), meat (14), fresh meat (14), raw fish (14), egg (12), clam (8), raw meat (8), beef (8), raw egg (7), animal meat (6), food (5), pork (5)</td>
</tr>
<tr>
<td>mên khaaw</td>
<td>fish (81), blood (44), seafood (20), menstruation (15), dead fish (13), fresh fish (13), meat (9), octopus (9), rotten fish (8), beef (7), fresh meat (6), raw fish (6), clam (5), fermented fish (5), egg (5)</td>
</tr>
<tr>
<td>hûn</td>
<td>oil (45), old oil (11), wasabi (11), lard (10), vegetable oil (9), pork cracklings (8), fried food (6), garlic (6), reused oil (5), old vegetable oil (5)</td>
</tr>
<tr>
<td>mên hûn</td>
<td>oil (37), old oil (18), vegetable oil (18), lard (11), pork cracklings (11), gas (8), used oil (8), mothball (7), reused oil (6), old vegetable oil (6), body (5), suicide tree (<em>Cerbera odollam</em>) (5), spoiled food (5)</td>
</tr>
<tr>
<td>áp</td>
<td>clothes (51), shoes (29), wardrobe (28), unventilated room (28), undried clothes (28), room (27), socks (18), cabinet (9), narrow room (8), shirt (8), unwashed clothes (7), wet clothes (7), closed room (6), storeroom (6), bedroom (5), undried shirt (5), old room (5), car (5), clothes in wardrobe (5)</td>
</tr>
<tr>
<td>mên áp</td>
<td>shoes (55), undried clothes (43), clothes (40), wardrobe (39), socks (34), undried shirt (9), shirt (9), unwashed shoes (8), unwashed clothes (7), wet clothes (7), room (7), restroom (6), socker (6), feet (6), bedroom (5), trousers (5), body odor (5), car (5), sweat (5)</td>
</tr>
<tr>
<td>áp chûn</td>
<td>undried clothes (48), restroom (23), shoes (22), clothes (19), socks (18), wet clothes (16), wardrobe (11), towel (9), unwashed clothes (8), undried shirt (7), refrigerator (6), air after rain (5), underwear (5), armpit (5), air (5), rain (5)</td>
</tr>
<tr>
<td>mên áp chûn</td>
<td>undried clothes (49), clothes (26), shoes (24), wet clothes (24), restroom (17), socks (17), undried shirt (11), stagnant water (7), wet socks (7), undried shoes (6), underwear (6), sweat (6), unwashed clothes (5), doormat (5), unwashed shoes (5), musty (“áp”) clothes (5)</td>
</tr>
<tr>
<td>sâap</td>
<td>cockroach (39), water buffalo (30), mud (15), person who didn’t shower (15), dead mouse (13), dog (13), dead animal (12), animal (12), body (8), unwashed dog (7), clothes (6), carcass (6), cow (5), sweat (5)</td>
</tr>
<tr>
<td>mên sâap</td>
<td>cockroach (37), dead animal (23), dog (21), corpse (19), unwashed dog (18), water buffalo (18), dead mouse (16), mud (13), person who didn’t shower (13), carcass (10), mouse excrement (9), body (9), cat (7), cockroach excrement (5), cow (5), sweat (5)</td>
</tr>
<tr>
<td>tû-tû</td>
<td>fart (75), socks (20), shoes (15), body (15), fermented fish (15), excrement (13), feet (12), dog excrement (7), unwashed clothes (7), unwashed socks (6), dead mouse (5), armpit (5), garbage (5), rotten thing (5)</td>
</tr>
<tr>
<td>mên tû-tû</td>
<td>fart (78), excrement (35), dog excrement (19), fermented fish (19), socks (17), feet (16), shoes (13), dead mouse (13), body (13), armpit (8), unwashed socks (7), garbage (7), rotten egg (6), person who didn’t shower (6), rotten food (5), unwashed dog (5), unwashed hair (5), shrimp paste (5)</td>
</tr>
<tr>
<td>hûm</td>
<td>perfume (120), flower (99), fabric softener (51), jasmine (32), soap (27), food (24), shampoo (23), lotion (16), powder (15), rose (12), baby powder (11), dessert (7), snack (6), detergent (5)</td>
</tr>
</tbody>
</table>
Term | ExEMPLARS
---|---
họm chuy | perfume (49), food (32), freshly cooked food (20), fabric softener (14), celery (8), flower (8), flower of suicide tree (*Cerbera odollam*) (7), holy basil stir fry (7), air freshener (7), holy basil (7), strong perfume (6), bread (6), coriander (6), suicide tree (6), clear soup (5), spicy stir fry (5), spices (5), soap (5), omelet (5)
họm huan | flower (65), perfume (47), jasmine (29), food (22), fabric softener (18), rose (13), traditional Thai perfume (10), orange jasmine (9), champak (*Magnolia champaca*) (9), plumeria flower (9), white champaka (*Magnolia x alba*) (8), dessert (6), ylang-ylang (5), pandan (5), aroma candle (5)
họm gun | coffee (74), bread (61), food (32), freshly baked bread (20), freshly cooked food (17), hot coffee (15), toasted bread (14), rice (13), tea (12), ovaltine (12), toasted snack (10), cake (9), freshly cooked rice (8), bakery (8), cooked rice (6), hot milk (6), flower (5), porridge (5), hot bread (5), cookie (5)
họm wian | dessert (69), cake (45), snack (25), candy (23), perfume (22), soft drink (19), honey (16), chocolate (14), Thai dessert (13), vanilla (12), ice cream (10), flower (10), fruit (9), coconut milk (8), milk (7), caramel (6), condensed milk (6), jasmine (6), cookie (5), *Uvaria siamensis* (5), banana (5), fabric softener (5), bua loi (glutinous rice balls in sweet coconut cream) (5), lipstick (5)

