

Heritage Dictionaries, Historical Corpora and other Sources: Essential And Negligible Information

Alina Villalva

Faculdade de Letras & Centro de Linguística da Universidade de Lisboa

E-mail: alinavillalva@campus.ul.pt

Abstract

Contemporary dictionaries are often the result of accumulating the contents of previous dictionaries, namely heritage or patrimonial dictionaries, which are those that set a landmark in the historical lexicography of a given language. This is certainly the case for European Portuguese dictionaries, which offer word lists containing words in usage as well as rarely used words, words that belong to other language variants and phonetic/orthographic and morphological alternates, being apparently unable to distinguish among relevant, puzzling, or simply useless information.

If contemporary dictionaries were to be redesigned, should lexicographic ancestry be ignored, or should it be dealt with otherwise? The discussion on new lexicographic models is beyond the range of this paper, but any lexicographic innovation must be grounded in solid lexicological analyses that cannot ignore the consideration of heritage dictionaries and historical *corpora*. These sources provide a large quantity of information requiring specialized interpretation and critical trimming, which may nevertheless be insufficient to fully grasp a thorough knowledge of the related words. This paradox, which is the focus of the present work, will be addressed by advocating the need to adopt a lexicographic protocol that may help to select the right amount of information and combine it with identical information from other languages.

Keywords: heritage dictionaries, historical corpora, European Roots

Studying words is a lengthy task that relies on the collection of information, and on the ability to interpret the relevant bits and to relate them in a meaningful way. Heritage dictionaries, crafted according to coeval word knowledge, convey some of this information, but these dictionaries are the privilege of only some languages and some moments in their history. Iberian languages, for instance, have had bilingual dictionaries since Nebrija's *Dictionarium*, published in 1492, but monolingual dictionaries, such as Bluteau's *Vocabulario Portuguez e Latino*, with ten volumes published between 1712 and 1728, arrived considerably later. Furthermore, heritage dictionaries need to be read in their time. Bluteau's long entry on *molher* 'woman' looks like misogynist libel to our eyes, but it has to be interpreted culturally. Notice that, in this entry, Bluteau invokes several classical 'authorities', like Salomon, Diphilus, Socrates, Democritus, or Tacitus, to substantiate a negative portrait of women, which is crowned by a quotation of Scaliger's definition of *femina* "natura sua infida est, suspicax, inconstans, insidiosa, simulatrix, supersticiosa, cui si potentia adjuncta est, fit intolerabilis". Bluteau's entry shows to posterity how men, and particularly clergyman like himself, saw women, and that in itself is an interesting piece of information.

Therefore, not only do heritage dictionaries fail to cover the whole existence of older words, but they also provide insights biased by what was then the received state of the art. Hence, other possibilities of data collection need to be considered. Historical *corpora*, despite covering only a sample of written production, do help to fill in lexicographic gaps, particularly those that precede the appearance of early dictionaries, and they allow us to crosscheck lexicographic inputs with real usage.

Unfortunately, historical textual databases are not yet as powerful as we would like them to be – the coverage has to be improved, the philological quality needs to be certified, and text annotation requires a heavy workload. Time and academic contributions will probably help to overcome some shortcomings of these databases, but professional and specifically trained manpower allocation is essential to the much needed upgrading.

This said, it becomes self-evident that the consideration of heritage lexicographic and historical textual information raises numerous methodological problems, particularly those that are related to the fact that dealing with these sources requires specialized interpretation and critical trimming. This is indeed a gigantic task, which may benefit from the adoption of a lexicographic model such as the European Roots prototype¹ that I have described on several occasions (cf. Silvestre & Villalva 2014, Villalva & Silvestre 2015). The ER dictionary prototype aims to serve as a principled source of information for general monolingual and multilingual dictionary makers, since it is intended to help to select the right amount of information and to combine it with identical information from other languages.

According to this model, lexicologists of each language must identify the set of heritage and contemporary dictionaries that must be considered. In the case of Portuguese, the list includes eight items: Cardoso (1569), Barbosa (1611), Pereira (1697) and Bluteau (1712-1728), available online at *Corpus Lexicográfico do Português* (CLP), Figueiredo (1913), also available online at *Dicionário Aberto*, and Morais Silva (1813) that can be consulted online at the website of Biblioteca Brasileira Guita e José Mindlin. Two other online dictionaries (*Infopedia* and *Priberam*) embody the contemporary stage. Each of these dictionaries represented a turning point in the Portuguese lexicographic panorama at the moment when they were released – they are not exempt from criticism, but they are the best there is. The same applies to historical textual *corpora*. For Portuguese, we may use *Corpus do Português* (CdP), that includes texts from the 13th to the 20th centuries, and *Corpus de Referência do Português Contemporâneo* (CRPC), that comprises texts (and some transcripts from oral speech) from the mid-19th century to present-day.

Though the information provided by these sources is enormous, it may nevertheless be insufficient to fully grasp a thorough knowledge of the related words. This is probably the most serious issue and deserves further discussion, and it will be the focus of this paper, as demonstrated by the complexity of the analysis of the Portuguese word *laranja* ‘orange’, some cognates like the Castilian *naranja* and the French *orange*, and some etymologically unrelated equivalents like the Greek *πορτοκαλί* [portoka’li]².

1 The Complexity of *Laranja*

Mabberley (2004: 483) claims that “the names of both citrus and orange are surrounded by a series of confusions, false etymologies and perhaps puns”. The Portuguese noun *laranja* is equally difficult to grasp, but it is a piece of the general puzzle that has not often been considered by itself. The widespread tendency to look at Portuguese as a variety of Castilian may have obfuscated any relevant specific information.

