Mapping the Language of Spices

A corpus-based, philological study on the words of the spice domain

Gábor Parti

Doctor of Philosophy

The Hong Kong Polytechnic University

— Initial Submission for Examination Purpose —

2022

⚓ Hong Kong

The Hong Kong Polytechnic University

Department of Chinese and Bilingual Studies

Mapping the Language of Spices

A corpus-based, philological study on the words of the spice domain

by

Gábor Parti

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

— Initial Submission for Examination Purpose —

August, 2022

Certificate of originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

_____ (Signed)

_____
Gábor Parti   (Name of Student)

Abstract

The majority of existing literature on spices is found in the areas of gastronomy, botany, and history. This study investigates spices on a linguistic level and aims to be a comprehensive linguistic account on the items of the spice trade. Some of these dried plant matter were highly desired at certain points in history, due to their attractive aroma and medicinal value, thus they were ideal products of trade early on. Cultural contact and exchange, and the introduction of new cultural items begets situations of language contact and linguistic acculturation, and so in the case of spices, we not only have a set of items that traveled around the world, but also a set of names. This domain is very rich in loanwords and *Wanderwörter*, but also supplies us with a myriad of cases where spice names are conventional innovations. To make it more interesting, the thesis compares English, Chinese, and Arabic, languages that represent major powers in the spice trade at different times. The thesis has two main parts. Part one identifies the spices under scope with a brief discussion on their botany and history, followed by a presentation of the geographic and linguistic diffusion of spices and their names. Basically, we track and explain word origins and subsequent spread by tracing the materials and the propagation of the accompanying Wanderwort. This part relies on philological literature, and tools from historical linguistics, such as etymological research. Part two examines the language of spices, the terminology and nomenclature related to the spice domain from linguistic-cognitive perspectives. On one hand, it is a systematic investigation on how humans name spices: what are the mechanism and motivations behind the naming principles, and how this relates to the salient sensory features of the products (strong gustatory, olfactory, or visual stimuli). On the other hand, it looks at to what degree spice terminology is used in daily language; which is proposed to be a gauge of a spice's embeddedness in a culture. This part relies on corpora and corpus linguistic tools. Conclusions are made on the connections between the physical properties of the spices, their patterns of diffusion, and the prototypical spices and their effect of naming principles. Besides being a novel and original approach to research and categorize spices from a linguistic point of view, this study offers new insights to our knowledge about (wandering) loanwords, and the effect of the highly sensory nature of spices in the naming process when adopted by a community. It is also intended to be a useful working database for future research, and aims to dispel some of the chaos and confusion surrounding spice names.

Acknowledgments

Contents

List of Figures

List of Tables

Glossary

---

[1]Merriam-Webster, n.d.

Acronyms

Symbols and Notation

*	reconstructed form
<	developed from
>	developed into
<?	uncertain development
†	obsolete
a.	*ante*, attested before the year
ca.	*circa*, around the year/century
fragrance	italic: lexical item, a word or phrase
[fragrance]	square brackets: gloss, literal meaning
'fragrance'	single quotation marks: meaning, sense
FRAGRANCE	small capitals: a concept

The Diffusion of Spices

1

I ɴ this chapter, I will present the findings on the diffusion of spices, by looking at the distribution of spice plants and their names. First, an overview about the spices geographical distribution will be presented, then, a discussion on their spread and *spreadability* will ensue. Lastly, I will present my findings on the diffusion of spice names, and how they relate to the botanical reality. The aim of this chapter is to have an understanding of how spices spread around the globe as informed by their names and etymologies, but at the same time supported by the evidences of physical diffusion.

1.1   The Geographic Distribution of Spices

In general, it is true that spices come from the hot and humid tropical regions, especially Asia. However, there are number of aromatic plants that originate from more temperate regions, here we should think about the umbelliferous plants of the West and Central Asia: asafoetida, fennel, cumin and caraway, and others, and we must not forget the three American spices: chile, vanilla, and allspice. Figure 1.1 shows the macroareas where the 24 spices originate.

Botanical databases, such as *Plants of The World Online* (POWO), often show distribution and give us the regions where a plant is *native* to, and where it has been *introduced*. "Introduced" means that the plant is not native in the area, but now grows wild due to human intervention—whether the plant escaped cultivation, or became naturalized after accidental introduction—or due

Figure 1.1   The distribution of spice plants by the macroarea of their native habitat.

to natural spreading. Looking at this information reflects on the plants' ability to adapt and grow in new places, but also hints on how human usage and transmission affected habitats. I have collected this information and used it to compare the spices in question. I have simply counted the native and introduced regions, and added them up. In figure 1.2, you can see the spices ranked by the total number of the regions they grow in, including both native and where the plants were consequently introduced. I would like to highlight that the highest ranks are occupied by aromatic plants that are also herbs, both in the botanical and in the culinary definition. This makes sense, since these plants—e.g., fennel, coriander, dill, fenugreek, etc.—are not only cultivated for their seeds, but the leafy green parts are made use of as well, so it is without question that the whole plant "travels" to new places, not only its product. People transplant their ingredients whenever they can, unless the primary goal of cultivation is purely profit.[1]

---

[1]The Dutch for example actively destroyed plant habitats, and wiped out whole islands—including the population—in
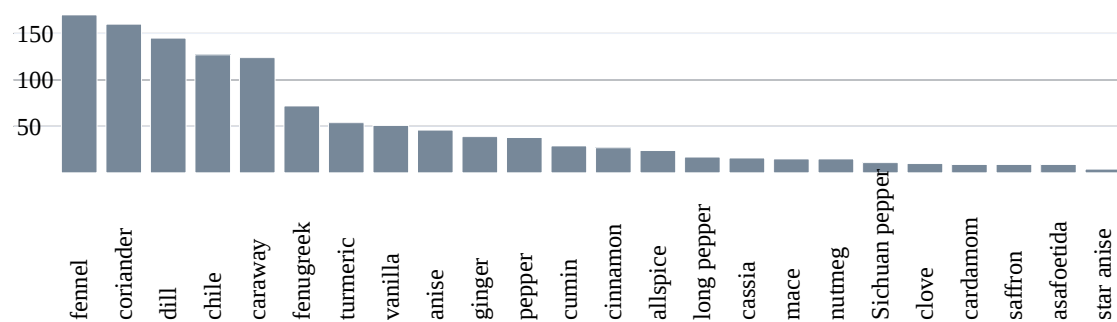
(a) Total number of growing regions.
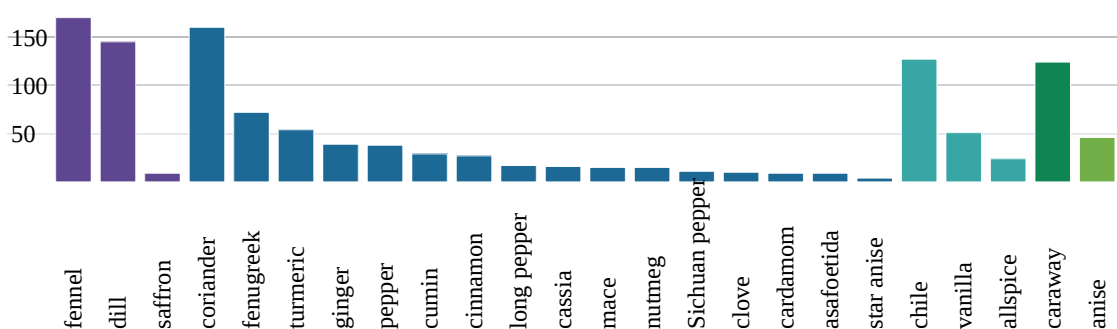
(b) Spices by total number of growing regions, grouped by macroarea.
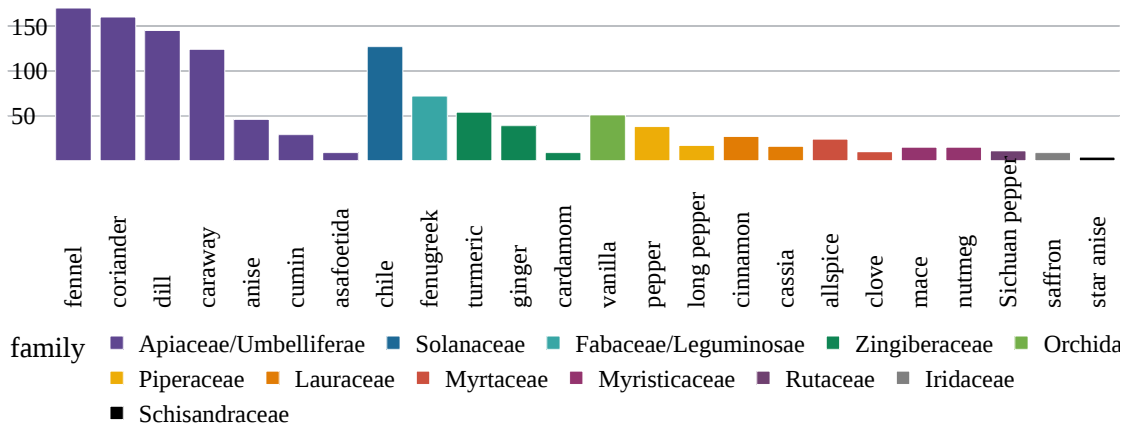
(c) Spices by total number of growing regions, grouped by family.

Figure 1.2   Spices ranked according to the total number of regions they grow in, native or introduced.

The far side of the ranking also shows the spices that do not grow extensively across many regions, regardless of how valuable or popular they are. Of course, behind this, are the complex issues of plant biology, ecology and the many factors that decide a plant's resistance to transplantation and if it can grow in new, alien environments. However, there is another point to notice here: labor. The lower ranks feature spices that are highly labor intensive to cultivate and harvest, including star anise,

the Spice Islands of Indonesia to generate scarcity and ramp up value during their monopol rule in the 17th century.

cardamom, and saffron, but the collection of asafoetida is cumbersome as well, and this also effects prices. Interestingly—and of course, closely related what was just said—all of these are products that are very specific plant parts, the pericarps (star anise, Sichuan pepper), dried oleo-gum-resin, (asafoetida), stigmas (saffron), and dried flower buds (cloves). Figure 1.2 also shows the same data, but grouped by macroarea and by plant family as well.
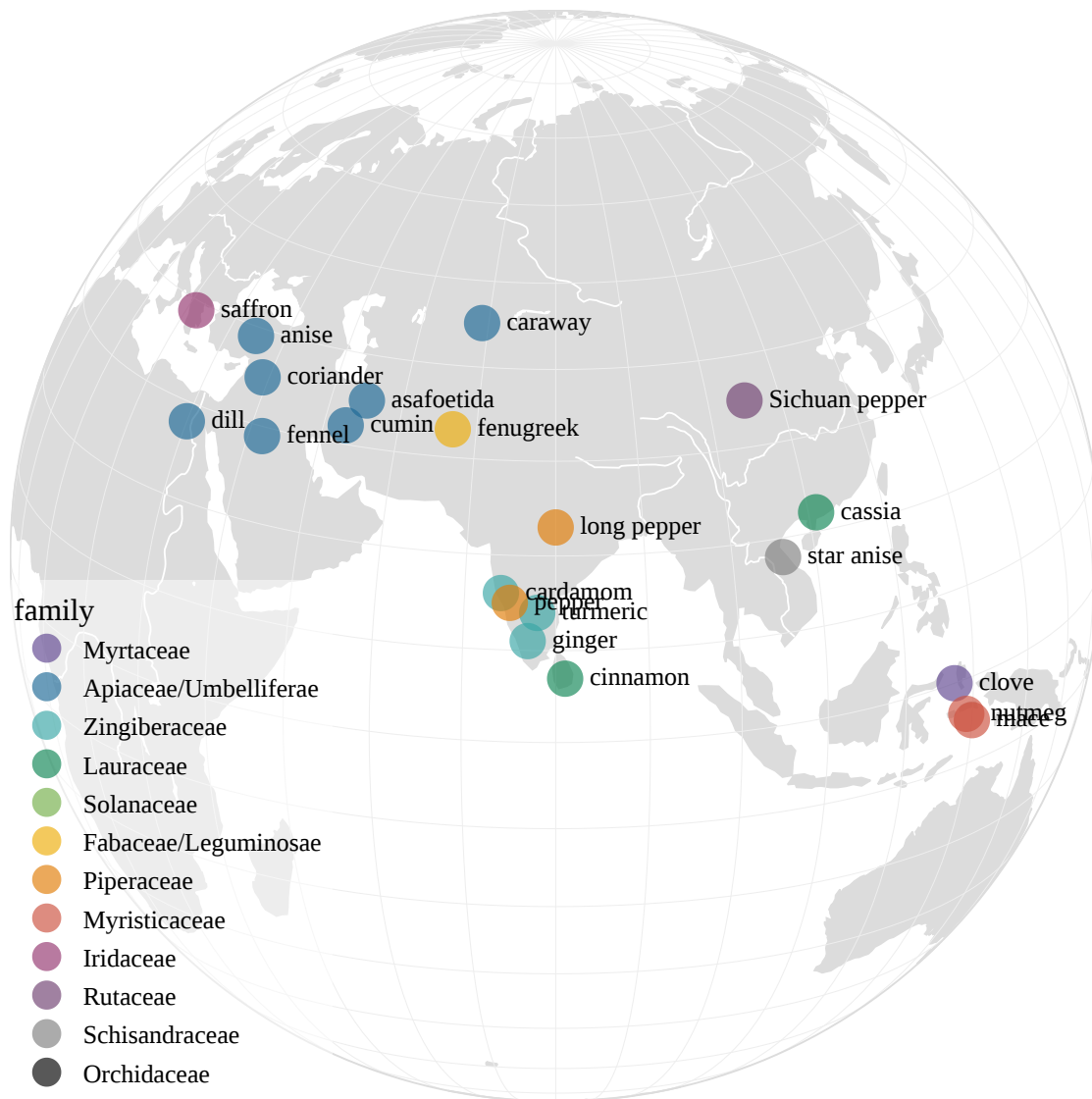
Figure 1.3   The approximate geographical origins of the spices in this thesis.

