



# Fragrance lexicon for analysis of consumer-generated perfume reviews in Russian and English



Larisa Nikitina

Faculty of Languages and Linguistics, Universiti Malaya, Kuala Lumpur 50603, Malaysia

## ARTICLE INFO

### Method name:

Method for developing a bilingual fragrance lexicon to analyze consumer-generated perfume reviews

### Keywords:

Olfaction  
Affective science  
Linguistic analysis  
Sensory linguistics  
Sensory lexis  
Online reviews  
Computer-assisted text analysis

## ABSTRACT

In recent years, there has been a rise in research on sensorium in various academic disciplines. Olfaction is recognized as a sense that is most closely linked to cognition, memory and emotion. Due to this unique feature, studies on various aspects of human olfaction are steadily gaining prominence in the humanities and social sciences. In order to understand how the olfactory modality is marked, several taxonomies and semantic spaces of olfactory terms have been developed. However, the focus has been on the general olfaction lexicon and there is a lack of systematic and comprehensive lexicons for fragrant smells. This article addresses this gap. It adopts a multilingual perspective and describes the process of developing a fragrance lexicon in two languages, Russian and English. A fragrance lexicon refers to a list of words that people might use to describe a perfume. The steps in the lexicon development included

- sourcing the lexical items in the two languages
- translating and cleaning the word lists
- revising and refining the lexicon

The fragrance lexicon presented in this article can be used to aid linguistic analyses of naturally occurring communications about perfumes, such as computational analyses of consumer-generated perfume reviews.

## Specifications table

Subject area:	Psychology
More specific subject area:	Affective science, Cognitive linguistics
Name of your method:	Method for developing a bilingual fragrance lexicon to analyze consumer-generated perfume reviews
Name and reference of original method:	N/A
Resource availability:	N/A

## Method details

It has been acknowledged that describing smells is notoriously difficult and many European languages have a limited olfactory lexicon [1]. A number of studies have been done to identify and document the olfactory lexis and produce semantic maps of the olfactory perceptions [2]. However, no attempts were made to develop a lexicon that pertains specifically to describing fragrances. This study addresses this gap. Besides its focus on fragrant smells, the main novelty of this study is a multilingual approach where the fragrance lexicon is developed in two languages: Russian and English.

E-mail address: [larisa@um.edu.my](mailto:larisa@um.edu.my)

<https://doi.org/10.1016/j.mex.2024.102627>

Received 8 June 2023; Accepted 21 February 2024

Available online 28 February 2024

2215-0161/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

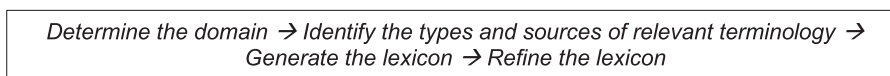


Fig. 1. Steps and stages in lexicon development.

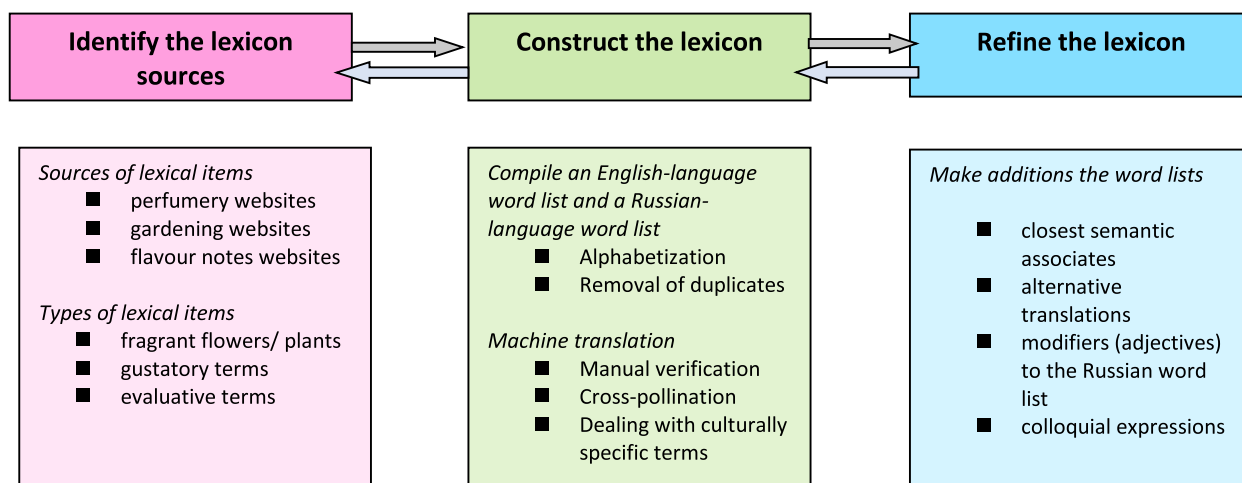


Fig. 2. Workflow for bilingual fragrance lexicon development.