The etymological information on *laranja* ‘orange’ conveyed by Cunha (1986, s.v. *laranja*), a Portuguese etymological dictionary, refers to an Arabic etymon (e.g. *nāranġa*) that originates in a Sanskrit word (e.g. *nāraṅġ*). Contemporary dictionaries, like *Infopedia* and *Priberam*, echo the same

1 This project is published at sites.google.com/a/campus.ul.pt/european-roots/. Landlex, a more recent project devoted to the study of the landscape lexicon, has been testing the ER prototype, particularly in German and Modern Greek, besides Portuguese (cf. sites.google.com/a/campus.ul.pt/landlex---lisbon-meeting/about-landlex).

2 I am indebted to Simeon Tsolakidis, from the University of Patras, who provided the information on Modern Greek *πορτοκαλί*.

information, regardless of how they spell these etymons³. Cunha also claims that the word first occurred in the 14th century, but this *a quo* terminus is not documented⁴. Since Portuguese was never in direct contact with Sanskrit, a vehicular language has indeed to be considered, but Arabic is probably not the best immediate candidate. In fact, language contact between Portuguese and Arabic existed for centuries, but it stopped after the definitive Portuguese conquest of the Algarve, in the mid-13th century, and even if we admit later commercial exchanges, we still need to explain how the initial consonant changed from the nasal [n] to the lateral [l], which is not a common phonetic change – this is a major setback for the Arabic loan hypothesis. Etymological research must be plausible, and, in this case, there is, at least, one missing link.

Where does Cunha's hypothesis come from? The etymological dictionaries by Corominas, though devoted to Castilian, were and still are very influential references for both Portuguese and Brazilian etymologists. Therefore, it makes sense reading Cunha's entry on *laranja* as an adaptation of Corominas (1981) entry on *naranja*, but the hypothesis of a parallel path for both words (i.e. *laranja* and *naranja*) is fragile, since it fails to explain the Portuguese phonetic output. Portuguese and Castilian etymologies are often quite similar, but in this case there is a slight distance to the closeness. Hence, the remote etymology of *laranja* may be solved, but there is no insight into how the word came to be what it is.

2 *Laranja* in Heritage Dictionaries

A survey of heritage dictionaries regarding the word *laranja* and some of its derivatives shows us that they are part of the entry list of them all, and they also evidence a growing descriptive complexity. Cardoso, Barbosa, Pereira and Bluteau all offer Latin equivalents to *laranja*, which seek to be explanatory, not a translation, since oranges were probably unfamiliar in the Roman Empire. Cardoso (1569) uses the generic noun *malum* (which meant fruit in Latin), and the adjective *medicum* to describe *laranja* and close explanations for the derivatives *laranjada* and *laranjâl*. In this entry, he adds the word *arantia(ae)*, unattested elsewhere to our knowledge, but it looks like a Latinized version of the Italian word *arancia*. Another single time, in the entry regarding *malus syria* he presents it as an equivalent to *malus medica*, both of which correspond to the Portuguese *larangeira* 'orange tree'. The reference to Syria invokes what Cardoso probably believed to be the regional origin of the species.

- | | |
|---|--|
| (1) Laranja. Malum medicum. arantia(ae).
Laranjeira. Malus medica. | Malum medicum / Medicum malum. A laranja.
Malus medica / Medica arbor. a laranieyra.
Malus syria. siue malus medica. A laranjeyra. |
| Laranjada. Medicatus(us).
Laranjal. Medicetum(i). | |

Forty years later, Barbosa (1611) almost replicates Cardoso's input. *Malum medicum* is the first Latin equivalent that he provides, but he also offers an alternative, i.e. *mala aurantia*, quoting the Italian Renaissance botanist Mattioli (1501-1577). It is worth noting that Pianigiani (1907) relates *arancio* to a Vulgar Latin *arancia* and *aurantia*, and he identifies a folk etymology link with Latin *aurum* 'gold' (cf. "accostato per etimologia popolare al Lat. AURUM"). Possibly, Mattioli inaugurated the adjustment of the Italian word to a Latin 'motivated' etymon, and Barbosa followed him. On the other hand, the mention to Syria disappears, but a close reference, to the *media regione*, shows up.

3 Cf. *Infopedia*: Do persa *narang*, «laranja», pelo árabe *naranjâ*, «idem»; *Priberam*: (árabe *narandja*).

4 Lorenzo (1968: 220) quotes an anthology of medieval documents where *laranja* occurs, in 1377.

- (2) Laranja. Malum Medicum. Vel, mala aurantia, Mathiolus in lib. 1. Dioscor. c. 131.
 Lorangeyra. Malus Medica. Arbor Medica. Plin. lib. 12. cap. 3. Dicta est Medica a Media regione.
 Laranjal. Locus malis Medicis consitus.