1.2   The Spreadability of Spices

When it comes to spices of commerce, there is a factor that greatly weighs in on their diffusion: I will call this *spreadability*. I wanted to have a basic understanding of what effect spices' ease or difficulty to spread can have on their diffusion, value, and global popularity, so I created a rudimentary metric based on geographical-botanical data. I have simply divided the sum of the introduced regions with
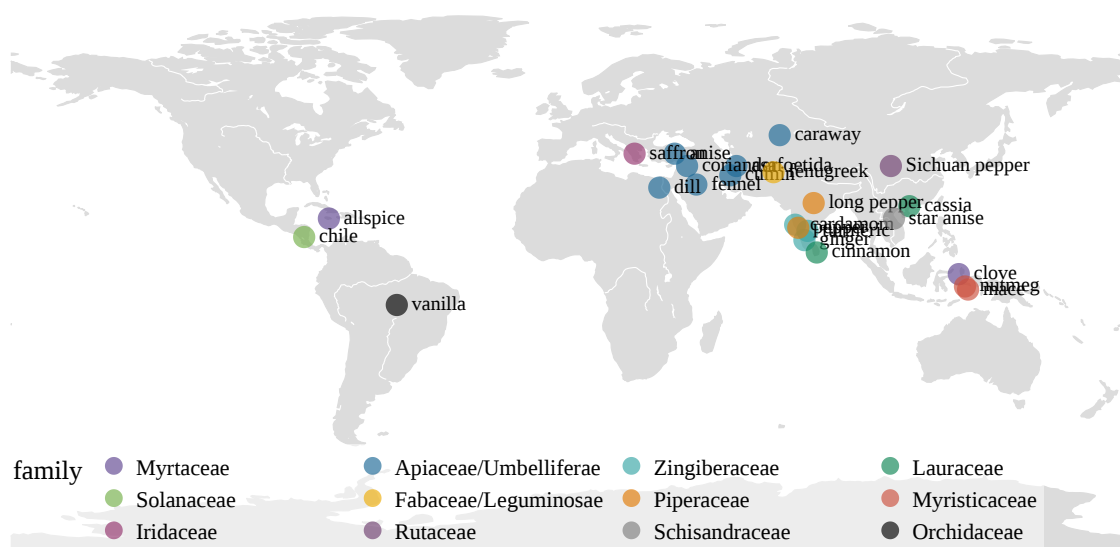
Figure 1.4   The approximate geographical origins of the spices in this thesis.

the sum of the native regions to serve as a crude indicator of how "well" a spice plant have spread. Intuitively, this index is about spice plants' ability and "ecological willingness" to spread to new regions, whether it is a result of human hands (by trade and transplantation) or nature (self-seeding, spread by birds, etc.) into neighboring areas.

$$\frac{\sum region_{introduced}}{\sum region_{native}} = spreadability\ index$$

This metric accounts for the initial difference between if a spice was minimally distributed (i.e. only found in one or two regions), or well distributed before being introduced to either a few, or many new places. Figure 1.5 shows the spices ranked by their spreadability index. The figure shows for example tumeric, originally from "one region" (India), is now found in 53 other regions, resulting in the highest score of 53. On the far side of the plot, we can find Sichuan pepper, whose main source, *Zanthoxylum bungeanum* is indigenous to 10 geographical zones in China, but only have been introduced to one region (Uzbekistan), getting a low score of 0.10.

The results of this plot—like any other—greatly depend on the data we feed to it, and like any other quantitative analysis, has its limitations. Although the regions in the POWO database are uniform, they are not clear-cut ecological zones, but rather based on administrative divisions of countries, and it is not perfect. While some large countries are divided to broad areas that represent different biodiversity zones, the borders are arbitrary. For example, the United States, Russia, and China are divided by states, provinces, or greater geographical areas (e.g., New South Wales, Central European Russia, China South-Central) India is just one unit, explaining the very high score of turmeric.[2] Nonetheless, in terms of general usefulness the index has some merit. If we look at the distribution

---

[2]Another limitation might be the age of this database as we find zones named Yugoslavia, or Czechoslovakia, but I doubt the biodiversity changed as much as political borders.

map of turmeric,[3] we will see that it did indeed spread far and wide, from Southeast Asia through West Africa to the Caribbean, and compared with Sichuan pepper[4]—which is still mostly limited to China—is much more well known globally.

(a) Spices ranked by spreadability.

macroarea ■ Asia ■ America ■ Med.; W. Asia ■ E. Med. ■ Eurasia

(b) Spices ranked by spreadability, grouped by macroarea.

family ■ Zingiberaceae ■ Piperaceae ■ Lauraceae ■ Solanaceae ■ Fabaceae/Leguminosae
■ Apiaceae/Umbelliferae ■ Myristicaceae ■ Myrtaceae ■ Iridaceae ■ Orchidaceae
■ Schisandraceae ■ Rutaceae

(c) Spices ranked by spreadability, grouped by family.

Figure 1.5 Spices ranked by their spreadability index, showing which spice plants spread to more regions, taking into account the initial state of their distribution. ($\frac{\sum region_{introduced}}{\sum region_{native}} = spreadability$)

---

[3]*Curcuma longa* on POWO: https://powo.science.kew.org/taxon/796451-1#distribution-map
[4]*Zanthoxylum bungeanum* on POWO: https://powo.science.kew.org/taxon/775625-1/#distribution-map

Figure 1.5 (b) and (c) show the spices ranked by their spreadablity index as well, broken down by macroarea and plant family. I have included the plant family groupings because it can be very interesting to those with affinity to the plant sciences, but truthfully this would be much more exciting when including more plants in these analyses.

1.3 The Linguistic Diffusion of Spices

Turning towards the linguistic side of In this section I will illustrate the linguistic diffusion of spices, through the investigation of spice names and their spread on spatial and temporal dimensions. In order to present these results in a convenient, reader friendly way, I will use geospatial mapping. The plots seen in this section are made by using the etymological data on spice terminology, collected and introduced in **??**.

Figure 1.6

anisun
hiltit
anjudan
karawiya
hal
qaqulla
darsini
qaranful
kuzbura
kammun
shibitt
shamar
zanjabil
darfilfil
basbas
jawz al-tib
fulful
zafaran
kurkum
faniliya
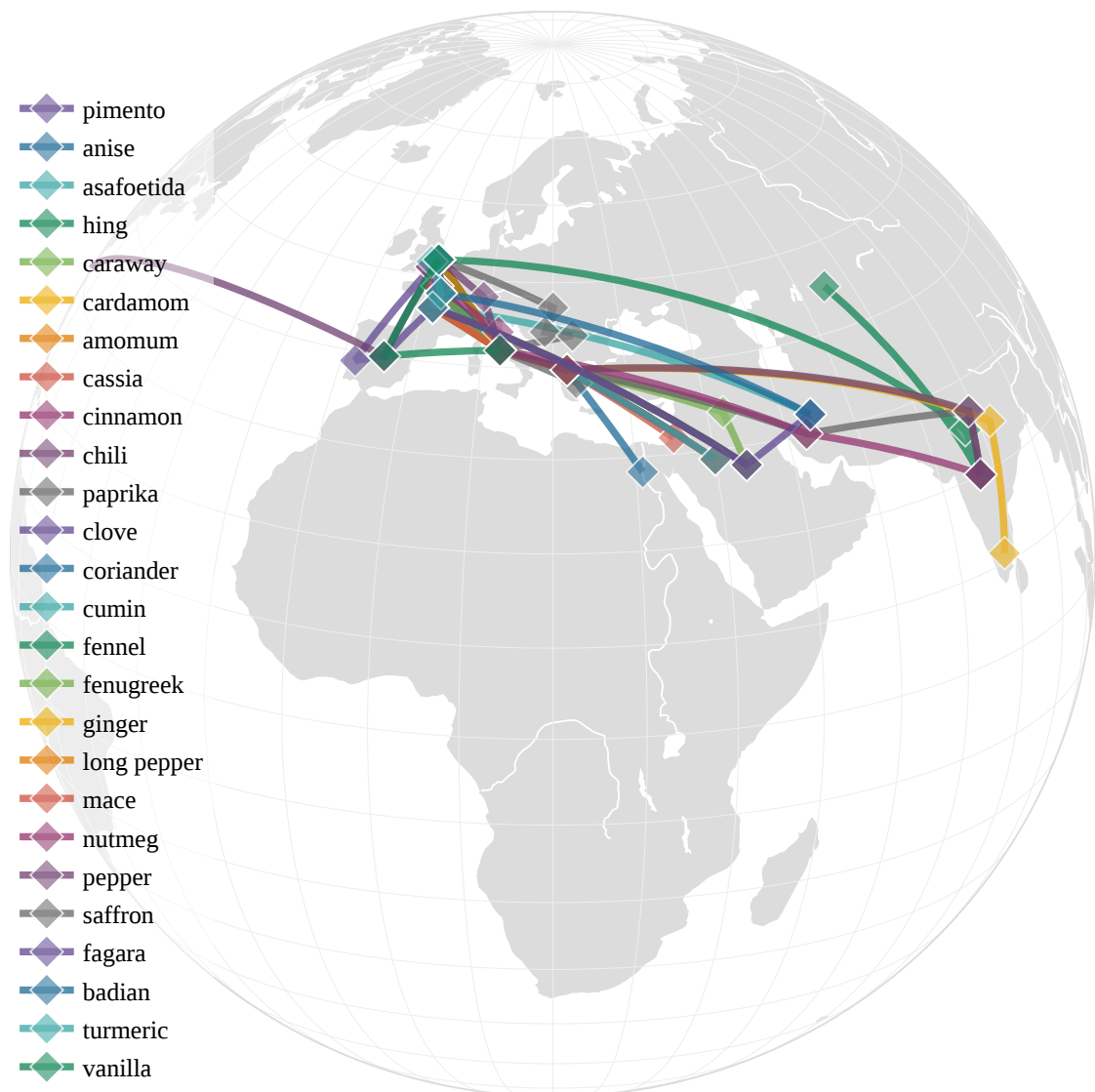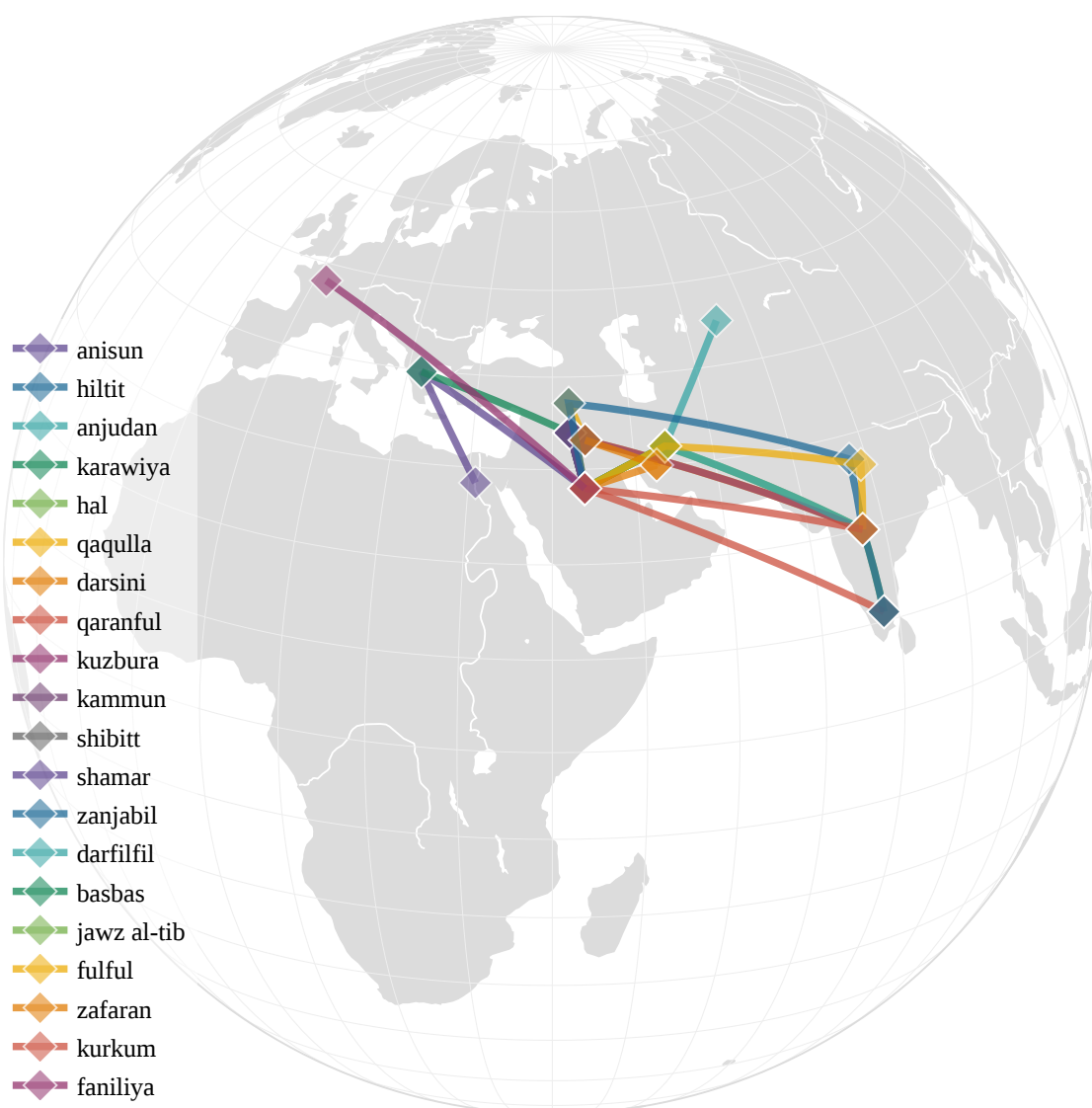
Figure 1.7