Participants typically generated one or two exemplars per term (M = 1.5). Each term received on average 133 exemplars, ranging from 79 exemplars for měn biuut up to 168 exemplars for měn khiaaw. The strongest agreement as to the most salient exemplars was perfume for họm, fish for khaaw, and body for měn priaaw. In cases where agreement was less pronounced, there was nevertheless one or more consensual exemplars listed. In addition, as expected, měn-headed compounds such as měn khaaw were associated with similar exemplars as the simplex unpleasant terms such as khaaw, though some subtle differences were present. A detailed analysis of the exemplars across the terms is provided in Section 6.

5 Distributional characteristics of smell terms

As mentioned briefly in Section 2, references to smell are relatively common in Thai discourse. Aside from the general term glin ‘smell’, a few other terms also feature among the 5,000 most frequent words in the Thai National Corpus: for example, họm ‘to be fragrant’ (rank 1,083), měn ‘to be stinky’ (rank 3,442), and dom ‘to sniff’ (rank 3,679). Although the high frequency of họm and měn is partly accounted for by compounds in which they feature, e.g., hua họm ‘onion’ (head be.fragrant), nāam họm ‘perfume’ (water be.fragrant), the overall frequency of Thai smell-related terms is still high in comparison to English, in which other smell-related terms like stink, stinky, or musty do not even make it to the top 5,000 list (with the exception of the verb to smell, rank 2,412) (cf. San Roque et al. 2015; Winter 2016; Floyd et al. 2018).

Below, we examine each term in more detail, consolidating the insights from the exemplar listing task with examples drawn from the Thai National Corpus to illustrate the use of smell terms. Related items with shared forms are discussed close to each other. We first present the term chún, followed by měn and all of the terms which can form compounds with měn, and finally end with the term họm together with all họm compounds.

5.1 Chún

Chún is linked to pungent, overpowering smells that may be experienced as too strong. As such, it can refer to smells that with normal intensity would be considered as pleasant, e.g., perfume, as well as various strong, potentially irritating smells, e.g., garlic and onion. The word chún originates from Khmer (RID: Royal Institute Dictionary, 2003). A typical context of use is provided in (10).

(10) lè mii glin phrik-thay chún ját
and exist smell black.pepper be.strong-smelling extremely
‘And there is an extremely strong smell of black pepper.’
(NWCOL179)

Chùn also has a metaphorical meaning referring to a feeling of anger, and similar to the English term *irritated*, combines the literal sense of physical irritation by odors, as well as irritation in the sense of an emotional state. The abstract sense of anger appears to be a metaphorical extension of the more concrete sense of physical irritation and constitutes a rarely reported case of a conceptualization of anger in terms of an odor-evoked sensation (cf. O’Meara and Majid, submitted).