Almost a century later, Pereira (1697) also relates *laranja* to *malum medicum*, and he adds a new Latin equivalent (i.e. *malum aureum*), thus making the metaphorical connection with ‘gold’ even more explicit. His Latin-Portuguese dictionary is yet more prolific. *Malum aureum* is presented as a variant of *malum aurantium* (he may well be just trying to ameliorate the Latin form of the expression), but Cardoso’s *arantia* (characterized as a new word) and Barbosa’s *aurantia* are somehow retrieved. Apparently, they were circling the Italian word and seeking for a plausible Latin etymon, a task doomed to failure that even considered the more bizarre word *anatarantium*, marked by Pereira as never found. Then, Pereira lists a series of *malum* variants that partly refer to prior knowledge (i.e. *assyrium*, *medicum*) and partly bring novelty (i.e. *hesperidum*, *citreum*). *Malum hesperidum* is probably a name influenced by Ferrari’s recent treaty⁵ that located the origin of these fruits in the mythological Garden of the Hesperides. The second innovation brings *citreum* into play, and a whole new series of interrogations raised by the equivalence between *laranja*, *cedromelon* and *cidra* that we will consider below, and *narantrium* and *nerantzium* that look like the Latinization of the Sanskrit etymon. Finally, Pereira accounts for two new meanings for *laranjada*. The first one is related to a strike with oranges, and the second meaning identifies an orange preserve. The amount of information found in Pereira is thus quite considerable, but not proportionally as instructive.

- (3) Laranja. Malum medicum. Malum aureum. Malum (Aurantium) aureum. A laranja. Jun. Arantium, ii, n. g. A laranja. Verbum novum. 1. b. Anatarantium, ii, n. g. A laranja. Non inveni. Aurantium, ii, n. g. A laranja. Amalth. Malum Hesperidum. Assyrium. Medicum. Citreum. A cidra, o limam, a laranja. Jun. *Cedromelon, i, n.g. A laranja, ou cidra. 1.c.2.b.3.l. Onom. M. *Narantrium, ii, n.g. A laranja. 1. l. Nicand. *Nerantzium, ii, n.g. A laranja. 1. b. L. G. B. Medica malus. Cidreira (a lorangeira) (Medicum malum. A cidra, ou laranja, & c.). 1.l.2.b. Virg. Georg. 2.!!

Laranjada. Mali medici ictus.

Laranjada, id est, conserva de laranjas. Mala Medica saccharo cocta.

Laranjal. Locus malis medicis cōsitus.

Bluteau’s dictionary began to be published less than twenty years after Pereira’s, but it is substantially different. Though it also makes use of Latin equivalents, the essence of his *Vocabulario* is practically monolingual, and it is tentatively encyclopedic. In the case of *laranja*, Bluteau (1712-28) resists choosing a Latin equivalent. He claims that *laranja* is a ‘known fruit’, then he invokes Virgil and, mostly, he quotes Ferrari, who gives a list of names and justifications, most of which failed to pass later inquiries:

laranja. Fruto conhecido. Alguns lhe chamaõ Malum aureum. Virgilio diz, Aurea mala. Outros dizem Malum citreum orbiculatum. Sobre os nomes Latinos, ou alatinados, que os Autores dão à laranja, diz o P. Ferrari nas suas Hesperides, liv. 1. pag. 43. Inter acida postremum poma sagacissimi conjectores, recentiore nomine appellant, vel Arantium ab Arantia, pomorum feracissimo Achaiae oppido;

5 G. B. Ferrari (1646) *Hesperides siue de malorum aureorum cultura et vsu Libri quator Io*. Rome: Sumptibus Hermanni Scheus.

quo mala Hesperidum primum Hercules tulisse credebatur, vel Aranium, quasi Ararium, id est, Persicum: est enim Aroa, ut ait Hellanicus, allique, Perfidis regio, vel certe Rantium, tamquam Raedum, hoc est, aeris colore fulvum. Vel Neratio inventore Neratium, vel eum veteri Nicondri Scholiaste Necrantzion, vel (quod etiam Hermulao placet) Narantium à Narantia, quae Ptolomei videtur esse naranga, ex qua idem cum Pausania existimat ab Hercule id ponu fuisse in Graeciam asportatum. Vel demum quia, ut modo diximus, relucet auri colore, aurengium, malum aurantium, unaque expuncta littera Arantium, & aureum malum, quod veteres Hespericum etiam vocavere sed nondum potuit malum aurantium auri quod nomine praefert luce, suos satis demonstrare natales.

Neither Moraes (1813) nor Figueiredo (1913) bring anything new to this discussion. Moraes, who lived in Brazil, describes the fruit and acknowledges a number of local subspecies: sweet oranges or from China; bitter tangerines, or seedless, and some other Brazilian varieties. The reference to Chinese oranges had occurred in a previous dictionary⁶, and it will reappear in Figueiredo, who even mentions *china* as a hyponym of *laranja*. This name for sweet oranges seemingly lasted in Portugal until the beginning of the 20th century, and in fact neither *Infopedia* nor *Priberam* make any mention of it.

Laranja
Fruto da laranjeira.

[...] Variedade de pêra portuguesa.
(Do ár. naranj)
hiperónimo de: china

LARANJA, s. f. Fruta d'arvore de espinho com casca de cor amarella, e gomos dentro: há laranjas doces, ou da China; azedas; Tangerinas, com embigo em baixo; selectas, ou sem caroço, mui doces: a Tangerina doce no Rio de Janeiro é diversa da Tangerina d'outras Colonias, e de sabor mui delicado. §. *Mya* — : peso das pendulas dos relgios de parede. *Mechan. de Marie.*

Figure 1: Moraes (1813) and Figueiredo (1913)

Figueiredo further mentions that *laranja* also refers to a variety of pear, which is probably a mistaken understanding of a variety of oranges that have the shape of a pear, which is commonly found in Brazil. *Infopedia* repeats the same information and it also presents the fruit as an hesperidium, which is an outdated botanical classification.