Lexical resources used to describe a smell often relate to its sources (e.g., lily-of-the-valley, coffee, skunk) and the emotional impact of the smell (e.g., divine, alluring, disgusting). The sources and the affective impact of a smell are two major domains in the olfaction-related communications and exchanges. In addition, descriptive and evaluative words are employed to describe a character of a smell. The former may include such lexical items as “floral”, “woody” and “spicy” while the latter can be “feminine”, “soothing” or “sexy”, though there is no clear demarcation between these two groups of words and they often overlap [3].

A lexicon on any subject can contain hundreds of words pertaining to the domain of interest. Therefore, some guidelines and steps in the process of lexicon development could be very helpful. The steps in developing the fragrance lexicon presented in this article were adapted from Drake and Civile [4] (see Fig. 1).

It should be noted that Fig. 1 shows a general linear progression of the work flow. The process of lexicon development, however, was iterative (see Fig. 2). The lists of words obtained from various sources were repeatedly checked, verified and further refined. More details of the steps and decisions taken at each stage of the lexicon development are given in the following section.

## Steps in the bilingual fragrance lexicon development

### *Sourcing lexical items and making the word lists*

The specific domain of interest was fragrant smells, particularly fragrances. This focus influenced the search for sources and the selection of lexical items. The workflow of bilingual lexicon development is shown in Fig. 2.

To give more details, at the earliest stage of lexicon development, the decision-making logic was as follows: Firstly, to describe a fragrant smell or a perfume, people often refer to various flowers and fragrant plants (e.g., “camellia”, “cedar”, etc.). Descriptions of perfume notes more often than not include such references. Therefore, the names of various flowers and plants were included in the lexicon.

Secondly, gustatory terms such as words pertaining to food (e.g., “chocolate”, “popcorn”), drinks (e.g., “tequila”) and taste (e.g., “sour”, “sweet”) often assist people in their attempts to communicate the ineffable olfactory perceptions. It has been noted that the gustatory and olfactory terms are neurally integrated [5]. Clearly, the references to food and drinks do not refer to the gustatory qualities of a fragrance; rather, they act as supramodal lexis that aids in communicating the olfactory sensations. Considering this fact and in order to enrich the fragrance bilingual lexicon, the “coffee flavour wheel” as well as gustatory flavours and sensory word lists were consulted [6].

It is also important to determine the emotional valence of a smell or divide smells into “pleasant” and “unpleasant” categories. Therefore, lexis pertaining to affective reactions, both positive (e.g., “wonderful”) and negative (e.g., “nauseating”), was included in this fragrance lexicon. The evaluative lexis pertaining to olfaction was sourced from various English and Russian websites, such as blogs for budding writers [7], websites specifically devoted to perfumes [8] and gardening and plants [9]. However, it should be noted that the affective reaction or valence of a smell can vary between individuals. Even though the majority of people might find a fragrant smell, such as that emanating from lily-of-the-valley flowers, to be pleasant and alluring, some people might find it repulsive.

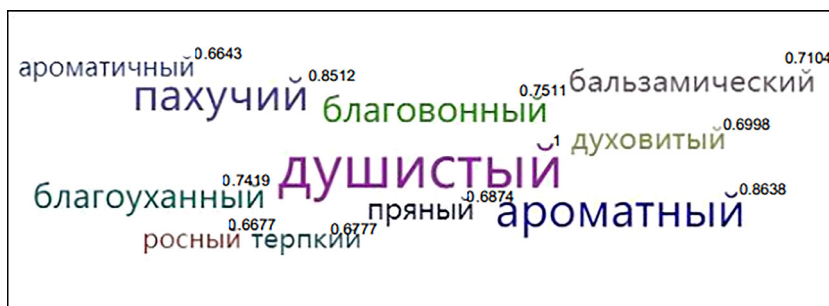


Fig. 3. Word cloud for “душистый”, Source: Russian National Corpus.

### Constructing the lexicon

All of the lexical items sourced from various websites were saved in two separate files: one file contained the Russian word list and the other file comprised the English word inventory. Each word list was then organized in alphabetical order. The alphabetization of the English language list was done with the aid of the MS Word “Sort” function. The alphabetization of the Russian words was done with the aid of the online widget “Russian Tools” [10]. Upon completing the alphabetization the duplicates in the word lists became easily noticeable and they were removed.