5.2 Mën

Compared to most other terms, mën has a relatively general meaning of an unpleasant smell. It elicited a wide variety of exemplars (see Table 2), which include bodily, environmental, industrial, as well as food smells. As mentioned above, in addition to being attested on its own, mën forms smell compounds with other terms, e.g., mën khaaw, mën āp, etc., where it explicitly marks the term as unpleasant. The most frequently named exemplars of mën include excrement, garbage, fart, and sewage (see (11)).

Aside from the more typical stative use, mën occurs in a dynamic frame with the experiencer as subject and the smelled item as direct object, as in (12).

(11) raw long ruā gan riāapróoy, thon mën
1PL descend boat together all.set bear be.stinky

smell remnants trash in boat throughout time
‘We all got onto the boat and put up with the stinky smell of trash remnants in the boat all through.’
(NACHM066)

(12) ngān kō glāp, naang sawāat eeng kō mën
so then return Mrs. Sawaat herself also experience.stink

smell that until bear not can
‘So, let’s go back. That smell is also so stinky to Mrs. Sawaat that she cannot stand it.’
(PRNV022)

Finally, mën has a metaphorical sense too. When combined with biūa ‘to be bored’ in a compound mën biūa, it marks an emotional state of being fed up with something or someone (e.g., mën biūa nāk-gaan-muāng ‘fed up with politicians’).

5.3 Mën khīaw / mën prīaw / mën biūut

Mën khaaw, mën prīaw, and mën biūut are three compounds representing the ‘mën + other term’ category. Mën khīaw (khīaw ‘to be green’) is prototypically linked with the smell of vegetables (e.g., Chinese kale, lettuce, long beans), as well as leaves, grass, the human body, sweat, and sewage. Mën prīaw (prīaw ‘to be sour’) is associated most strongly with body odors, the odor of armpits, sweat, and spoiled foods. It is similar to mën biūut (biūut ‘to be spoiled’), which is used primarily with reference to the smell of spoiled and rotten foods, e.g., milk and rice, but mën prīaw is said to be less strong and, according to some speakers, is a milder quality that arises before an object becomes mën biūut. Example (13) illustrates the use of mën biūut in context.

(13) siup būrii phlaw-phlaw nā khā, mây ngān
smoke cigarettes moderately PART PART not so
Khawa and mën khaaw are usually glossed as ‘fishy’, and indeed, fish is by far the most prototypical exemplar of these terms, mentioned more than twice as often as the second most frequent exemplar – blood. In addition, khaaw and mën khaaw are used to refer to the smell of seafood and raw meat. Example (14) below illustrates a standard use of khaaw. The combination of fish and blood odors lexicalized in a single odor term has been reported for a number of languages within Southeast Asia, e.g., pale in Maniq (Wnuk and Majid 2014; Wnuk 2016), pðih/piʔey in Jahai (Burenhult and Majid 2011), anglis in Amis (Lee 2014), laŋtəs in Kavalan (Lee 2010), and elsewhere, e.g., in the Amazon – wiya in Yora/Yaminahua (Shepard Jr. 1999).

Example (14)

bay khàa ɜ̃n hò plaa ping hây glin hɔ̌m
leag galangal be.soft wrap fish grill give smell be.fragrant
dàp glin khaaw dày dii
suppress smell be.fishy can well
‘Wrapping grilled fish in a soft galangal leaf can give it a fragrant odor and suppress the fishy smell well.’

(NWCOL119)

Khawa is also used in an expression khɔɔŋ khaaw to describe savory food, opposite of khɔɔŋ wǎan ‘sweets, desserts’. In addition, khaaw has a metaphorical sense and can be used to mean ‘a tainted or degenerated reputation’, as in raakhii khaaw ‘sexual stigma’.

5.5 Hûun / mën hûun

Hûun and mën hûun are most strongly associated with the smell of oil, typically oil used for frying, especially when the oil is not fresh. The terms are also related to the smell of other fats, e.g., lard, as well as fried food. Since hûun and mën hûun carry the implication that the smells are not desirable, the terms are frequently used in admonitions and instructions to prevent rancid odors from coming about, as in (15).