In sum, so far we have a hint about the remote Sanskrit origin and the vehicular Arabic etymon, and a fluctuating list of Latin equivalents that have no existence in classical Latin. However, the equivalence between *laranja* and *cidra* initially established by Pereira opens a new line of research.

3 *Cidra, Cidrão and Cidreira in Heritage Dictionaries*

Quite unexpectedly, we find out that Cardoso (1569) assigns to *cidra* the Latin equivalent that Pereira (1697) also used for *laranja* (i.e. *malum citreum*). The same applies to Barbosa (1611), who also

⁶ Folqman (1755) “LARANJA, [...] Laranja da china, Malum aureum sinense. § Laranja azeda, Malum aureum acidum”

brings the name *malum Hesperium*, again found in Pereira (1697) for *laranja*, though he only provides the equivalence *malum citreum* for *cidra*. Finally, Bluteau (1712-28) characterizes both *cidra* and *cidrão* as fruits from a tree called *cidreira*, and, as to Latin equivalents, he registers *malum citreum* or *malum medicum*. The contamination between *laranja* and *cidra* becomes evident – semantically, they are quite equivalent.

Cidra may eventually be the word of Latin origin that could not be found for *laranja*. Indeed it has a Latin origin (cf. Cunha, s.v. *cidra*), but what can be found in Latin is not what might be expected. Lewis and Short (1879) claim that the Latin word *citrus* is probably “a mutilation of κέδρος, *cedrus*”, and they assign two meanings to it, which suggests that the naming issue is ancestral:

- I. *The citrus*, an African tree (hence Atlantis silva, Luc. 10, 144, and Massyla robora, Stat. S. 3, 3, 94), whose very fragrant wood (v. citrum) was used in making household furniture, and was prized very highly
- II. *The citron-tree* (also called malus Medica, Persica, etc.), Citrus Medica, Linn., whose fruit and leaves were laid between the folds of clothing to preserve it from worms

In all probability, the trees that the Romans knew as *citrus* were *cedars*, coniferous trees in the *pinaceae* family⁷, and they may have later adopted the same name *citrus* to refer another tree that became a member of the *rutaceae* family. Smith (1859) endorses Fée’s claim⁸ that “for a long period [...], the *citron* was without any specific name among both the Greeks and Romans”.

*CITRUS (κιτρία or κιτρία), the Citron-tree. For a long period, as Fée remarks,⁴ the Citron was without any specific name among both the Greeks and Romans. Theophrastus merely calls it *μηλέα Μηδική ἢ Περσική*. Pliny⁵ styles it the Median or Assyrian Apple-tree, “*Malus Medica sive Assyriaca*.” At a later period, *μηλέα Περσική* became a name appropriated to the Peach-tree, while “*malus Assyriaca*” ceased to be used at all: the designation of the Citron-tree then became more precise, under the appellation of *malus Medica* or *Citrus* (*μηλέα Μηδική, κιτρία*). Of all the species of “*Citrus*,” that which botanists term, *par excellence*, the Citron-tree of Media, was probably the first known in the West. Virgil⁶ gives a beautiful description of it, styling the fruit “*felix malum*.” This epithet *felix* is meant to indicate the “happy” employment of the fruit as a means of cure in cases of poisoning, as well as on other occasions; while the *tristes succi* indicate, according to Fée, the bitter savour of the rind, for it is of the rind that the poet here points out, as he thinks, the medical use: he makes no allusion to the refreshing effects of the citron, but only to its tonic action; and this latter could not refer to the juice, the properties of which were not as yet well known. Some commentators think that, when Josephus speaks of the apple of Persia, which in his time served as “*hadar*,” he means the citron. This, however, cannot be correct. It would seem that he merely refers to a remarkable and choice kind of fruit, which was to be an offering to the Lord; so that *hadar* cannot be ‘the Hebrew for

the citron-tree or its produce.’ Neither is there any ground for the belief that the Jews in the time of Moses were acquainted with this tree.⁷—Virgil⁸ says that the fruit of the citron was a specific against poison, and also that the Medes chewed it as a corrective of fetid breaths, and as a remedy for the asthma. Athenæus⁹ relates a remarkable story of the use of citrons against poison, which he had from a friend of his who was governor of Egypt. This governor had condemned two malefactors to death by the bite of serpents. As they were being led to execution, a person, taking compassion on them, gave them a citron to eat. The consequence of this was, that though they were exposed to the bite of the most venomous serpents, they received no injury. The governor, being surprised at this extraordinary result, inquired of the soldier who guarded them what they had eaten or drunk that day, and being informed that they had only eaten a citron, he ordered that the next day one of them should eat citron and the other not. He who had not tasted the citron died presently after he was bitten; the other remained unhurt!—Palladius¹⁰ seems to have been the first who cultivated the citron with any success in Italy. He has a whole chapter on the subject of this tree. It seems, by his account that the fruit was acrid, which confirms what Theophrastus and Pliny have said of it, that it was not esculent. It may have been meliorated by culture since his time.³

Figure 2: Citrus (Fée 1835)

Both these authors quote Theophrastus⁹, who described the species under the name of *Median or Persian Apple* and lists its properties (a fragrant tree, with thorns and inedible fruits, beard at all seasons, leaves that may be used to keep moths away from linen, a breath freshener, and a remedy for deadly

7 Cf. Smith (1984) “[...] the tree called citrus (a species of cedar or juniper), the wood of which was highly esteemed by the Romans for furniture.”

8 A. L. A. Fée is a French botanist who published *Flore de Virgile, ou Catalogue raisonné des plantes citées dans ses ouvrages*, in 1835.