In the next step, the English word list was translated into Russian and, vice versa, the Russian word list was translated into English. The translation was done using Google Translate and Bing Microsoft Translator. This step allowed for cross-pollinating and enriching the fragrance lexicons in each of the two languages. Specifically, lexical items that appeared in one list of words but were missing from the other were added (with appropriate translations) to the respective list.

The machine translation of the terms was checked, verified, refined and amended manually, where needed, by the researcher, who is fluent in Russian and English. At this stage, mistakes in the translated lists of words were identified and corrected. For example, the Russian word “багряник” was translated into English as “purple”. This erroneous translation was manually replaced with the correct English equivalent, “redbuds”. This was the most laborious part of the bilingual fragrance lexicon development. This is because the most precise and adequate translation options had to be selected and retained among several alternative terms proposed by the machine translator. In cases where not one but several word options were correct and all of them were pertinent to the domain of interest, then all of the proposed options were included in the target language word list. These terms were then translated into the other language.

There were several culturally-specific terms for which no precise equivalent exists in the other language. For example, some Russian drinks (e.g., “морс”, which is a fruit drink prepared from cranberries or lingonberries) and sweets (e.g., “зефир”, which is a soft fluffy confectionary) could not be machine translated into English. Wherever possible, I substituted such lexical items with their closest equivalents in the target language. Thus, the word “зефир” was replaced in the English list with the word “meringue”. Though the smell of the two sweets is not exactly the same (in fact, they might not have any distinctive smell), nevertheless, the pleasant mental associations that arise from these pastel-coloured sweet treats can be similar.

### Refining the lexicon

At the stage of refining the translated lists of words, I consulted available online dictionaries and other resources available in Russian and English. Particularly useful for refining and enriching the Russian fragrance lexicon was the Russian National Corpus [11] and its widget “похожие слова” (“similar words”). This function generates a word cloud that contains the closest semantic associates of a specific lexical item; the proximity coefficients can be obtained by placing the mouse over a word of interest. Fig. 3 shows the word cloud for the word “душистый” (“fragrant”, “perfumed”) including each lexical item’s coefficients of semantic proximity to the main term. To give more details, the main term of interest (e.g., “душистый”) is assigned value “1”. Values closer to “1” show greater semantic proximity of a term to the main term of interest. Also, words that have greater similarity of usage with the word “душистый” are displayed in larger fonts. A more detailed explanation of the widget is available on the Russian National Corpus website [11].

Also, further refinements were made to the Russian version of the fragrance lexicon. To be more specific, firstly, I added the alternative names for some plants (e.g., “лабазник” was added besides the existing word “таволга”). Secondly, the adjectival forms were added to the noun forms. To give some examples, the modifier “трюфельный” was added to the word list besides the existing noun “трюфель” (truffle); the modifier “кофейный” was listed in addition to the noun “кофе” (coffee). This was done in order to allow for differentiating during a subsequent computer-assisted linguistic analysis whether the source of smell (e.g., expressed by the noun “кофе”) or the smell character (e.g., expressed by the adjective “кофейный”) is being referred to. Some colloquial expressions that Russian speakers might employ when talking about perfumes or smells in general were added to the Russian version of the olfactory lexicon. These included such words as “парфюманьяк” (literally “perfume maniac”), “деревяшки” (literally “pieces

of wood”), “свежак” (a slang word for a “fresh”- or citrus-smelling perfume or cologne), “химоза” and “химозный” (slang words referring to a “chemical” smell).

The finalized versions of the fragrance lexicon in Russian and English can be found in the Appendix. In total, there were 1024 words in the Russian word list and 773 words in the English version. The English version is shorter than the Russian inventory because in English, the same word form indicates a noun (e.g., hot coffee) and a modifier (e.g., the coffee aroma). In Russian, the noun “coffee” (“кофе”) has a different form from the modifier “кофейный”. Therefore, in the Russian version of the lexicon, both the noun forms and the adjectival forms are included.

## Concluding remarks

The fragrance lexicon presented in this article can be used for computer-assisted analyses of texts pertaining to fragrances and perfumes, such as perfume blogs and consumer-generated perfume reviews. A range of natural language processing techniques can be used for such analyses [12].

There are some limitations to this endeavour. One of them is that the researcher did not pursue the aim of listing all possible words that could be used to describe a smell or olfactory reaction. This would be an unsurmountable task considering the rich lexical and metaphoric means that people employ to describe their olfactory perceptions. Notwithstanding, the word lists of olfactory terms presented here contain an extensive lexicon to aid linguistic analyses of fragrant smell descriptions, including computational analyses of consumer-generated online reviews. Other limitations are that the lexicon was developed in two languages, Russian and English, and only to assess fragrant smells such as perfumes, all of which narrow the range of languages and types of odours that can be fruitfully assessed using this inventory.