Example (15)

thàa thɔst thîng wày, mây khuan go̰n sɔɔŋ wàn
if deep-fry discard put away not should exceed two day
phró jà mîi glin mën hûun,
because will have smell be.stinky be.rancid
mây nàa-gin
not appetizing
‘If you deep-fry and put (food) away, it shouldn’t be for longer than two days because it will have a rancid stinky smell and won’t be appetizing.’

(NACMD075)

5.6 Àp / mën àp / àp chúun / mën àp chúun

Àp and mën àp are probably best glossed with the English term ‘musty’. The most commonly listed exemplars for both àp and mën àp include clothes, shoes and closed small spaces, e.g.,

pàak hêng mën buut
mouth be.dry be.stinky be.spoiled
‘You shouldn’t smoke too much, so your mouth isn’t dry and stinky.’

(NACMD085)
unventilated rooms and wardrobes. When occurring on its own, ăp is not necessarily interpreted exclusively as an odor descriptor, as it can also simply mean ‘to be stuffy’. The odor sense is sometimes indicated explicitly by occurrence with glin ‘smell’ (cf. Example (16)) or within the compound mën ăp, but this is not required in order to get an olfactory interpretation. In addition to its abstract senses, ăp is also a noun denoting a concrete object—‘a small flat container with a cover, e.g. for face powder, medicated wax, dentifrice in cake or powdered form, etc.’ (Haas 1964), which could be a source that the odor term ultimately derives from, especially since there is a semantic connection between the source (closed container) and the odor type (musty, stuffy).

Ăp also features in a variety of non-smell compounds, e.g., tôk ăp ‘to fall, sink into poverty’ (fall be.stuffy/musty), ăp pan-yaa ‘to run out of ideas’ (be.stuffy/musty wisdom), but these metaphorical extensions link to its container sense, not the olfactory sense of ăp.

(16) ,chúay yâp-yâng gaan gọt rai fun, békhtiria, lê chúa-raa
help prevent NMLZ arise mite dust bacteria and fungus
talɔt-jon glin ăp thii máy phung-phasisɔŋ
as.well.as smell be.musty which not be.desirable
‘It helps to prevent dust mites, bacteria, fungus, as well as the musty smell which is not undesirable.’
(NWCOL118)

Âp chûnn and mën ăp chûnn (chûnn ‘to be moist, damp’) are similar to ăp and mën ăp, but are more specific and focus primarily on odors arising in the presence of excessive dampness, such as the smell of undried clothes, restroom, and shoes. The sentence in (17) provides an example of use.

(17) phró khâng-nay khɔng-khâang ăp, mii
because inside somewhat be.stuffy exist
glin ăp chûnn chuan ụt-àt máy nóoy
smell be.musty be.moist induce be.uncomfortable not little
‘Because it was a bit stuffy inside, there was a musty damp smell which made her feel a bit uncomfortable.’
(PRNV002)

5.7 Sâap / mën sâap

Sâap and mën sâap are terms denoting unpleasant smells associated with certain animals, e.g., cockroaches, dogs, and water buffalo, as well as carcasses/corpse, human body, and mud. Cockroaches – listed most often in the task – are by definition sâap-smelling since the Thai term for ‘cockroach’ mâlɛŋ sâap means literally a sâap-smelling bug. The sentence in (18) contains further examples of animals associated with this odor quality.

(18) núa phi, gè, lê tua jaamarii pen tón, núa
meat goat sheep and CLF yak is example meat
sât tâang-tâang thii wāa maa nii lúan mii glin sâap
animal various which say come this all have smell stink
‘For example, meat of goat, sheep, and yak. The various animal meats that I just named have an unpleasant rank smell.’
(BIOO24)

5.8 Tù-tù / mën tù-tù
Tù-(tà) (also tù-tù; Haas 1964) is a bound adjective, or as Haas states, a restricted modifier (cf. Haas 1946), limited to specific contexts. It is not a fully independent term, and functions primarily as a modifier of mên in mên tù-tù, but can also occur in a phrase glín tù-ttà. The terms tù-tù and mên tù-tù denote a faintly/slightly unpleasant smell and show considerable overlap in exemplars with mên (see Section 6). According to the exemplar listing task, the most salient exemplars of these terms include farts, socks, shoes, body odors and fermented fish. Tù-tù is also common in metaphorical use to convey a sense of suspiciousness, e.g., thú-dìn sò-pò-go glín tù-tù ‘land from ALRO (Agricultural Land Reform Office) is suspicious’ (Source: https://www.posttoday.com/columnist/479847; cf. “smell” in English and Basque, Ibarretxe-Antuñano 1999).