9 The Greek botanist Theophrastus (3rd century BC) is the author of *Historia plantarum*.

poison ingestion). Theophrastus claimed that the tree grows in Media and Persia, but he describes how to propagate it, which seems to indicate its introduction in the Italic peninsula. A few centuries later, Pliny the Elder¹⁰ is probably the first to acknowledge the term *citrus* that he links to the *nata Assyria malus*. He treats it as a ‘really exotic tree’, as he also instructs on how to plant it¹¹.

Thus, the much later association between *cidra* and *laranja*, drawn both by Pereira (1697) and Bluteau (1712-28) needs to be interpreted under the light of the terminological confusion induced by the introduction of new species, and we may conclude that the understanding of word meanings may require well more than lexicographic research – hopefully these other sources will be part of historical corpora¹².

4 Laranjas and Cidras in Historical Corpora

Data from *Corpus do Português* show that the first attestations for *laranja* occur in the 15th century, and they are quite rare until the 19th century and especially the 20th. Three of the most remote attestations refer the existence of *laranjas* in distant locations. The first one reports an offer of *laranjas* and *cidrões* by the king of Mombaça to Vasco da Gama, during his inaugural trip to India, by sea, in 1498.

Ao domjnguo de rramos mandou o Rey de Mõbaça ao capitam moor hû carneiro e mujtas **laranjas** e cidrões e canas daçuquar (*Diário da viagem de Vasco da Gama*, 1498)

The second attestation reports the existence in San Tomé of ‘orange trees’ (brought from Portugal) that produce ‘oranges’ apparently bigger than expected:

Larangeyras ha muytas e as trouuerom de Portugal e dam fruto **laranja** tam grande como grande çydra de Portugal // Em esta ylha de Sam Thome ha muytas cidras e tammanhas como a barriga de pote de meo almude Limões ha muytos e tam grandes como cidrões em Portugal Limas muytas tam grandes como as cidras de Portugal (*Códice Valentim Fernandes*, 1506-1510)

The third attestation concerns China and the discovery of local sweet oranges:

Ha muitas e muito boas **laranjas**, ha tres generos de **larãjas** doces a quaes milhores, hûas que tem ha casca muito delgada, que quasi sabem a uvas, outras que tem ha casca grossa e crespa tamalaves bicaes mui sabrosas, que lhe comem casca e tudo: outras maiores que as demais que tem ha casca em meo, nem muito grossa nẽ muito delgada: estas sam somenos por serem muito docicadas. (*Enformação das cousas da China*, 1520)

All these attestations value *laranjas* positively – they are a gift from a king, and they are praised for their size and for their taste. Yet the very first attestation brings *laranjas* as ammunition, similar to stones and certainly inedible:

Hûus lançauom pedras, outros **laranjas**, e outros cospiom contra ele (*Crónica de D. Fernando*, 1431-43)

The first attestation of *cidra* also dates from the beginning of the 15th century and occurs in a Portuguese culinary treaty. It is the recipe of a sweet preserve, similar to marmalade:

10 Pliny the Elder (AD 23 – 79) is a Roman naturalist, author of the *Naturalis Historia*.

11 Cf. www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.02.0137%3Abook%3D12%3Achapter%3D7.

12 It should be noted that, in contemporary Portuguese, there are two homophonous words [‘sidrɐ’]: according to *Infopedia*, *cidra* (from the Latin *citrea*) is the name for a fruit that is green and bigger than a lemon, and *sidra* (from the Hebraic *schēchar*) is the name of a beverage obtained from fermented apple juice. French and English equivalents for the Portuguese *sidra* are written with an initial <c> (cf. Fr. *cidre*, Eng. *cider*), though <s> would be more expectable, given the etymon *sicēra*. The graphic contamination can also be documented in Portuguese: *cidra* occurs quite frequently meaning ‘apple cider’.

Pera fazer diacidrão Escolherão muyto bõas cidras & bem feytas que nã Sejam quejmadadas da jeadã nẽ Verdoemgas. (*Tratado de cozinha portuguesa*, 1400?)

Lorenzo (1968: 220) documents *cidra* in the same text where he found *laranja*, some 30 years sooner than the 15th century, and since this is an enumeration, *cidras* and *laranjas* must be different fruits.

Romãas e laranjas e limõoes e çidras

In sum, Lorenzo (1968) locates both words in the final quarter of the 14th century, in the same document and in the same sentence, referring to different entities, and *CdP* allows us to identify two meanings for *laranja* and another closely related meaning for *cidra*, but none of them help us to understand why ‘oranges’ are called *laranjas* in Portuguese.

5 Cross-linguistic Evidence

The very essence of the European Roots project is to facilitate cross-linguistic inquiries through similar monolingual descriptions. The Portuguese word *laranja* is formally close to the Galician *laranxa* and the Basque *laranja*, and they all contrast with the Castilian proximity to the Arabic (Sanskrit) source (cf. Cast. *naranja*). The Italian output *arancia*, like the French *orange*, is also diverse – both these languages lack the initial consonant. These phonetic distinctions are almost microscopic, but they may help to understand how such a variety originated and how they all relate to each other.

The dictionary of the Real Academia Española claims that *naranja* comes from the Hispanic Arabic word *naranġa* and the *Corpus del Español* provides evidence for the word in a document from 1418. Since the Nasrid Kingdom of Granada survived until 1492, direct language contact between Castilian and Arabic is quite plausible until the end of the 15th century. Therefore, the Arabic origin hypothesis for the Castilian word looks like a sound hypothesis, but it sheds no light as to its equivalents in other Romance languages.