1.4 The Case of Cinnamon

1.5 One

This chapter aims to give an overview on the terminology used by various languages when referring to cinnamon. These words are connected to the spread of material culture, and a (not-so) specific plant product used and coveted for its aroma, used as spice and medicine. Known by humans for millennia, cinnamon is now present essentially on a global scale, and by exploring its names in multiple languages, we can reconstruct its linguistic genealogy. These results also tell a story; they tell us an account on the linguistic history of *cinnamonic* words, their origins, diffusion, and ultimately, the story of cinnamon. We can infer information on the trade routes and the peoples who transmitted it, and identify the cultures that used and diffused knowledge on it.
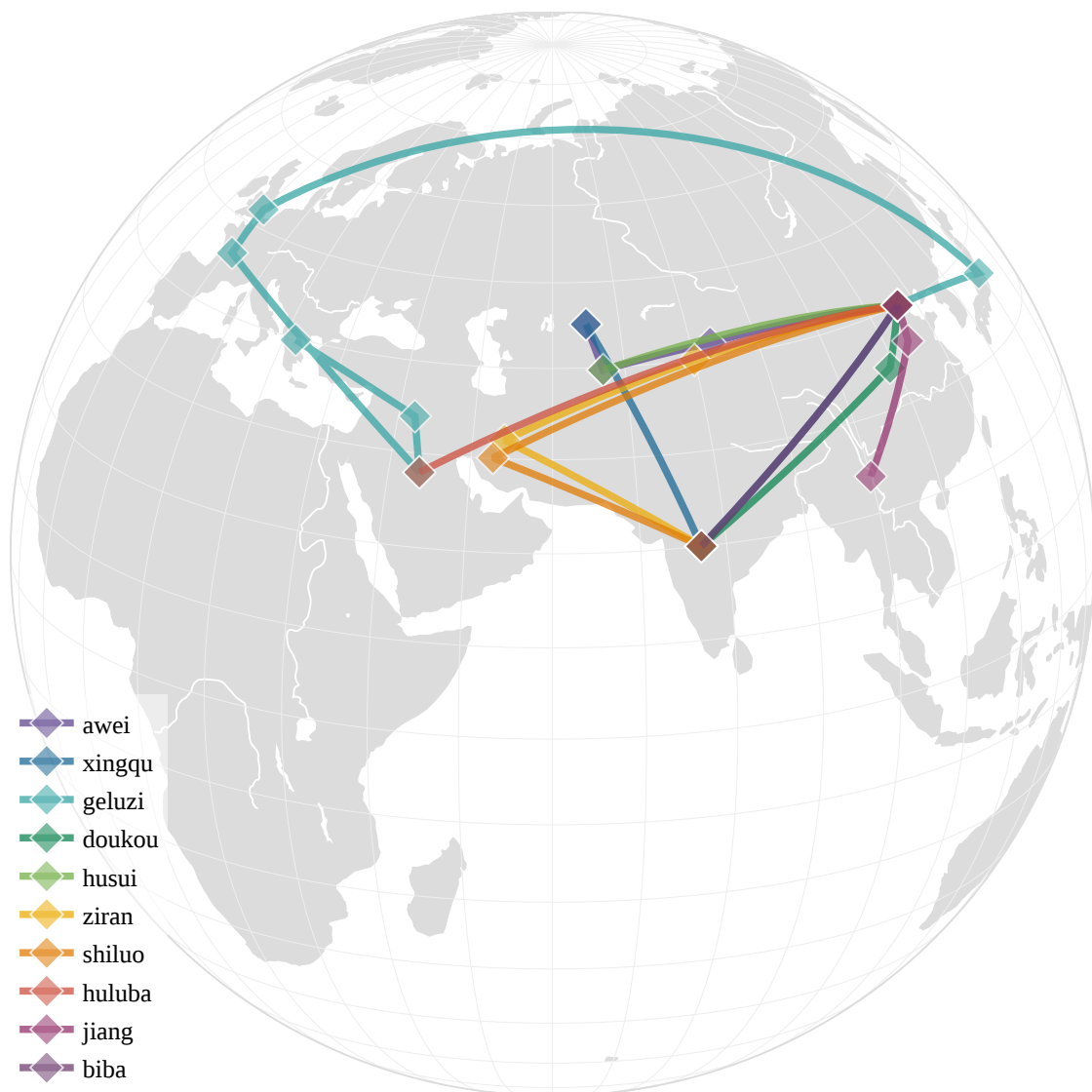
Figure 1.8

To those of us who interested in the spread of words, especially Wanderwörter and their underlying cultural, historical, and geo-political significance, the map of tea might come to mind. This is a map that shows the journey of words for tea (either from Sinitic *cha* or Minnan *te*), and their distribution in a sample of the world's languages. The point of this map is that it clearly shows if the name for tea arrived by overland trade or via a sea route. This peculiar phenomenon is a feature on its own (138A) in *The World Atlas of Language Structures* (WALS), and have been described in a chapter by Dahl (2013).[5] Discussions and maps of the land vs. sea distribution of tea terminology have since made it into popular science magazines and articles, made rounds on Twitter, and hence relatively well known.[6] On a more scientific note, the distribution of tea words are discussed in detail by (Mair & Hoh,

---

[5]The accompanying map is available online at https://wals.info/feature/138A#2/25.5/143.6

[6]See for example Sonnad (2018) in Quartz: https://qz.com/1176962/map-how-the-word-tea-spread-over-land-and-sea-to-conquer-the-world/ or Netchev and Macquire (2022) in the World History Encyclopaedia: https://www.worldhistory.org/i

Words for tea

group
◆ te
◆ cha
◆ other
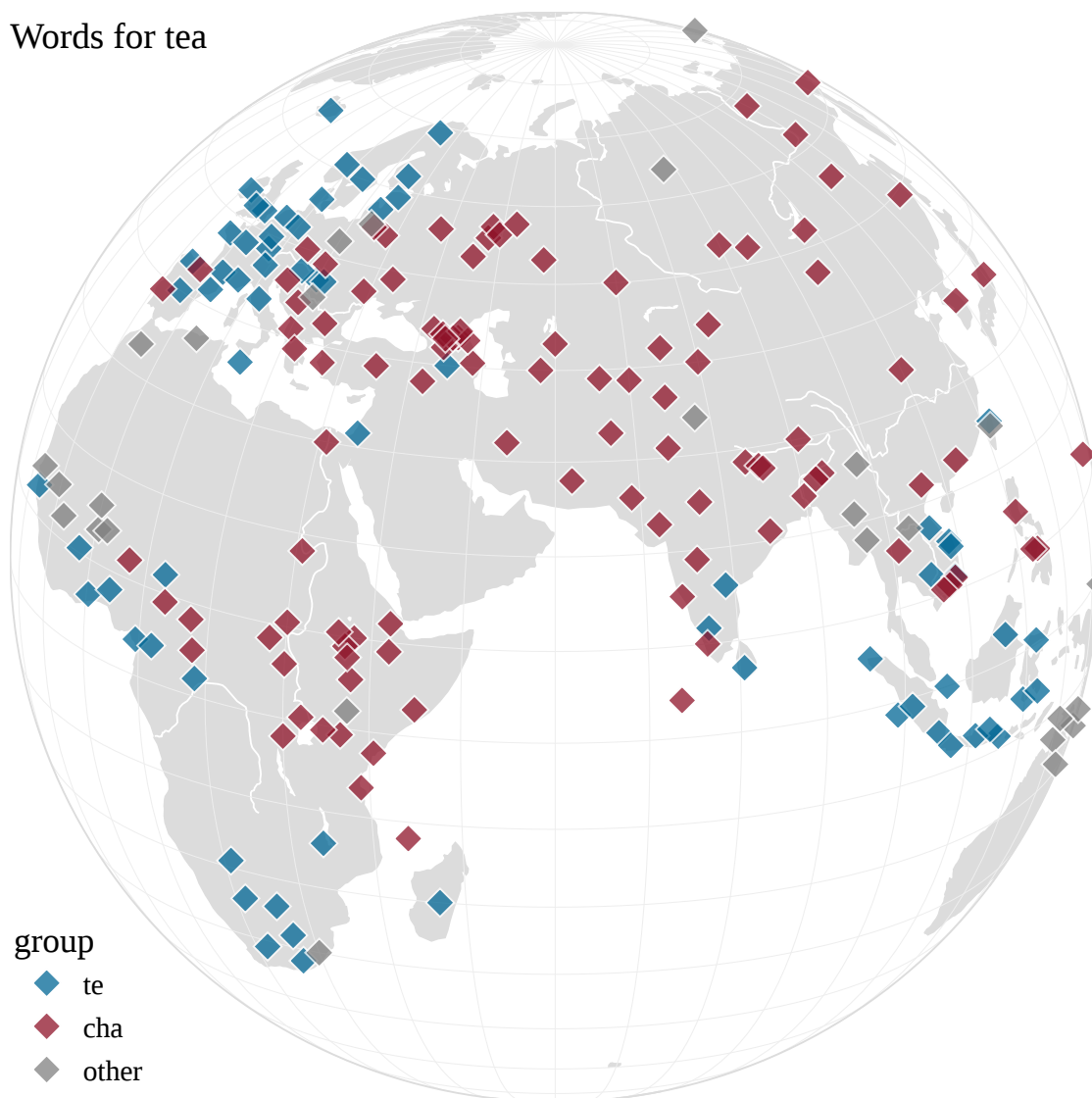
Figure 1.9   Distribution of words for tea from Sinitic *cha* and Minnan *te*, based on the data around the globe.

2009, pp. 261–270) in an appendix titled *A Genealogy of Words for Tea*, with including a discussion on historical phonology.