5.9 Hổm

Hổm, just like mên, has a general meaning and is used to refer to a large variety of pleasant odors. A typical use is illustrated in (19). It is frequent in discourse, and highly productive in derivation, participating in numerous compounds denoting various object names, e.g., hũa hổm ‘onion’ (head be.fragrant), náam hổm ‘perfume’ (water be.fragrant), náam man hổm ráhdoi ‘essential oils’ (water be.oily be.fragrant vaporize), glàay hổm ‘Cavendish banana’ (banana be.fragrant), tòn hổm ‘green onions’ (stalk be.fragrant), etc. The term is also used in metaphorical expressions such as núa hổm ‘in demand, popular’ (meat be.fragrant).

Similar to mên, hổm is sometimes used in a dynamic frame with the experiencer subject. In addition, hổm combined with gěm ‘cheek’ denotes the activity of sniff-kissing, a cultural practice for expression of affection widespread in South and Southeast Asia (Hopkins 1907; Schapper 2017), cf. (20).

(19) khâu hũck vàa khâuw khâu hổm eeng
3SG say that rice 3SG be.fragrant by.itself
‘(S)he says his/her rice is fragrant by itself (without adding anything).’
(PRNV024)

(20) phũo jà gòöt luuk, hổm gěm
father would hug child sniff-kiss cheek
‘The father used to hug the child and sniff-kiss (him/her).’
(BIO031)

5.10 Hổm grùn

Hổm grùn (grùn ‘to be smoldering’) is used to describe smells of warm foods and drinks, most prototypically coffee and freshly baked bread, as in (21).

(21) raw tòp-tháay múum glaang-wan dúay gaafee glín
1PL end.up.with meal daytime with coffee smell
hổm grùn
be.fragrant be.smoldering
‘We ended the lunch with a fragrant-smelling coffee.’
(BIO023)

5.11 Hổm chũy

Hổm chũy (chũy ‘to be wafted quickly in large quantity (of smoke and odor only)’) refers to fragrant smells hanging in the air. In the exemplar listing data, it is most strongly associated with perfume, food, fabric softener and flowers (see (22)).

(22) lũak sẽn, sày gâak múu grãhiam
parboil noodles put.in residue pork garlic
khūik hōm chūy
mix be.fragrant be.wafted
‘Parboil noodles, put in pork cracklings and garlic, and mix so it’s fragrant.’

5.12 Hōm hūan
Hōm hūan (hūan ‘to go back, turn back’) is a term most strongly linked to the smell of flowers, perfumes, jasmine, and food, as in (23). While jasmine and flowers were also listed for several other terms, hōm hūan stood out in eliciting the largest number of specific types of flowers/flowering plants, e.g., orange jasmine, champak (Magnolia champaca), plumeria, white champaka (Magnolia x alba), and ylang-ylang, suggesting the term has a particularly strong association with the floral scents. In literary use, it can metaphorically refer to other kinds of pleasant experiences recollected from the past, e.g., sāmphāt nān glāp hōm hūan ‘the touch was gentle/sweet’ (PRNV111).

(23) nōng-jin dōenthhaang phān ráan wāphōm thī sōng glīn
Nong-Jin travel pass shop waffle which send smell
hōm hūan tē jāmūuk krāng-lēw-kràng-lāw
be.fragrant go.back kick nose again and again
‘Nong Jin went by a waffle shop which was emitting a recurring fragrant smell kicking the nose again and again.’

5.13 Hōm wāan
Hōm wāan (wāan ‘to be sweet (to the taste)’) is associated with pleasantly smelling sweet food, e.g., desserts, cakes, snacks, candy, etc., as well as non-food objects, e.g., flowers (see (24). The term also has a metaphorical sense and is used to refer to pleasant and satisfying events or feelings, as in adīt an hōm wāan ‘good old days’ (past that be.fragrant be.sweet) or itsārā an hōm wāan ‘sweet freedom’ (freedom that be.fragrant be.sweet). Note that in this sense, hōm wāan appears to function as a coordinate compound, i.e., ‘fragrant and sweet’, in contrast to its smell-related sense exemplified in (24), whereby wāan is a modifier of hōm and the compound is better understood as meaning ‘sweet-smelling, fragrant in a sweet way’.