The *Trésor de La Langue Française* claims that the French word *orange* was preceded by the compound *pome (d') oreng*, a calque from the Italian *melarancio*¹³, which may explain the absence of the initial consonant in *orange*, if we succeed in understanding why it is absent from Italian (cf. *mel(a)+arancio*). Pianigiani (1907) offers the useful suggestion that a rebracketing of the source word (i.e. the Arabic *narangi*) took place when it entered a southern Italian dialect: the initial consonant [n] was interpreted as an indefinite article. Since Italian and Arabic were in contact, at least in Sicily, this hypothesis also appears to withstand skepticism. Only Portuguese, Galician and Basque words lack a logical hypothesis:

- (4) Sanskrit *nāraṅga*
 - > Arabic *nāraṅġ*
 - > Hispanic Arabic *naranġa*
 - > Castilian *naranja*
 - > Mediterranean Arabic *narangi*
 - > Southern Italian dialect (*u*)n *arancia*
 - > French *orange*
 - ? Portuguese *laranja*
 - ? Galician *laranxa*

13 Many French words of Arabic origin arrived through Italy (cf. It. *zucchero* > Fr. *sucre*). The Portuguese word *açúcar* and the Castilian *azúcar* have the same Arabic origin but, as usual in both these languages, they incorporate the Arabic determiner (i.e. *as-sukkar*).

6 Oranges from Portugal

At this stage, a further cross-linguistic survey may be worth considering. According to Triantafyllidis' dictionary, the Modern Greek equivalent for *orange* is πορτοκαλί [portokalí], a borrowing from an Italian word, *portogallo*. The Modern Greek word, meaning 'sweet orange'¹⁴, is first attested in 1669. Similar names can be found in many other languages, mainly located in Northern Africa and in the southern and western vicinity of the Black Sea, as well as in some Italian dialects:

(5) Arabic	عِلْ اِقْتِرْبَل (alburtuqaliu) ¹⁵
Georgian	ფორთოხალი (p'ort'okhali)
Greek	πορτοκαλί (portokalí)
Macedonian	портокалова (portokalova)
Romanian	portocaliu

So, quite paradoxically, Portuguese has a peculiar word for orange, and the name for orange in a number of languages evokes the name of Portugal. On Reddit, we can find at least two discussions about orange names. The first one¹⁶, entitled "The word for the fruit orange in various European languages" presents a colorful map (cf. Figure 3) with the distribution of orange names according to their linguistic origin. It certainly contains inaccurate details (the discussion reveals several discrepancies with speakers from many of these languages), but it helps to visualize the geographic distribution of these etymological word families. In Central and Western Europe, the Sanskrit/Arabic word dominates, either in its simple form (cf. Castilian *naranja*, Italian *arancia*, Portuguese *laranja*, French *orange*) or in a compound preceded by a descendant of the Latin word *pomus* (cf. Polish *pomarańczowy*). Then Germany, Northern Europe, Eastern Europe and Asia typically adopt a compound formed by *Apfel* (the German translation of *pomus*) and a modifier that related to China (cf. German *Apfelsine*). Finally, North Africa, Greece and some neighboring languages choose words related to Portugal (cf. 5). The second map¹⁷ draws a hypothesis for the spreading of the words for orange (cf. Figure 4).



Figure 3: The word for the fruit orange in various European languages.

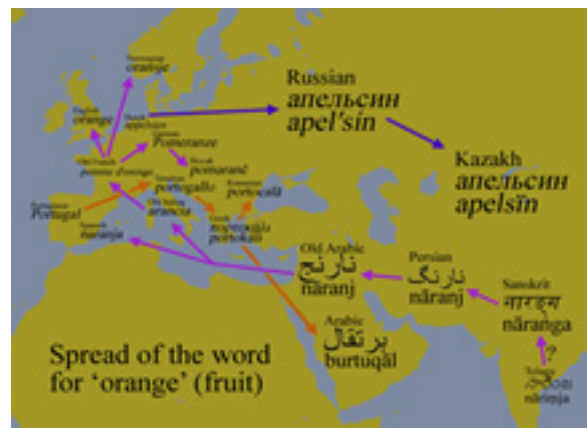


Figure 4: Spread of the word for the fruit orange.

None of these maps help to understand the specific output of the Portuguese word (i.e. *laranja*), but they do help to set a path for the dissemination of the word related to Portugal. According to the second map, the loan started in Venetian and travelled south in Italy, reached Greek and from there

14 Simeon Tsoladikis, whom again I wish to thank, conveyed this information.

15 Just for sweet oranges. Sour oranges are called *nāraṅg*.

16 Accessed at www.reddit.com/r/linguistics/comments/1cmhsv/the_word_for_the_fruit_orange_in_various_european/ [28/03/2018].

17 Accessed at www.reddit.com/r/etymology/comments/5mrfao/spread_of_the_word_for_orange_the_fruit_oc_963_733/ [28/03/2018].

it spread to Romanian and to Arabic. Apart from the fact that *Portugal* is not the name for *orange* in Portuguese, as this map suggests, it is also questionable to set a dissemination route based on no evidence. Is it plausible that a Venetian name for orange (i.e. *portogallo*) moved into Greek and from Greek to Arabic, Macedonian, Romanian, Turkish, Georgian and maybe some more languages or dialects in close-by regions? Though no references are provided, Wikipedia's Italian entry on *citrus sinensis*¹⁸ gathers some dialectal information that is displayed in the following diagram:



Figure 5: Orange in some Italian dialects.