   Cinnamon as a spice is relatively well known around the world, and the history of its diffusion goes back to thousands of years, with words attested as early as the Bible itself, as it was discussed in **??**. This is in contrast with the story of tea, in the sense that the international spread of tea is a relatively recent process in the economic history of plant products and colonial powers, and so we have a much clearer picture on the exact ways it was transmitted. Although tea-drinking in its homeland was practiced from time immemorial, and trade allowed it to spread regionally on networks, such as the Tea Horse Road, its present global domination is a result of 17th-century European fascination and large scale shipping. While the tea map illustrates the long haul trade connections of the time,

such as those between Europe and the Far East, the map of cinnamon shows traces of an older, more gradual spread that happened in stages, outlining a more geographically contiguous development, and incremental trade networks. The propagation of cinnamonic Wanderwörter mirrors the historical processes, and just as the story of cinnamon, the words' origins are sometimes obscured by the sheer time-depth that is covered.

1.6   Methods and Data

Informative geospatial visualizations such as figure 1.9 above are a powerful tool in conveying the information about spread and distribution of words, and they can also help us to notice patterns and connections faster and easier than studying long tables of words, especially when the distributions are more complex than the somewhat neat duality of tea. In this case study, I will attempt a classification for the words for cinnamon by looking at clusters and categorizing them according to their source, to see what the distribution of names today can tell us about the spread and history of cinnamon.

Because words for cinnamon or other spices are not included as features in balanced typological datasets, such as WALS (tea is an exceptional feature in this database), I have attempted a manual collection of words for cinnamon based on dictionary entries. As a starting point, I have crawled data from the Wiktionary (https://en.wiktionary.org), which is the closest resource we currently have to an open- and crowd-sourced multilingual dictionary. Similarly to the Wikipedia, the Wiktionary is edited and reviewed by the community, which has both advantages and disadvantages. On one hand, information on the Wiktionary is free, broad in scope, it usually represents the public consensus, and often well cited. On the other hand, it is not always complete, the available languages do not represent a balanced sample from a typological point of view, and the information can sometimes be ill-informed or deprecated. In any case it is a rich resource to start with.

For cinnamon, first I scraped the translations for the word *cinnamon* in the sense 'spice' (Wiktionary, n.d.), and cleaned the data using regular expressions. After this, I have performed a round of manual checking where I fixed obvious mistakes in word forms and transliterations by consulting other dictionaries and reference works, in the languages and scripts I felt competent to do so. I proceeded to add a few missing translations with the help of other lexicographical resources and the Google Neural Machine Translation engine's Python API (Wu et al., 2016).[7] Then, I analyzed each word in terms of etymological origin, and assigned them to categories. For example, words derived from Greek *kinnámōmon*, such as Lithuanian *cinamonas* or English *cinnamon* constitute one category, and words derived from Persian *dârčin*, such as Turkish *tarçın* or Hindi *dālcīnī*, make up another. I continued this categorization for all instances, and created a new category for every group that has at least three attested members. Instances that do not belong to any group or undetermined were assigned to "other". Finally, I merged this dataset with language data obtained from the databases of both WALS (Dryer & Haspelmath, 2013) and Glottolog (Hammarström et al., 2022) to prepare for geospatial plotting. The datasets were handled using the `pandas` library in Python, and the visualizations were created using the `plotly` Python library (McKinney, 2011; Plotly Technologies Inc., 2015).

---

[7]https://pypi.org/project/googletrans/

1.7 Results and Discussion

Figure 1.10 shows the results of the analysis above, on a geographical scatter plot. As it can be seen, there are six groups in total: canela, kinnamon, korica, qirfa, darchin, and gui, with a seventh one — other — containing those that do not belong to any of these. It is also noticeable that the groups that were manually identified form geographical clusters, for example, the gui group appears in East Asia, while the canela group is mainly found in Europe. Lastly, I would like to draw attention that the "other" group has a high number of members in regions where cinnamon (or cassia) is native. The canela group represent words that derived from Latin, the kinnamon group contains words going back to Greek, and the korica group represent mostly Slavic languages. Qirfa words are derived from Arabic, darchin gathers terms from the Persianate world, and gui embraces some terms from the Sinosphere. Let us now look at these categories one by one.
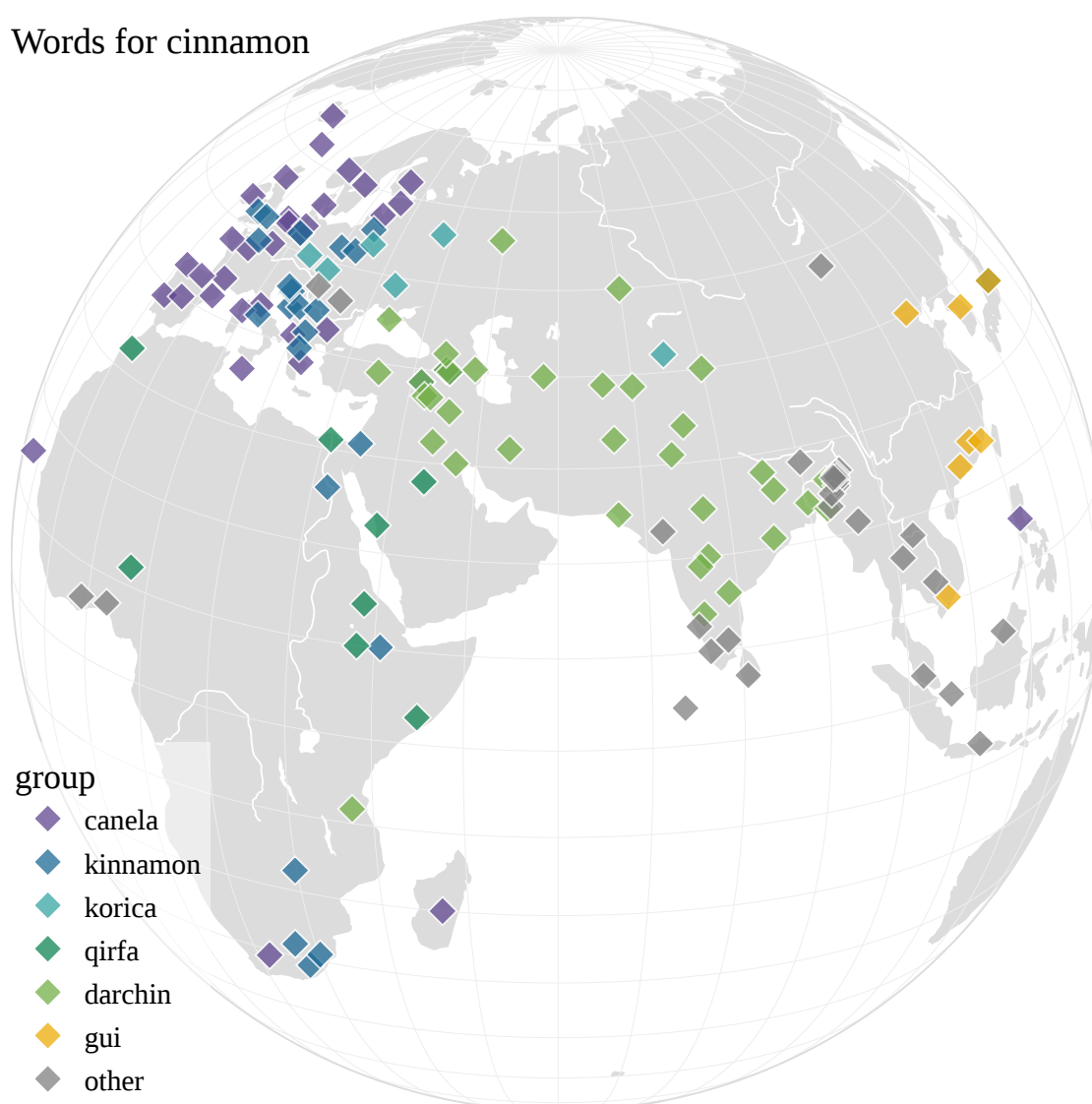
Words for cinnamon

group

◆ canela

◆ kinnamon

◆ korica

◆ qirfa

◆ darchin

◆ gui

◆ other

Figure 1.10   The distribution of *cinnamonic* words in a few languages around the globe.

> **Note 1.7.1.** For a full, interactive and explorable version of the plot, please visit the following link: http://htmlpreview.github.io/?https://github.com/partigabor/phd-test/blob/main/cinnamon.html.[a] The interactive plot can be rotated, zoomed in and out, and the groups of data points can be isolated with a double-click on the group name/icon. Hovering over a data point will bring forward further information on the term, its transliteration, associated language and language family.
>
> ---
>
> [a]For an annotated version, please visit http://htmlpreview.github.io/?https://github.com/partigabor/phd-test/blob/main/cinnamon_annotated.html

1.7.1   The canela group

Words belonging to this group are cognates of Spanish *canela* and its variants in Romance languages, which have been formed with the diminutive of Latin *canna* 'reed, cane'. It is named so after the curled shape of the cinnamon sticks resembling a little, hollow reed-pipe (*Oxford English Dictionary* (OED), "cannel"). Latin *canna* itself is a loanword from Greek κάννᾱ *kánna* 'reed, pole', which is probably a borrowing from a Semitic language (cf. Arabic قناة *qanāh* 'hollow spear, cane; conduit, canal', Hebrew קָנֶה *qāneh* 'stalk, reed, cane', Aramaic ܩܢܝܐ *qanyā* 'id.'[8]) (OED, "cane"). According to Beekes and van Beek (2010, p. 636) the Greek word is from "Babylonian-Assyrian" (Akkadian) 𒄖 *qanû* 'reed', which may come from "Sumerian-Akkadian" (Sumerian) 𒄀 *gin* 'id.' (cf. Roth et al., 1968/2004, vol. 13, p. 85), and proceeds to give Ugaritic *qn* and Punic *qn'* as further Semitic attestations.

The distribution of this group is overwhelming in Europe, which seems to echo the strong influence of Latin vocabulary, especially in the developing Romance languages. One example would be Old French *canele* (modern *cannelle*), which was formed within French from *canne* 'cane', and first attested in the first half of the 12th century from an epic poem describing a fictional expedition of Charlemagne to Jerusalem[9], and the local vendors selling cinnamon, pepper, and "other fine spices" (*Trésor de la Langue Française informatisé* (TLFi), "cannelle")[10]. The TLFi explains that this word exists in most romance languages and it is impossible to determine its progress, and also notes that the medieval Latin is not attested in the 'cinnamon' sense. Either French or Italian was the usual donor for other European languages, take for example Dutch *kaneel*, or Finnish *kaneli* through Swedish *kanel*. Spanish *canela* is attested around 1250, from "Italian" (Medieval Latin) *cannella* (Corominas, 1987, p. 125; Gómez de Silva, 1985, p. 98). Due to later colonization by European powers, many of these terms spread elsewhere, e.g.: Tagalog *kanela* from Spanish, or Haitian Creole *kannèl*.

[†]*Cannel*, also earlier as *canel* had entered English usage in the 13th century from French, but is now obsolete. It existed in Early Modern English up until the 18th century, and was gradually replaced by *cinnamon* (also arriving through French), which was first attested in the first half of the 15th century

---

[8]https://cal.huc.edu/oneentry.php?lemma=qnh+N&cits=all

[9]*Le Pèlerinage de Charlemagne* [*The Pilgrimage of Charlemagne*], or *Voyage de Charlemagne à Jérusalem et à Constantinople* [*Charlemagne's Voyage to Jerusalem and Constantinople*], (c. 1140).

[10]https://www.cnrtl.fr/definition/cannelle

(see Etymology **??**). Neo Latin *canella* also appeared for a brief time, but its meaning as 'cinnamon' waned, and now it is used in botany to refer to a plant genus.

In many other languages of Europe the opposite happened, and an existing word from Greek was replaced by the Latin term. Even Modern Greek uses *kanéla*, re(?)-borrowed from Italian *cannella*, instead of the Ancient Greek *kinnámōmon*.

1.7.2   The kinnamon group

This group centers around Ancient Greek *kinnámōmon*, most possibly a loanword from a Semitic language as I discussed in **??**. *Kinnámōmon* is the source of words for cinnamon in many European languages (e.g.: German *Zimt*, Lithuanian *cinamonas*, and English *cinnamon*), prominently in Central Europe and the Middle East. In most cases, these words represent an area where words derived from Latin cannella (or one of its descendants) did not replace the earlier word derived of *kinnámōmon*. This group also contains South Slavic languages in the Balkan linguistic area (e.g. Slovenian *cimet*, Serbian цимет *cimet*) where it arrived via the earlier German term *Zimmet* (now *Zimt*), and therefore it diverges from West and East Slavic branches for this lexical item. It reached Southeast Europe in the 16th century (Snoj, 1997, s.v. cimet)[11], from which we can assume that cinnamon started to arrived here from the West during this turbulent time in the Balkans, in the middle of the Ottoman Empire's European expansion.