(24) dōk gēev an mīi glīn hōm wāan
flower orange.jasmine that have smell be.fragrant be.sweet
‘The orange jasmine flower that has a sweet fragrant smell’

(ACHM001)

6 The structure of the Thai odor lexicon.
Following a descriptive analysis of meaning extensions, we set out to investigate how these smell terms were related to one another. To do so, we used a novel approach to shed light on the internal structure of the lexicon by quantitatively depicting the similarity of the terms to each other. To gauge similarity, we used the Chao Jaccard index, an abundance-based similarity
measure commonly employed in taxonomic and ecological research for comparing species composition and biodiversity (Chao et al. 2005).1

In this case, we used the index to compare smell terms by examining how exemplars (elicited in the experiment described in Section 4) were distributed over terms. We calculated the extent to which there was overlap in exemplars between two terms relative to all exemplars provided for those terms. The index was defined by the following equation: \( J = \frac{U\cdot V}{U+V-U\cdot V} \), where U is the frequency of overlapping exemplars in the first set (smell term 1) normalized by the total number of responses in this set, and V is the frequency of overlapping exemplars in the second set (smell term 2) normalized by the total number of responses in that set (see further below). The values of the index range between 0 and 1, where 0 indicates complete lack of similarity and 1 indicates perfect similarity (full intersection) between two terms.

As a simple example, if we were to calculate the similarity of hōom and hōom hūan and assume for the moment that they had elicited only to the top 3 exemplars in Table 2, we would get the U value of 0.81 for hōom, \( U = \frac{120+99}{120+99+51} \), and the V value of 0.79 for hōom hūan, \( V = \frac{65+47}{65+47+29} \), which would yield a final similarity value of 0.67 (since two of the listed exemplars are overlapping). If, on the other hand, we compared another two terms, e.g., âp and tū-tū, again restricting the comparison to the top exemplars for illustrative purposes, the index value would be 0.10, reflecting lower similarity because there is only one overlapping exemplar.

Of course, the actual calculations were different since all exemplars were taken into account. By comparing all exemplars across all terms, we obtained a single similarity value (ranging from 0 to 1) for all pairs of terms (see Supplementary material), and thereby constructed a similarity matrix which was then used as input to a hierarchical cluster analysis. In this case, the cluster analysis was performed with the average-linkage-between-groups clustering method which does not presuppose the number of resulting clusters. Figure 1 illustrates the main groupings uncovered by this analysis. In this dendrogram, the length of branches indicate similarity (such that terms connected by the shortest branches are most similar in meaning). While branch length and the subgroup organization reveal similarity, the position is irrelevant as long as the connections remain preserved (e.g., as in a Calder or Miró hanging mobile).

The highest-level grouping divides the terms into two large categories: terms denoting pleasant smells with the word hōom, and a much larger group of terms denoting unpleasant smells with the word mën recurring across sub-branches. Quite predictably, perhaps, most semantic similarity is found between pairs of items with shared forms, in particular the mën-headed compounds and their corresponding monolexemic terms, i.e., mën khaaw and khaaw, mën hūan and hūan, mën sāap and sāap, mën tū-tū and tū-tū, mën âp chūn and âp chūn, mën âp and âp, but also hōom and hōom hūan. This global configuration, along the fact that hōom and mën systematically feature in pleasant- and unpleasant-denoting compounds, reinforces the idea that hōom and mën represent generic pleasant and unpleasant terms. In addition, the fact that pleasantness is underlying the most important distinction in this lexical field is in line with previous studies of odor vocabularies (van Beek 1992; Wnuk and Majid 2014), including in languages without much lexical elaboration of smell (Viberg 1984; Winter 2016).