Though I am not a specialist on Italian dialectology and cannot ascertain the reliability of these data, it is interesting to note that words related to Portugal occur in a larger number of dialectal spots, but not near Venice (as hypothesized in Figure 5) nor in Sicily. Though the history is still tentative, Dugo and Di Giacomo (2002: 8-9) present interesting information. They remember that the Arabs occupied the Mediterranean territories left free by the Romans and have introduced into these new plants and agricultural techniques. Sour oranges were used to adorn mosques and gardens, for their scent. Then, the same authors refer that “the first mention of the common or sweet orange was found in a historical book by Hugo Falcano, who lived in Sicily from 1154 to 1169” and he also mentions that “the term ‘arangias’ was used by Blondus Flavius in a description (thirteenth century) of the citrus in Amalfi and Naples”. Dugo and Giacomo (2002:9) also state that “Nicolò Speciale (15th century), who wrote a book on the siege of Palermo, reported that sour orange was grown in Sicily and its fruits were called ‘arangias’ by the Sicilians”. It is quite tempting to establish a link between the Sicilian word *arangia* and the Portuguese *laranja*, particularly if we admit the hypothesis of a methanalysis process intervening over the sequence *l’arangia*. After all, other cases of methanalysis are documented for fruit names in Portuguese¹⁹ (cf. Lat. (*pruna*) *damascea* > *d’amascea* > Pt. *ameixa* ‘prune’, by the

18 Cf. it.wikipedia.org/wiki/Citrus_sinensis.

19 Cf. A.Nascentesarchive.org/stream/DICIONARIOETIMOLOGICORESUMIDODALINGUAPORTUGUESAANTENORNASCENTES/DICION%C3%81RIO%20ETIMOL%C3%93GICO%20RESUMIDO%20DA%20LINGUA%20PORTUGUESA%20%20ANTENOR%20NASCENTES_djvu.txt.

deglutination of what appeared to be a preposition; Lat. *prunum* ‘plum’ > Pt. *abrunho*, by the agglutination of a putative definite article). So, though we do not have clear evidence of the Sicilian origin of the Portuguese word, we do have an indirect hint about this immediate etymological link, probably made possible by means of commercial relationships between Portuguese and Sicilian merchants before the 14th century:

- (6) Sanskrit *nāraṅga*
 > Arabic *nāraṅġ*
 > Mediterranean Arabic *narangi*
 > Sicilian (*u*)*n arangia*
 > Portuguese *laranja*

Finally, we need to take into account the contribution of a Portuguese agronomic researcher regarding the introduction of citrus plants in Portugal. Ferrão (1992: 167) claims that citrus coming from the Orient arrived in Europe via the Mediterranean Sea, brought by the Arabs. Consequently, citrus plants reached the Iberian Peninsula well before Portugal existed as a country, and penetrated the territory into the north as far as the ecology allowed them. The same author also claims that the oranges of such origin were not necessarily bitter – oranges that grow in milder climates tend to be sweeter than those that manage to survive in less favorable conditions. Algarve was, and still is, a very propitious environment for the cultivation of sweet oranges. On the other hand, still according to Ferrão (1992: 167), the Arabs spread citrus species down the African Western Coast and this is why the Portuguese navigators found oranges in Gambia, and, as we have seen above, Vasco da Gama also found oranges along the East African Coast in 1498. So apparently the introduction of sweet oranges in Europe by the Portuguese navigators who had brought them from China lacks historical support.

Ferrão (1992) considers that the oranges the Portuguese navigators found in India and China were sweeter than those they knew in Portugal, because of the ecology of these different locations. But they nevertheless brought them to Portugal, grew them, and started large-scale export to Europe, advertising them as a new cultivar. Italy was probably a good market, since Portugal used to buy and sell several goods to Italian cities, such as sugar to Genoa and Florence.

7 Conclusion

Words that we know and use have a shared existence – shared by all those that also know and use them or have done so in the past. However, shared existence does not always entail a shared knowledge. Words that are the legacy of each generation to the following may undergo a change in the process, and that change often occurs without explicit notice.

Dictionaries may help to trace semantic changes, because they document words, but dictionary-making processes are not primarily concerned with that aim. Therefore, dictionaries may give us some information on past and present word meanings, but this is generally not enough to fully understand them, and it may even be inaccurate. The body of all sorts of texts that human ancestry has produced until now, if preserved and incorporated in resourceful databases, is crucial to complement the information that heritage dictionaries offer us, and to validate the latter’s contents, or not.

However, even if textual databases, here including lexicographic databases, were as comprehensive as we would like them to be, it would still be necessary to manage the output results of any given search, and the amount of information thus obtained may be, as we have seen in the case of *laranja*, quite difficult to manage. The ancestry of oranges and orange trees in Europe is rather challenging to trace, because the words that were and are used to refer to them conceal more than what they show. In

this paper, I have collected the information that is available in heritage and contemporary dictionaries and in historical corpora as well as other relevant sources, which amounts to a large set of data that still fails to provide convincing explanations for all our questions.

This case study helps to demonstrate that we need to develop a sophisticated descriptive protocol that will allow us to precisely identify the relevant data and to be able to relate what we find productively. The European Roots prototype aims to respond to those needs, as it considers essential evidence and discards negligible information, but it is particularly useful in that it allows cross-linguistic analysis, since all languages may use the same template.