1.7.3   The korica group

The korica group contains languages that use words derived from the inherited Slavic lexicon, in this case the East and West Slavic branches. Proto-Slavic **korica* 'bark' is a derivative of **korà* 'bark'[12], the suffix *-ica* is diminutive. Old Church Slavic *koricę* meant 'cinnamon', and further cognates are Russian *koríca* 'id.', Ukrainian коріця *korýcja* 'id.' (East Slavic), Czech *skořice* 'id.' (West Slavic). In other cases, words derived from **korica* can mean 'bark, crust' (e.g. Serb-Croatian) or 'cover (of a book), binding' (e.g. Bulgarian) (Derksen, 2008, p. 235). Due to the influence of Russian during Soviet times, some Central Asian Turkic languages ended up with a foreign words in their vocabularies, e.g. Kirghiz корица *korica* **??**.

1.7.4   The qirfa group

The qirfa group contains languages from Africa and the Middle East, whose words for cinnamon were borrowed from Arabic *qirfa*, for example Hausa *kirfa* (Newman, 2007, p. 114) and Amharic ቀረፋ *qäräfa* (Leslau, 1996, p. 74).

1.7.5   The darchin group

Names for cinnamon in this category originate from Persian, as it was explained in **??**. According to the data this cluster has the largest geographical extent, and by number of instances constitutes the

---

[11]Fran — https://fran.si/193/marko-snoj-slovenski-etimoloski-slovar/4285437/cimet?View=1&Query=cimet
[12]*Proto-Indo-European* (PIE) **(s)kor-* 'to cut' **??**

largest group, almost head to head with the group of canela. Darchin represents the earliest stage of cinnamon's westward spread from South, Southeast, or East Asia, depending which cinnamon or cassia we think became the first cinnamon of commerce. Consulting the plot we can witness the huge influence Persian had in this step of transmission to the Middle East and Central Asia. We can also see that central and north Indian languages use a loanword from Persian, which can be explained by the Persianate[13] societies that resulted from the Islamic conquest of India, starting from the 13th century. The first sultan to ravage the land, Mahmud of Ghazni was a Persianized *mamluk* Turk, who laid the foundations with his raids in the 11th century for a series of Muslim dynasties on the Indian subcontinent, culminating in the Mughal Empire (1526–1857) and what we define today as Indo-Persian culture (Eaton, 2019, p. 33).

1.7.6 The gui group

The gui group contains terms from the Sinosphere, words that borrowed the Sinogram 桂 *gui* (see ??), such as Japanese 桂 *kei* 'cinnamon or cassia tree', synonym with 肉桂 (肉桂) *nikkei*, Korean 계 *gye* as 계피 (桂皮) *gyepi* and 육계 (肉桂) and the Sino-Vietnamese quế. This shows that the the Chinese transmitted their cassia to their immediate neighbors East and Southwest, together with the word and character for it. However, there is little evidence for trade in cinnamon between China and Southeast Asia in early history, Wang (1958) does not give any information on it in his "The Nanhai Trade: A Study of the Early History of Chinese Trade in the South China Sea." (Wang, 1958) This makes sense if we remember that all regions active in the South China Sea maritime trade — from Guangdong to Sumatra to Lanka — had their own source of cinnamon, and traders would only transport it westwards.

1.7.7 Others

We can see that the category of "other" is prevalent in areas where cinnamon of various kinds is native and therefore these languages often have native words to refer to it. Many words from these group are derived from the meaning of 'tree bark, skin, peel' Malay/Indonesian *kulit kayu manis* [bark-wood-sweet] 'sweet wood bark', where *kulit* 'skin, bark' is often omitted, or Dhivehi *fonithoshi* [sweet-bark].

Hungarian *fahéj* [tree-bark] is made by compounding and was attested in ca. 1395 (Zaicz, 2006, s.v. fahéj),

Romanian *scorțișoară*[14], is perhaps modeled after Slavic **korica*.

scortea,

scortum

*(s)ker- ("to cut")

1.8 Two

So what does this tell us exactly? It shows that in East Asia Chinese, especially the Chinese writing had influence over its neighbors...?

1.9 Conclusion

1.10 Limitations

The Names of Spices

Now that the detailed introduction of the spices is complete, let us examine these spice names comparatively as three sets representing the nomenclature in English, Arabic, and Chinese. This chapter constitutes the analysis and findings part of the thesis, and will thematically introduce certain aspects of the terminology of the spice domain, guiding the reader from a general overview towards more nuanced probes that can be derived from the results. The aim of this chapter is to showcase the many ways we can interpret, analyze, and visualize the data.

2.1 Overview: Figures and Statistics

As a result of the data collection set forth in ??, the database now contains 369 spice names. Of these, 159 are in English, 87 are in Arabic, and 123 are in Chinese; figure 2.1 shows this distribution. The total number is the accumulation of the lengthy process of carefully compiling the nomenclature for the set of spices as defined at the beginning of the thesis, which consists of 24 different spices. The data collection methods were detailed in ??.

On average, a spice has 14 names, where the max is 44 (chile), the min is 4 (fenugreek and mace). Figure 2.2 show the top ten and the bottom ten spices that have the most and least number of names including all three languages. This measurement might raise some eyebrows, but in fact it is a very good indicator of which spices are more complex in their nomenclature, and therefore

Figure 2.1 The distribution of spice names across the three languages.

which are the most "problematic" to untangle. As we can see, spices that boast with many names include the chili pepper, Sichuan pepper, cassia, false cardamoms, and allspice. These are—not incidentally—the very items that I have dedicated substantially more pages to than some of the other spices, due to issues about their identity or the complexity and richness of their nomenclature. This seems to go hand in hand with matters of biodiversity: chile has countless varieties that have spread to faraway corners of the earth, and now it is a hobby in its own right to cultivate, breed, and crossbreed hot chili pepper cultivars. As we saw, Sichuan peppers species are used across vast regions, and it can cause headache to pin them down exactly, their "boundaries" are not that well defined, and it also needed some explanation to isolate the various sources of cassia types.

On the other hand, spices with the lowest number of names are presumably the most straightforward items, take for example cloves, or vanilla. But What makes a spice "straightforward", or in other words, simple? In my opinion, it is their uniqueness and recognizability. Indeed, if we reflect on our investigation on vanilla in the last section of the previous chapter, we have already established
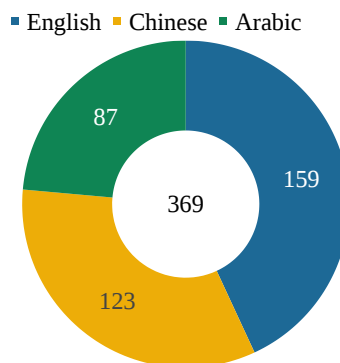
that it is a rather special item: there is no other spice that is made from the fruits of an orchid—it is unique. Or, if we think of cloves, they are unmistakable in their shape and in many language they are known by their shape (see ??). These two items are also very well circumscribed in terms of their geographic origins. Although now cultivated in multiple tropical regions, vanilla is known to be from the jungles of Central America and Brazil, there is no doubt about its origins. The native habitat of cloves is even more narrow, as it is only indigenous only to North Maluku and the "spice islands" of Makian, Ternate, and Tidore. We see nutmeg and mace as well among the bottom five items with the least amount of names, and we should notice that nutmeg and its mace are also from this region, they were exclusively found on the Banda islands of Maluku, and nowhere else until the second half of the 18th century. Now, it makes a bit more sense to look at these same charts deconstructed by language, this can be seen on figure 2.3. The most conspicuous feature of these pie charts is that chili has the most names, across every language.
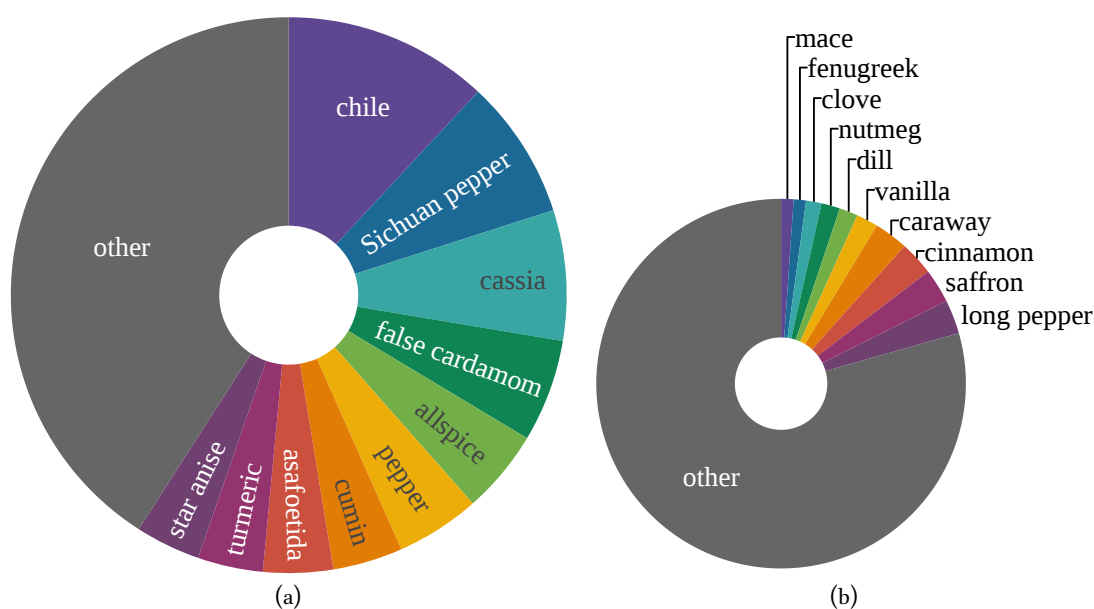
Figure 2.2   Top 10 spices with the most number of names (a), and bottom 10 spices with the least number of names (b).

2.2   The Attestation of Spice Words

One of the most exciting part of this thesis, is the data that was collected regarding dates of attestation. In other words, I tried to find out the earliest possible mentions for each spice, then combine this information in a way that enables us to see the diffusion of spices throughout the history of a language and culture. This kind of information is a valuable indicator, as it shows the approximate times of the earliest contact with the material. In essence, we can see the history of the spice trade in the words, how and when they arrived, which spices were the earliest to be recorded, and which ones make the latest additions in our vocabularies and spice cabinets.

Figure 2.3 Top 5 spices with the most number of names, broken down by language.

From the nearly 400 names, I have chosen a selected few that—for lack of a better word—I marked with "default". These are the name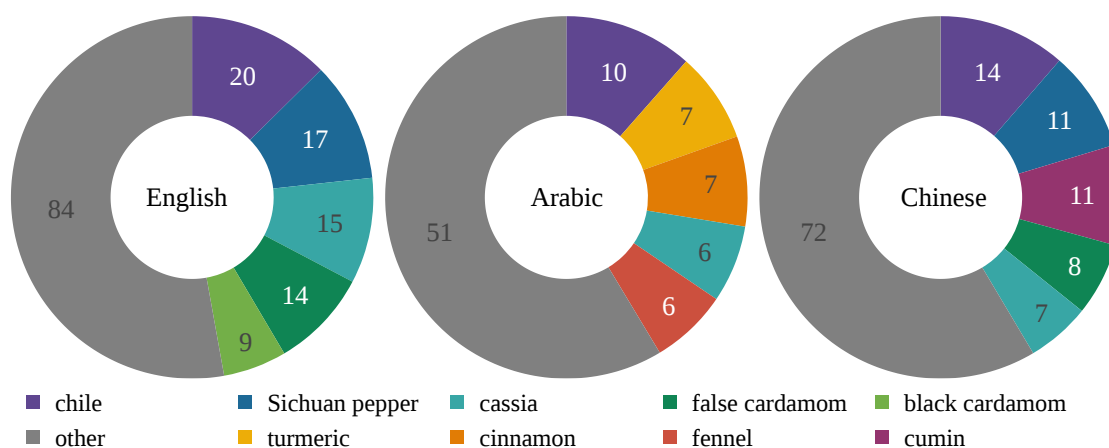s most people are familiar with, the most people search by, and the most prevalent in corpora. These are the names that act also as keys in my datasets. For the attestation visualizations, I only used the default terms, and a couple of historic terms precede the contemporary names. This allows for more clean overview, and a way to compare the attestations in the three languages.

The following plots should give a bird's eye view of the history of the spice domain, and its mark on vocabulary. In figures 2.4 to 2.6, you can see the timeline of the spice nomenclature language by language. Obviously, these figures will show the native spices that are to be found the closest to the homeland of the ancestors of English, Arabic, and Chinese speakers, are recorded first. This means dill and fennel in English, saffron and fenugreek in Arabic, and Sichuan pepper and cassia in Chinese. The figures also show which are the early products of transnational trade, that spread first despite its origins are distant and unknown to the early speakers of the language. These include pepper and ginger. In the final, overlapping plot ??, there is a chance to compare the main attestation periods for these items, followed by a compact version of the same plot (??) accompanied with a histogram to better see which periods have seen the emergence of new spice words, indicating flourishing (scientific) literature and trade.