---

1 In ecological research, the index is used to compare pairs of sets, e.g., an assemblage of seedling species vs. assemblage of tree species in a forest, in order to establish their level of similarity.
Turning to the clusters uncovered by the analyses in more detail, the first sub-grouping within unpleasant smells includes two clusters. The ãp-cluster is clearly distinguished from the other terms and is most strongly associated with musty smells of wet clothes, shoes, and unaired spaces. The remaining broad cluster includes a variety of odor terms, all of which share two exemplars: garbage and body odor. Within this broad cluster, we can distinguish further sub-clusters. The khaaw-cluster, associated primarily with fishy and bloody odors, branches off from the garbage-body-odor group, which itself splits further into two sub-clusters: the top one with hùun, mèn hùun, chœn and mèn khíaaw, shares several exemplars among which sewage and suicide tree were the most frequent, and the bottom cluster with tù-tù, mèn tù-tù, mèn, mèn prìaw, mèn büut, sàap, and mèn sàap, linked to body (body odor, person who hasn’t showered, armpits, socks) and decay (dead animal, garbage). This body-decay cluster is further split into the sàap-cluster, primarily linked to animal smells, and clearly distinct from the larger cluster of tù-tù, mèn tù-tù, mèn, mèn prìaw, mèn büut, associated with a somewhat broader selection of body- and decay-related smells, including also spoiled and rotten food. The small cluster of tù-tù, mèn tù-tù, mèn shows high similarity, as indicated by short branches, and is most strongly linked to odors of excrement, fart and feet, while the parallel cluster of mèn prìaw and mèn büut relates primarily to spoiled food. The “unpleasant” branch includes several smaller clusters (not labeled on the figure): hùun and mèn hùun (oil), tù-tù and mèn tù-tù (fart, socks), mèn ãp chùun and ãp chùun (undried clothes, shoes), mèn ãp and ãp (clothes, shoes, wardrobe).

The “pleasant” cluster shows much less differentiation compared to the “unpleasant” cluster. Pleasant odor terms appear to share a wide range of exemplars, including food, cosmetic, and natural smells (coffee, flower, food, rose, jasmine, perfume, powder, cake, fabric.
softener, curry, and bread). It is not the case that a specific category of pleasant smells – for instance, food smells – is clearly differentiated from the rest. Rather, the pleasant smell terms appear to capture additional properties, as is also apparent from examining the meaning of the second element in the compound, e.g., coming from a hot source, hanging in the air, coming back, etc. The high internal similarity among terms denoting pleasant smells is not unexpected since many languages show less diversity in pleasant compared to unpleasant smell terminology (e.g., Lee 2010; O’Meara and Majid 2016). This could be the case for a number of reasons: in certain contexts unpleasant smells are more perceptually salient than pleasant smells, e.g., unpleasant smells have been shown to be more easily detected in low-intensity concentrations (Rabin and Cain 1989). In addition, unpleasant smells are remembered better than pleasant odors (Larsson et al. 2009). Further data are needed to verify the causes and robustness of this apparent cross-linguistic tendency.

7 Discussion

We have shown here that talk about smell in Thai is relatively common, as indicated by high incidence of smell-related vocabulary in the Thai National Corpus, and the relevance of smell across a number of cultural domains. More importantly, we illustrated the rich smell terminology of Thai, which has not been described in detail previously. We found there are at least seven monolexemic terms for referring to smells, and around a dozen lexicalized compounds. The present study is also one of the few to examine the structure of the smell lexicon. It did so in a novel way: instead of asking speakers to make explicit judgments of the similarity in meaning between smell terms (cf. Wnuk and Majid 2014; van Beek 1992), it asked speakers to list exemplars, and used the co-occurrence of exemplars over terms as an indication of the underlying semantic similarity. This approach reveals the relationships between terms, which constitute an integral part of their meaning (Fillenbaum and Rapoport 1971; Lyons 1977; Majid 2015). We have thus been able to go beyond simply uncovering the extensions of words; we have tapped into aspects of the intensional semantics by revealing the relations between terms.

Our findings largely converge with previous work in that the primary semantic dimension structuring the field in Thai was pleasantness, similar to what has been found for Maniq (Wnuk and Majid 2014). Further clustering beyond the main pleasant-unpleasant division also showed striking similarity to what has been reported in other languages. For instance, the “musty” cluster was found in both Thai and Kapsiki, and the Thai “body/decay” and “body/decay/spoiled food” clusters were remarkably similar to the Kapsiki “absolute inedibility” cluster (rotting food; feces; smith-food, i.e. foul-smelling food such as horsemeat) (van Beek 1992). These results are telling since they reveal that similar kinds of odors are grouped together in unrelated languages, suggesting these groupings might be especially salient. Data from further languages would shed further light onto this topic and help establish a typology of smell terms, consonant for that posited for color (Majid and Kruspe, 2018).