Despite these limitations, it is hoped that this study has expanded the availability of research instruments for conducting linguistic analyses of olfactory perceptions, particularly of fragrant smells. The bilingual fragrance lexicon can be used in its own right or in combination with other pertinent lexicons and dictionaries, for example those available in LIWC-22 software [13]. This lexicon can be translated into other languages, with some modifications that undoubtedly will be required considering the differences in cultural norms and in linguistic means for expressing olfactory perceptions.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

No data was used for the research described in the article.

## Ethics statements

*This work involved neither human subjects nor animal experiments. It did not use data from social media platforms.*

## Acknowledgments

This study was supported by the UM International Collaboration Grant – SATU Joint Research Scheme (Grant number ST026–2023). The image used in the graphical abstract is by KamranAydinov on Freepik. I dedicate this work to my aunt Lydia, with whom we spent many happy hours talking about perfumes and other beautiful things in life.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.mex.2024.102627](https://doi.org/10.1016/j.mex.2024.102627).

## References

- [1] A. Majid, Human olfaction at the intersection of language, culture, and biology, *Trends Cogn. Sci.* 25 (2) (2021) 111–123, doi:[10.1016/j.tics.2020.11.005](https://doi.org/10.1016/j.tics.2020.11.005).
- [2] C. Chrea, D. Grandjean, S. Delplanque, I. Cayeux, B. Le Calvé, L. Aymard, M.I. Velasco, D. Sander, K.R. Scherer, Mapping the semantic space for the subjective experience of emotional responses to odors, *Chem. Senses* 34 (1) (2009) 49–62, doi:[10.1093/chemse/bjn052](https://doi.org/10.1093/chemse/bjn052).
- [3] B. Moeran, The colours of smell: perfume advertising and the senses. In: *The American Anthropological Association Annual Meeting 16–20 2011*, Montreal, Canada. [https://research-api.cbs.dk/ws/portalfiles/portal/58915747/Brian\\_Moeran\\_Advertising\\_The\\_Colours\\_of\\_Smell\\_2011.pdf](https://research-api.cbs.dk/ws/portalfiles/portal/58915747/Brian_Moeran_Advertising_The_Colours_of_Smell_2011.pdf) (accessed January 15, 2023).
- [4] M.A. Drake, G.V. Civile, Flavor lexicons, *Compr. Rev. Food Sci. Food Saf.* 2 (1) (2003) 33–40, doi:[10.1111/j.1541-4337.2003.tb00013.x](https://doi.org/10.1111/j.1541-4337.2003.tb00013.x).
- [5] B. Winter, Taste and smell words form an affectively loaded and emotionally flexible part of the English lexicon, *Lang. Cogn. Neurosci.* 31 (8) (2016) 975–988, doi:[10.1080/23273798.2016.1193619](https://doi.org/10.1080/23273798.2016.1193619).
- [6] World Coffee Research, Sensory Lexicon 2017. <https://worldcoffeeresearch.org/resources/sensory-lexicon> (accessed April 20, 2023).
- [7] M.G. White, Descriptive words for scents: list of smell adjectives. <https://grammar.yourdictionary.com/style-and-usage/descriptive-words-for-scents.html> (accessed April 20, 2023).
- [8] *Fragrantica*, (2023). <https://www.fragrantica.com/> (accessed April 8, 2023).

- [9] Fermilon.ru, (2023). <https://fermilon.ru> (accessed April 8, 2023).
- [10] Russian Tools, (2023). <https://www.russiantools.com/en> (accessed April 20, 2023).
- [11] Russian National Corpus, (2024). Word cloud of the term “душистый”. <https://ruscorpora.ru/word/main?req=%D0%B4%D1%83%D1%88%D0%B8%D1%81%D1%82%D1%8B%D0%B9&gr=A> (accessed February 9, 2024).
- [12] L. Nikitina, F. Furuoka, F. Words of scents: a linguistic analysis of online perfume reviews, *Linguist. Vanguard* 9 (1) (2023) 205–214, doi:[10.1515/lingvan-2022-0135](https://doi.org/10.1515/lingvan-2022-0135).
- [13] J.W. Pennebaker, R.L. Boyd, R.J. Booth, A. Ashokkumar, M.E. Francis, *Linguistic Inquiry and Word Count LIWC-22*, Pennebaker Conglomerates (2022) <https://www.liwc.app>.