2.3   The Analysis of Spice Nomenclature

This chapter will present the analysis on these spice names, and try to answer the main question: How do people name spices, specifically, new spices that they came into contact with? Immediately, we can think of two ways: languages either borrow, or conceive a name. But how does this naming process work exactly? What are the underlying mechanisms and critical factors that influence the naming, and how does the nomenclature reflect the contact situation? How does borrowing work, and how languages invent new names for novel materials and substances? In an attempt to give answers to these questions, I will take a bottom-up approach and look at examples from the data I collected to
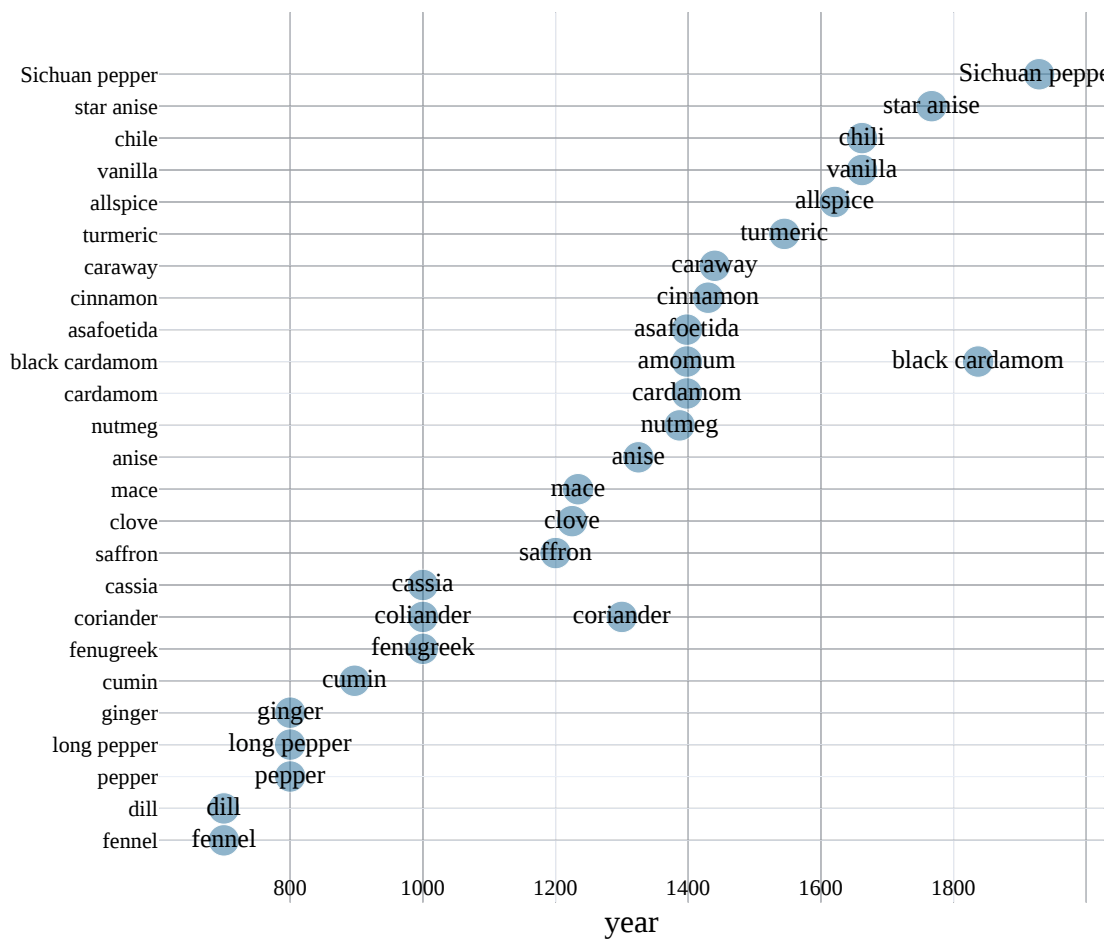
Figure 2.4

arrive to some conclusions.

2.3.1 Terminology

During the analysis, I will take into account the term's (a) analyzability, their (b) borrowed status, and inspect the ways spice terms are generated using (c) prototype words and distinguishing words.

Analizability

Analyzability of words is an idea from the 20th-century philological movement and method *Wörter und Sachen* (words and things in German), which had a big influence on linguistics and ethnography. Outlined by Hugo Schuchardt and based on the titular journal started by Indo-Europeanist Rudolf Meringer in 1909, it proposed the close study of the etymology of words together with the artifacts/concepts (Ortutay, 1977–1982).

"Ohne Sachwissenschaft keine Sprachwissenschaft mehr!" There is no linguistics anymore without the study of material culture!

Basically, the more opaque a name is in terms of morphological analysis, the longer it is assumed
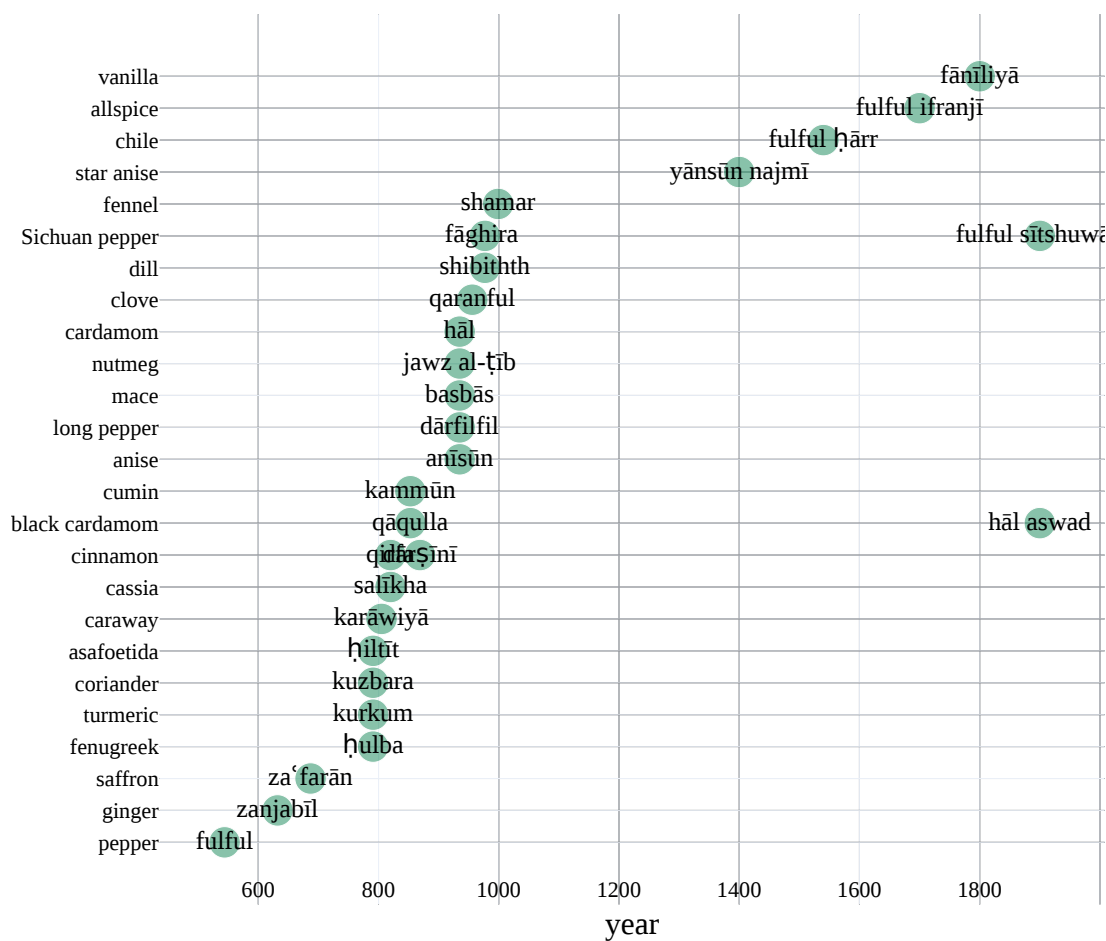
Figure 2.5

to be present in the language. A basic example would be *York* (monomorphemic) vs. *New York* (analysable), which provides a potential chronology for the concepts the words signify. This approach was incorporated into historical linguistic research and philology, often studied in parallel with findings in archeology. SOURCE??

Haspelmath and Tadmor (2009a, p. 12) also used the term "analyzability" in the creation of their loanword database (*The World Loanword Database* (WOLD)) as a first step to assess a word's loanword status, although — to the grief of — in a purely linguistic way.

If the word is morphosyntactically complex, "it was almost certain that it was created by speakers of the language rather than borrowed from some other language" — we can read. The authors also state that these are not considered loanwords, even when they contained borrowed elements.

Borrowed Word or Native Invention

Closely related to analyzability, is the question if a term is borrowed or not.
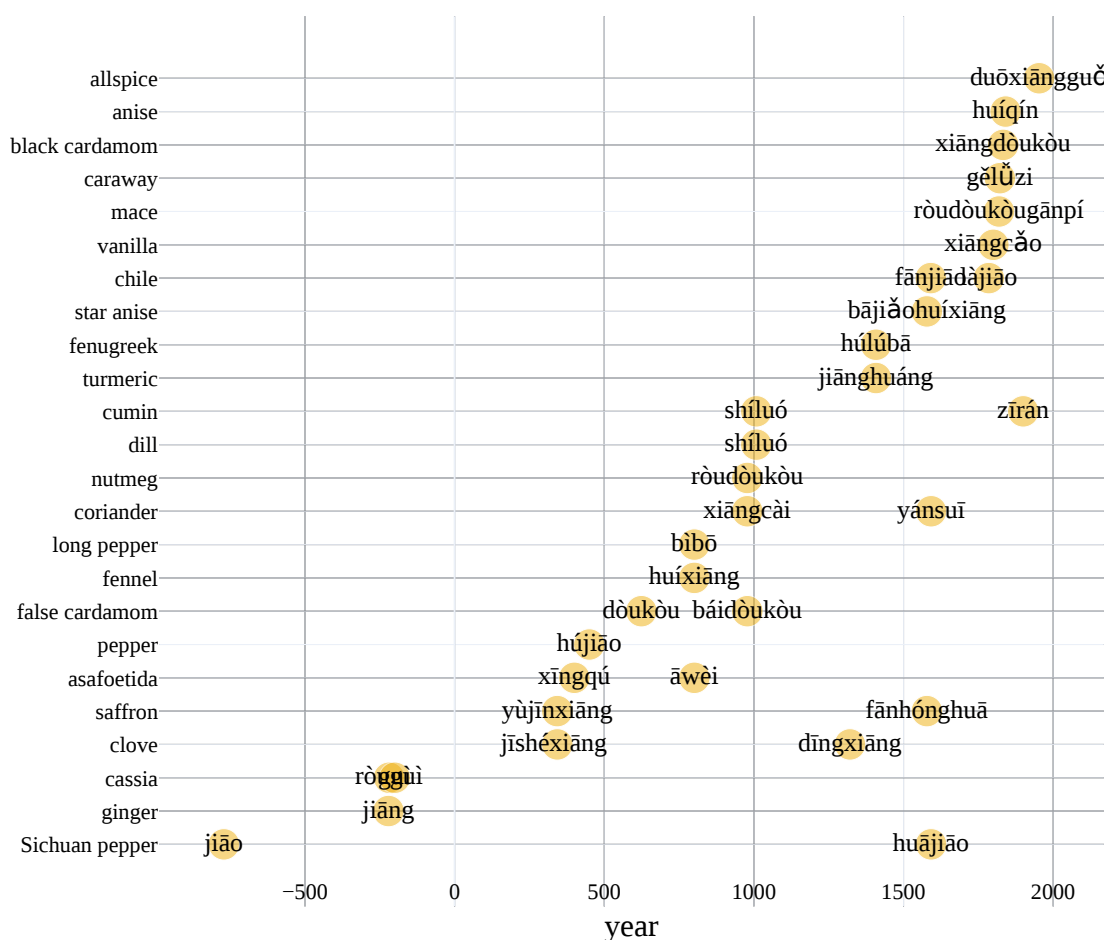
Figure 2.6