This article is a first in-depth exploration of smell vocabulary in a language spoken by a speech community of tens of millions of speakers. Although similar detailed studies in other large languages seem to be lacking, Thai does not seem to be the only such case. Smell terms might in fact be a relatively common feature within the linguistic area of Mainland Southeast Asia (MSEA) and the adjacent insular areas (Philippines, Taiwan) not only in small, but also large languages (cf. Burenhult and Majid 2011). For instance, there are dedicated reports on smell terms among SEA small-scale communities, e.g., Aslian and Formosan (Majid and Burenhult 2014; Wnuk and Majid 2014; Lee 2010), and the larger Formosan and Philippine languages (Lee 2010; Lee 2014; Blust 1988). If we inspect dictionaries of some major languages of the area, we find the number of smell-denoting terms in each of them is well above the rarum threshold of “two or three terms”, cf., e.g., Burmese (1996), Lao (Kerr 1972;
Patterson et al. 1994; see also Majid, Roberts, et al. 2018), Vietnamese (Bùi 1992), Khmer (Headley et al. 1997), Tagalog (Ramos 1971; Calderón 2007), and Malay (Wilkinson 1926; Wilkinson 1932). In Cantonese, there are at least six specific terms for odors, although De Sousa (2011) notes that their knowledge is decreasing in younger generations. These languages span at least four of the five families spoken in SEA: Sino-Tibetan, Tai-Kadai, Austroasiatic, and Austronesian, suggesting a wide distribution cross-cutting genealogical boundaries (cf. Enfield 2005; Enfield and Comrie 2015). Such wide-ranging distribution within a single diffusion area casts doubt on whether they would qualify as a case of rarum in the absolute sense (Plank 2001; Wohlgemuth and Cysouw 2011). Their presence in other areas of the world, notably in Africa and Latin America, further suggests they are also not an instance of relative rara (i.e., a phenomenon that is frequent locally, but rare globally). Further descriptive work remains to be done to get a wider coverage of the domain of smell in SEA and obtain a full typological picture for the area.

8 Conclusions

The Thai data discussed here shows smell terms are not limited to languages spoken by small-scale societies. This sheds new light onto what we know about smell lexica generally and suggests large population size alone is not incompatible with having elaborate smell terminology. In the Western context, this lends further credibility to the idea that cultural factors such as deodorization and the derogation of smell could be among the critical forces behind the lexical gap pointed out for English and other European languages (Weisgerber 1928; Classen et al. 1994; McGann 2017). While these forces may have affected Thailand to some degree as well, there are a number of factors favorable to foregrounding olfaction in daily life among Thai speakers, e.g., the rich olfactory environment of a tropical climate zone, as well as the cultural relevance of smell in the context of food, medicine, and religion (cf. Majid et al. 2017). The resulting attention smell receives is reflected in the frequent mention of smell-related words in the corpus with the word glin ‘smell’ being among the 1,000 most frequent words of Thai.

In addition to addressing these broader themes, this article provides valuable insights into our understanding of smell lexica and showcases a new methodological approach to investigate the semantics of smell terms. Most notably, we demonstrate that the traditionally employed method of exemplar listing can lead to novel insights not only with respect to extensional meaning, but may also be a window into the intensional meaning by virtue of revealing the internal structure of the lexicon. As such, this is another method that can be exploited to bring new dimensions to our understanding semantic fields, particularly of smell.

Abbreviations

1 first person
3 third person
CLF classifier
ASP aspect
NEG negative
NMLZ nominalizer
PL plural
PART particle
SG singular

Acknowledgements: This work was supported by The Netherlands Organization for Scientific Research: NWO VICI grant “Human olfaction at the intersection of language, culture and biology” 277-70-011 (PI Majid). We thank Seth Phiriya for data collection and coding, Sasinee
Khuankaew for help collecting data, Krittanon Thotsagool for critical discussions, and Yuma Ito for comments.

References


O’Meara, Carolyn, Susan Smythe Kung & Asifa Majid. in press. The challenge of olfactory ideophones: Reconsidering ineffability from the Totonac-Tepehua perspective. *International Journal of American Linguistics*.


Yeshurun, Yaara & Noam Sobel. 2010. An odor is not worth a thousand words: From multidimensional odors to unidimensional odor objects. *Annual Review of Psychology* 61. 219–2