References

- Barbosa, A. (1611) *Dictionarium Lusitanico Latinum*. Bracharae : typis, & expensis Fructuosi Laurentij de Basto. Accessed at: clp.dlc.ua.pt [28/03/2018].
- Cardoso, J. (1569-1570) *Dictionarium Latinolusitani cum & Vice Versa Lusitanico Latinum*. Conimbricae: Joan Barrerius. Accessed at: clp.dlc.ua.pt [28/03/2018].
- Bluteau, R. (1712-1728) *Vocabulario Portuguez e Latino*. Coimbra: Collegio das Artes da Companhia de Jesu. Accessed at: clp.dlc.ua.pt/DICIweb/default.asp?url=Home [28/03/2018].
- CdP = *Corpus do Português*: 45 million words, 1300s-1900s. Accessed at: www.corpusdoportugues.org/x.asp [28/03/2018].
- CdE = *Corpus del Español*. Accessed at: www.corpusdelespanol.org/x.asp [28/03/2018].
- CLP = *Corpus Lexicográfico do Português*. Accessed at: clp.dlc.ua.pt [28/03/2018].
- Corominas, J., Pascual, J. A. (1981). *Diccionario Crítico Etimológico Castellano e Histórico*. Madrid: Gredos.
- CRPC = *Corpus de Referência do Português Contemporâneo*. Accessed at: www.clul.ul.pt/pt/recursos/183-crpc#-qp [28/03/2018].
- Cunha, A. G. (1994) *Dicionário Etimológico Nova Fronteira da Língua Portuguesa*. Rio de Janeiro: Nova Fronteira. 2ª edição.
- Fée, A. L. A. (1835) *Flore de Virgile, ou Catalogue raisonné des plantes citées dans ses ouvrages*. Bibliothèque classique latine ou Collection des auteurs classiques latins. Available at: gallica.bnf.fr/ark:/12148/bpt6k298102/f108.item.r=citrus [28/02/2018].
- Figueiredo, C. de (1913) *Novo Dicionário da Língua Portuguesa*. Porto, Typ. da Empr. Litter. e Typographyca Accessed at: *Dicionário Aberto*
- Infopédia = *Dicionário da Língua Portuguesa da Porto Editora*. Accessed at: www.infopedia.pt/linguaportuguesa/ [28/03/2018].
- Lewis, C. T. & C. Short (1879) *A Latin Dictionary*. Founded on Andrews' edition of Freund's Latin dictionary. Revised, enlarged, and in great part rewritten edition. Oxford: Clarendon Press. Accessed at: www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0059 [28/03/2018].
- Lorenzo, R. (1968) *Sobre Cronologia do Vocabulário Galego-Português (Anotações ao 'Dicionário Etimológico' de José Pedro Machado)*. Vigo: Editorial Galaxia.
- Mabberley, D. J. (2004) Citrus (Rutaceae): A Review of Recent Advances in Etymology, Systematics and Medical Applications. *Blumea* 49: 481–498. Accessed at: repository.naturalis.nl/document/565086 [28/03/2018].
- Moraes = Moraes Silva, A. (1789) *Diccionario da lingua portugueza*. Lisboa, na Of. de Simão Thaddeo Ferreira. 2nd edition (1813). Accessed at: www.brasiliana.usp.br/pt-br/dicionario/. [28/03/2018].
- Nebrija, E. A. (1492) *Lexicon hoc est Dictionarium ex sermone latino in hispaniensem*. Salamanca.
- Pereira, B. (1697) *Prosodia in Vocabularium Bilingue, Latinum et Lusitanum*. Eborae: Typographia Academiae Raskin. Accessed at: clp.dlc.ua.pt [28/03/2018].
- Pianigiani, O. (1907) *Vocabolario Etimologico della Lingua Italiana*. Florence. Accessed at: www.etimo.it/?pag=hom [28/03/2018].
- Priberam = (2008-2013) *Dicionário Priberam da Língua Portuguesa*. Accessed at: www.priberam.pt/dlpo/chave [28/03/2018].
- Real Academia Española, *Diccionario de la Lengua Española*. Accessed at: dle.rae.es/?id=DgIqVCc [28/03/2018].

- Silvestre, J. P. & A. Villalva (2014) A Morphological Historical Root Dictionary for Portuguese. Proceedings of the XVI EURALEX International Congress: The User in Focus, 967-978. Accessed at http://www.euralex.org/elx_proceedings/Euralex2014/euralex_2014_074_p_967.pdf [29/03/2018].
- Smith, W. (1854) *Dictionary of Greek and Roman Geography*, illustrated by numerous engravings on wood. London: Walton and Maberly.
- Smith, W. (1859) *A Dictionary of Greek and Roman Antiquities*. Little, Brown, and Co. Available at: archive.org/details/adictionarygree05smitgoog [28/2/2018].
- TLFI = *Trésor de la langue Française informatisé*. ATILF - CNRS & Université de Lorraine. Accessed at: www.atilf.fr/tlfi [28/03/2018].
- Triantafylidis, M. *Dictionary of Standard Modern Greek*. Accessed at www.greek-language.gr/greekLang/modern_greek/tools/lexica/triantafyllides/search.html?lq=%CF%80%CE%BF%CF%81%CF%84%CE%B-F%CE%BA%CE%AC%CE%BB%CE%B9&dq= [28/03/2018].
- Villalva, A. & J. P. Silvestre (2015) Filling gaps in dictionary typologies: ROOTS – a morphological historical root dictionary. Silvestre, J. P. & A. Villalva, eds. (2015) *Planning non-existent dictionaries*. Centro de Linguística da Universidade de Lisboa & Universidade de Aveiro.