Prototype and Distinguishing Words

2.4 The Case of Star Anise

Let us consider the nomenclature of star anise in the three languages (see ??). In English, there is the default *star anise*, which is a native invention, obviously after the fruit's unmistakable appearance. On a rare occasion, we have information on the exact time of star anise's arrival to England, which is dated to 1588, as it was introduced in ??. The same idea for a name is found in most European languages, either influenced by 16-17th-century spice dealer terminology, or devised on their own conviction, looking at its recognizable shape. I used the word "native", even though the phrase is obviously mixed from an etymological point of view: *anise* is a loanword ultimately from Greek. However, when faced with this type of phrases, I consider that at the time of the contact situation, *anise* was already part of the English lexicon — as well as *star* — therefore, this phrase was coined within English, and deemed as a native creation. This practice is consistent with the approach took by the team of Haspelmath and Tadmor (2009b) at WOLD. English also has the term *Chinese anise*, which is a phrase consisting of *anise*, again, and *Chinese*, referring to star anise's geographical location and the origin of
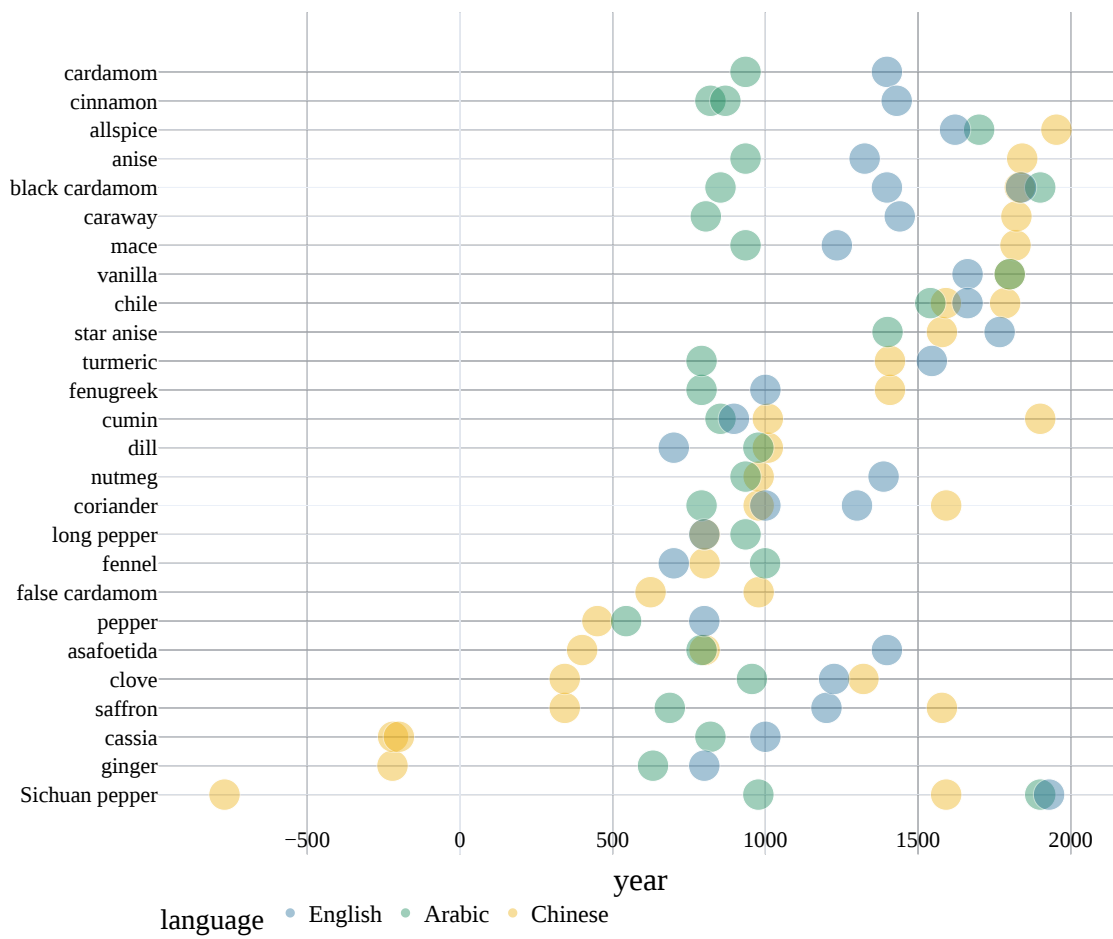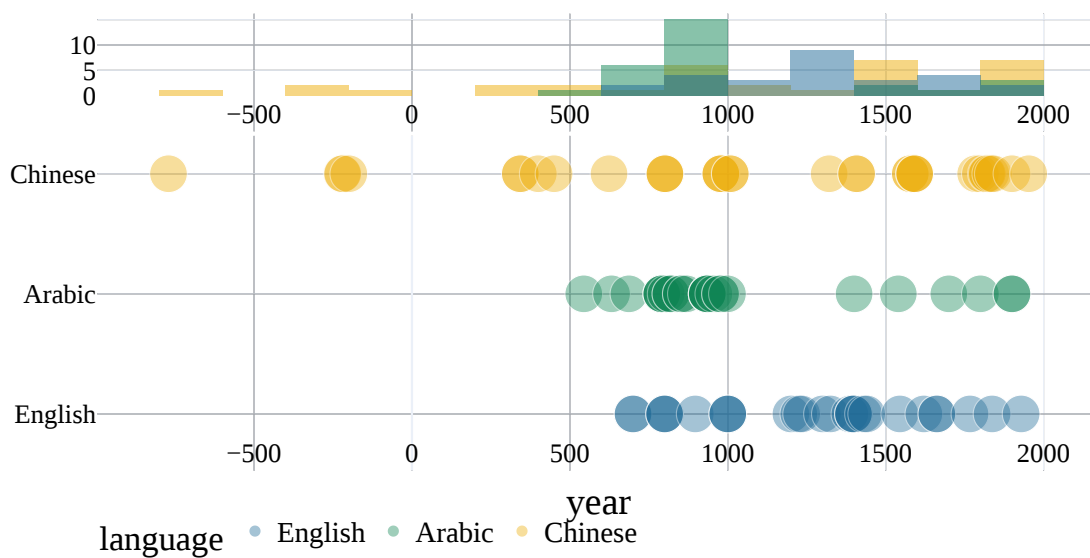
Figure 2.7

Figure 2.8

its procurement for the English. Both phrases utilize the term *anise*, which refers to the small anise seeds of the Mediterranean, used as a spice, and flavouring for liqueurs and confectionary (see ??). Why is there a connection to anise? The two plants could not be more different, they are geographically distant, they are botanically unrelated. The only thing that connects them is their highly similar flavor profile, dominated by the volatile oil anethole, the same nauseating and sweet chemical compound that is found in fennel and licorice. And so, for the Europeans who were familiar with anise and its taste, the novel product reminded them of anise's aroma. Hence, the names are in part inspired by taste/plant chemistry, defining anise as a prototype spice and protoype term. To avoid confusion, (the existence of which will be clear to anyone who tries to do a brief search about anise or star anise), distinguishing words are used for the new material. These modifiers are attached to the head word, and in one case inspired by the spice's shape, on the other hand referring to its geopgraphical origin. The existence of a Chinese star anise could be explained by the fact that there is a Japanese star anise as well, a similar looking but poisonous fruit and tree, *Illicium anisatum*. In short, the two phrases have different ways to identify this spice. English also has a now archaic form referring to star anise: *badian* from French, which arrived via a land route through Persian, perhaps a phonetic loan from Chinese, but there is no documentary evidence for this (see Etymology ??).

Arabic *yansūn najmī* [star anise] was devised along similar lines, using a native Arabic word for 'star', the prototype word is anise, and the more interesting instances are to be found in neighboring Persian. *Bādyān khatā'ī* or *khatāyī* [star anise] is star anise, while *bādyān rūmī* [Roman anise] is anise.[1] *Bādyān* alone could also refer to fennel.[2] This shows, that in Persian, the prototype word was *bādyān*.

As for Chinese, we do not find any loanword among the terms used to refer to star anise, all names are local "inventions". The modern "proper name" for star anise is *bājiǎohuíxiāng* [eight-horn-hui-spice], where [eight-horn] means 'octagonal', and [hui-spice] is fennel, therefore it can be translated as 'octagonal fennel', or 'eight-horned fennel'. An other name, *dàhuíxiāng* 'big-fennel' strengthens the assumption that in Chinese, *huíxiāng* 'fennel' is the prototype. Again, the flavor profiles of fennel and anise are basically identical, hence the connection (and confusio). The formal Chinese names of star anise are not attested in historical corpora as we discussed in ??, and I assume that the vernacular name of *bājiǎo* [eight-horn] was first applied to star anise, and the formal name was modelled later driven by the plant sciences. In modern dialects star anise is also referred to as *huíxiāng* 'hui-spice' (historically 'fennel') and *dàxiāng* 'big-spice'. In modern *Traditional Chinese Medicine* (TCM), fennel is referred to as *xiǎohuíxiāng* 'little-hui-spice', contrasting the two spices that are confounded due to their taste, using size. In fact, the Chinese 大/小 *dà/xiǎ* 'greater/lesser' contrast is not necessarily a marker of size, but a semantic tool to convey unmarked/marked, or proper/imitator.

To summarize the points I intended to make above: First, I determined if the words and phrases are analyzable (morphologically, syntactically, semantically), then I examined those names further, while also stating why a specific item is unanalyzable. E.g., *badian* as a loanword does not carry any useful information for an English speaker that is not familiar with the word, it cannot be dissected

---

[1] Hayyim, 1934–1936, vol. 1, p. 197.
[2] Steingass, 1892, p. 140.

or interpreted alone. Next, I looked at the borrowed status of the names to determine if the word or phrase is borrowed, or devised locally. E.g., the Chinese names are native "lexical creations", while English and Arabic use a non-native headword (*anise/yansūn*) and a native distinguishing word (*star/najmī*). Finally, I have looked at the inspirations behind these lexical inventions, and identified the rationale and motivation behind them. For phrases and compound words, we can separate a prototype word (headword), and a distinguishing word (modifier). In each case, we can discern the reasons why that prototype word was used, what feature of the prototype item (referent) is the most salient. The same is true for the distinguishing word(s). For example, *star anise* is named so after (1) similarity in taste + (2) shape; and *Chinese star anise* is named so after (1) similarity in taste + (2) shape + (3) geographic origin. In table 2.1, you can see a concise overview of the analysis of star anise terminology.

Term	Gloss	Analyzability	Borrowed	Prototype	Modifier
star anise		analyzable	native	similarity in taste	shape
badian		unanalyzable	borrowed		
Chinese anise		analyzable	native	similarity in taste	origin
Chinese star anise		analyzable	native	similarity in taste	shape + origin
yansūn najmī	star anise	analyzable	native	similarity in taste	shape
bājiǎo	octagonal	analyzable	native	shape	
bājiǎohuíxiāng	octagonal-fennel	analyzable	native	similarity in taste	shape
bóhuíxiāng	ship-fennel	analyzable	native	similarity in taste	shape
dàhuíxiāng	big-fennel	analyzable	native	similarity in taste	size*
dàliào	big-ingredient	analyzable	native	function	size*

Table 2.1 Cap

In this sense, the space names are layered. Intuitively, the more layers a spice name has, the more distant the item was culturally, and on the converse, the less components there is to a term, more familiarity with the substance is presumed (e.g., anise vs. star anise in English). Therefore, spice names' modifiers can be categorized according to what salient feature contributed to the naming the most, and in this specific case, it is star anise's distinct shape. As we will later see, shape is just one of many properties that can distinguish/identify a spice, for others, different properties are salient, including color, taste, smell, and the geographical origin we mentioned. Furthermore, these names reflect on the materials' physical qualities, and the perception and importance of a spice for various sensory modalities in the human experience: vision, gustation, olfaction, etc.
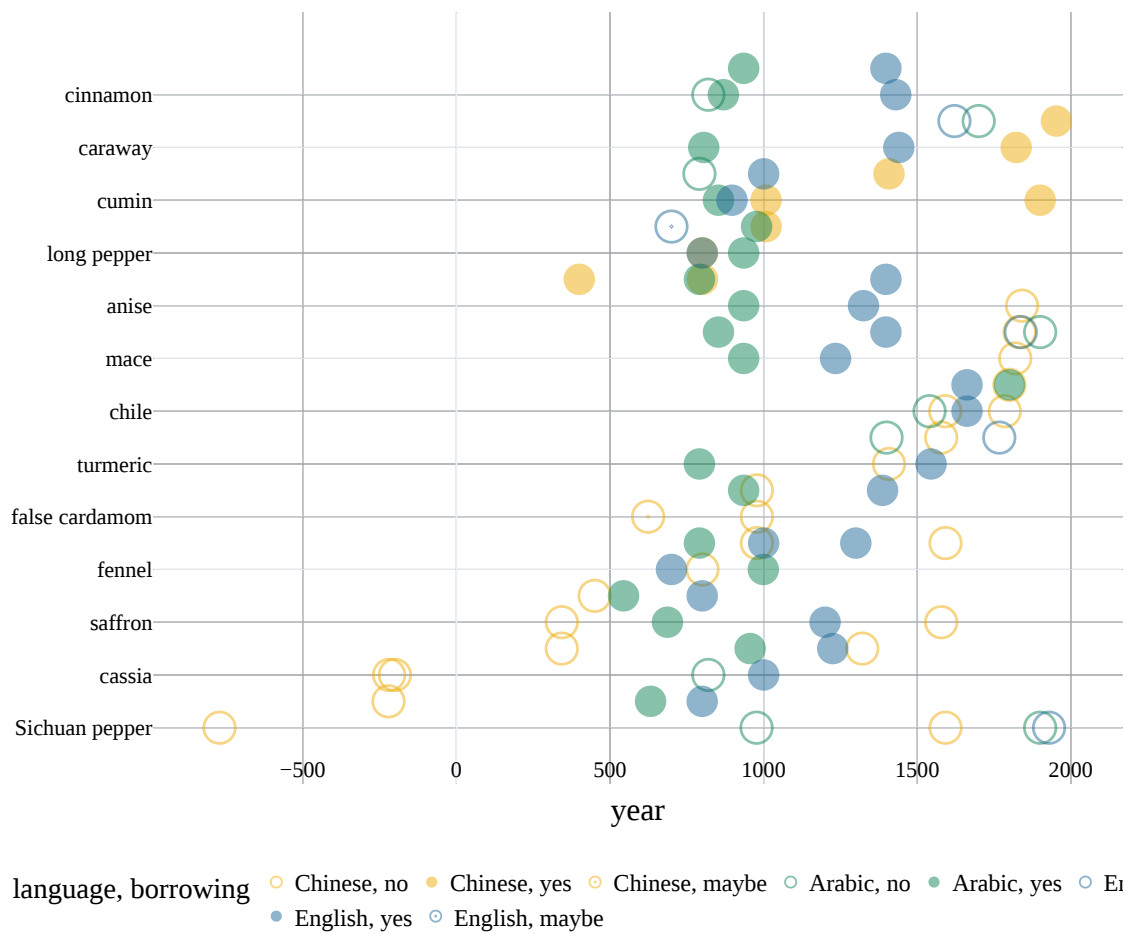
2.4.1 Borrowed

2.4.2 Donor Languages

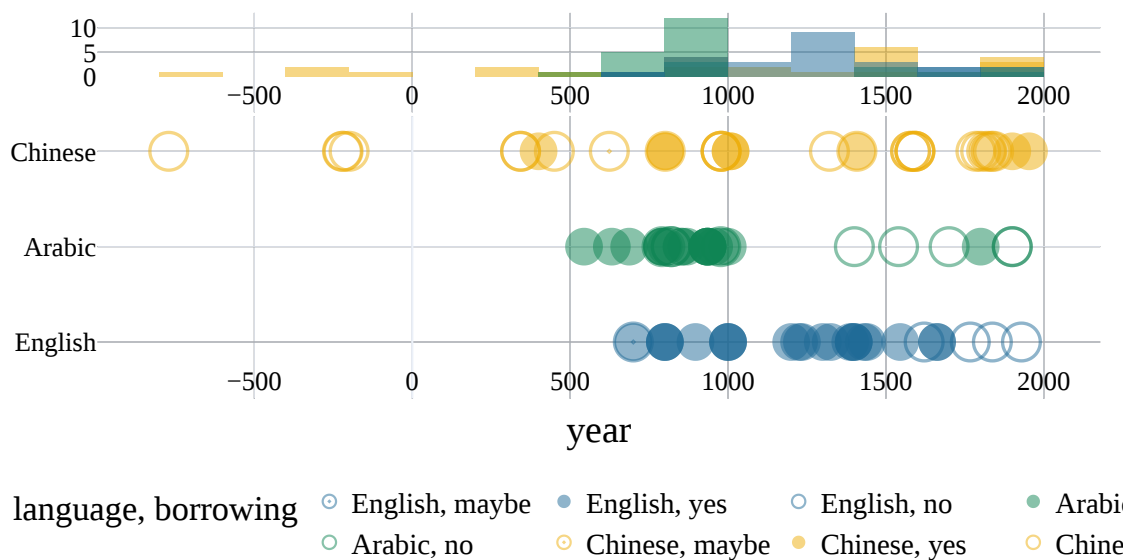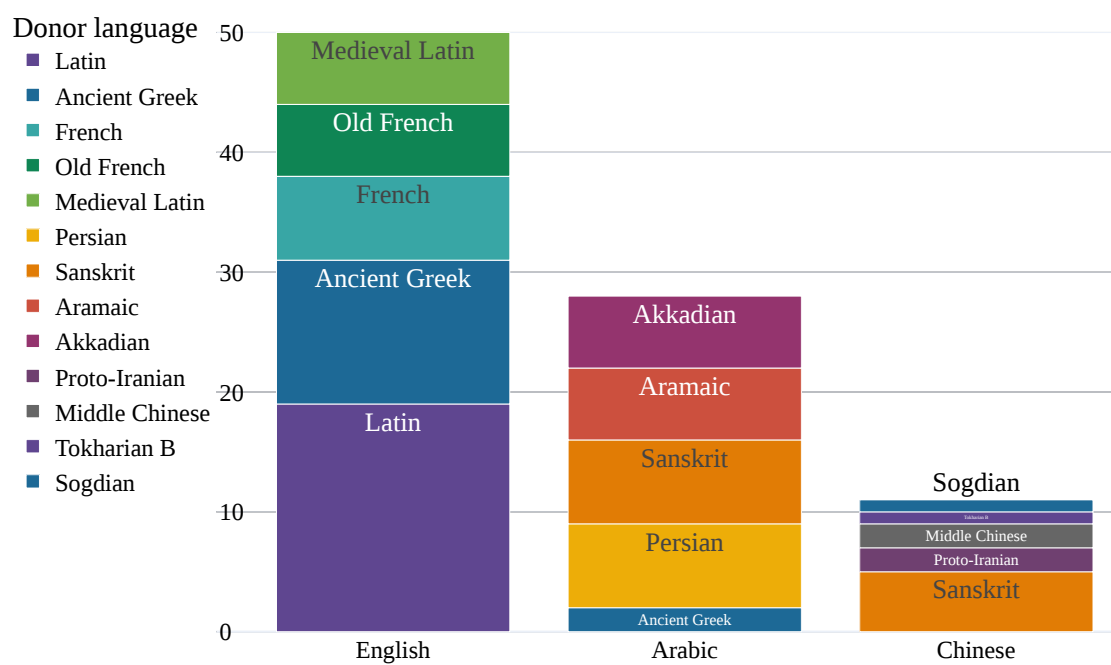Figure 2.9   Borrowed spice terms across the three languages

Figure 2.10

Figure 2.11

Bibliography

Beekes, R. S. P., & van Beek, L. (2010). *Etymological Dictionary of Greek*. Brill. (Cited on p. 12).

Corominas, J. (1987). *Breve Diccionario Etimológico De La Lengua Castellana* [*Brief Etymological Dictionary of the Castilian Language*] (3rd ed.). Editorial Gredos. (Cited on p. 12).

Dahl, Ö. (2013). Tea. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/138 (Cited on p. 8).

Derksen, R. (2008). *Etymological Dictionary of the Slavic Inherited Lexicon*. Brill. (Cited on p. 13).

Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *WALS Online*. Max Planck Institute for Evolutionary Anthropology. https://wals.info/ (Cited on p. 10).

Eaton, R. M. (2019). *India in the Persianate Age*: *1000–1765*. Allen Lane. (Cited on p. 14).

Gómez de Silva, G. (1985). *Elsevier's Concise Spanish Etymological Dictionary*. Elsevier. (Cited on p. 12).

Green, N. (Ed.). (2019). *The Persianate World*: *The Frontiers of a Eurasian Lingua Franca* (1st ed.). University of California Press. https://doi.org/10.2307/j.ctvr7fdrv (Cited on p. 14).

Hammarström, H., Forkel, R., Haspelmath, M., & Bank, S. (2022). *Glottolog 4.6*. Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.6578297 (Cited on p. 10).

Haspelmath, M., & Tadmor, U. (Eds.). (2009a). *Loanwords in the World's Languages*: *A Comparative Handbook*. De Gruyter Mouton. https://doi.org/doi:10.1515/9783110218442 (Cited on p. 21).

Haspelmath, M., & Tadmor, U. (Eds.). (2009b). *The World Loanword Database* (*WOLD*). Max Planck Institute for Evolutionary Anthropology. https://wold.clld.org/ (Cited on p. 22).

Hayyim, S. (1934–1936). *New Persian-English Dictionary, Complete and Modern, Designed to Give the English Meanings of over 50,000 Words, Terms, Idioms, and Proverbs in the Persian Language, as Well as the Transliteration of the Words in English Characters. Together with a Sufficient Treatment of All the Grammatical Features of the Persian Language*. Librairie-imprimerie Béroukhim. https://dsal.uchicago.edu/dictionaries/hayyim/ (Cited on p. 24).

Leslau, W. (1996). *Concise Amharic Dictionary*. University of California Press. (Cited on p. 13).

Mair, V. H., & Hoh, E. (2009). *The True History of Tea*. Thames & Hudson. (Cited on p. 8).

McKinney, W. (2011). Pandas: A Foundational Python Library for Data Analysis and Statistics. *Python for high performance and scientific computing*, *14*(9), 1–9 (Cited on p. 10).

Merriam-Webster. (n.d.). *Merriam Webster's Unabridged Dictionary*. Retrieved November 12, 2021, from https://unabridged.merriam-webster.com/ (Cited on p. xv).

Netchev, S., & Macquire, K. (2022, April 14). *Movement of "Tea" & "Cha" Around the Globe*. World History Encyclopedia. Retrieved May 23, 2022, from https://www.worldhistory.org/image/14112/movement-of-tea--cha-around-the-globe/ (Cited on p. 8).

Newman, P. (2007). *A Hausa-English Dictionary*. Yale University Press. (Cited on p. 13).

Ortutay, G. (Ed.). (1977–1982). *Magyar Néprajzi Lexikon* [*Hungarian Ethnographic Lexicon*]. Akadémiai Kiadó. http://mek.niif.hu/02100/02115/html/ (Cited on p. 20).

Plotly Technologies Inc. (2015). *Collaborative Data Science*. https://plot.ly (Cited on p. 10).

Roth, M. T., Biggs, R. D., Brinkman, J. A., Civil, M., Farber, W., Reiner, E., & Stolper, M. W. (Eds.). (2004). *The Assyrian Dictionary of the Oriental Institute of the University of Chicago* (4th ed.). The Oriental Institute. https://oi.uchicago.edu/research/publications/assyrian-dictionary-oriental-institute-university-chicago-cad (Original work published 1968). (Cited on p. 12)

Snoj, M. (1997). *Slovenski Etimološki Slovar*. Zalozba Mladinska Knjiga. (Cited on p. 13).

Sonnad, N. (2018, January 11). *Tea If by Sea, Cha If by Land: Why the World Only Has Two Words for Tea*. Quartz. Retrieved May 23, 2022, from https://qz.com/1176962/map-how-the-word-tea-spread-over-land-and-sea-to-conquer-the-world/ (Cited on p. 8).

Steingass, F. J. (1892). *A Comprehensive Persian-English Dictionary: Including the Arabic Words and Phrases to Be Met With in Persian Literature*. Routledge & K. Paul. https://dsal.uchicago.edu/dictionaries/steingass/ (Cited on p. 24).

Wang, G. (1958). The Nanhai Trade: A Study of the Early History of Chinese Trade in the South China Sea. *Journal of the Malayan Branch of the Royal Asiatic Society*, *31/2*(182), 1–135 (Cited on p. 14).

Wiktionary. (n.d.). Cinnamon. In *Wiktionary*. Retrieved May 26, 2022, from https://en.wiktionary.org/w/index.php?title=cinnamon&oldid=66914125 (Cited on p. 10). Page Version ID: 66914125.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., … Dean, J. (2016, October 8). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. https://doi.org/10.48550/arXiv.1609.08144 (Cited on p. 10).

Zaicz, G. (2006). *Etimológiai Szótár: Magyar Szavak és Toldalékok Eredete* [*Etymological Dictionary: The Origin of Hungarian Words and Affixes*]. Tinta Könyvkiadó. (Cited on p. 14).