Université de Lille, France
École Doctorale Sciences de l'Homme et de la Société, Lille Nord de France
UMR 8163 | Savoirs, Textes, Langage

TARK, *SHPERLACK*, *BURFIP*, AND OTHER ALIEN BAD WORDS: INVESTIGATING A SOUND-MEANING ASSOCIATION IN ENGLISH AND FRENCH SWEAR WORDS

TARK, *SHPERLACK*, *BURFIP*, ET AUTRES GROS MOTS ALIENS : ÉTUDE D'UNE ASSOCIATION SON-SENS DANS LES GROS MOTS DE L'ANGLAIS ET DU FRANÇAIS

Robin VALLERY

Thèse soutenue le 13 décembre 2024 en vue de l'obtention du grade de
Docteur en Langues et Littératures Anglaises et Anglo-Saxonnes

COMPOSITION DU JURY

Maarten LEMMENS (PR, Univ. de Lille, FR)	directeur
Paolo MAIRANO (MCF, Univ. de Lille, FR)	co-encadrant
Denis JAMET-COUPÉ (PR, Univ. de Lyon, FR)	rapporteur
Anne PRZEWOZNY-DESRIAUX (PR, Univ. de Toulouse, FR)	présidente du jury, rapporteur
Benjamin BERGEN (PR, Univ. de Californie San Diego, USA)	examinateur
Maria FLAKSMAN (MCF, Univ. de Bamberg, DE)	examinatrice
Maïa PONSONNET (CR HDR, Univ. de Lyon, FR)	examinatrice

“This is stressful! [laughter]

[…]

[laughter] Now, it was fun!”

The first English-speaking participant
in our alien language experiment, 2021

Acknowledgements

To my supervisors Maarten Lemmens and Paolo Mairano, I would like to thank you for this five-year journey. Despite the numerous hardships of Covid, constant overwork at an understaffed university, and the consequences that those had on our ability to work efficiently on this thesis, this piece of linguistic research is now arriving at a safe harbour. Thank you for the patience, kind words, constant support, and your complementary approaches to my work.

For Maarten in particular, I would like to take this as an opportunity to tell you how importantly you contributed to making me the person I am today. When I was an undergraduate student here in Lille, I was unsure whether I wanted to become a linguist or translator. I was good at both linguistics and translation, and could not really make up my mind. I enjoyed linguistics classes much, but after I followed your undergraduate optional courses, I thought something new: "*This* is the kind of linguistic research I want to make." Here is the result, nine years later. The passion you put into teaching to students and in guiding budding researchers like me makes all the difference, and brings much solace. I felt reenergised and more confident in the future after every one of our meetings. Please never stop caring.

For my co-supervisor Paolo, thank you for always keeping your door open, for answering me as quickly and efficiently as you could despite how busy you were, and for the energetic and honest conversations we had about my work.

I also would like to thank the members of the jury Benjamin Bergen, Maria Flaskman, Denis Jamet-Coupé, Maïa Ponsonnet, and Anne

Przewozny-Desriaux, for the time you have taken to evaluate this, and for the precious feedback that you will provide and that will improve this research further.

I would like to also thank the people at the STL lab for the guidance, feedback, and support they gave throughout these five years of teaching and researching, in particular to Caroline Bouzon, Giuditta Caliendo, Bert Cappelle, Ilse Depraetere, Samantha Laporte, Chris Pinon. Caroline, I thank you for allowing me to teach phonetics, for helping me to do so as best you could, and for lending an ear to my numerous questions and suggestions. Bert and Ilse were particularly helpful as members of my Comité de suivi de thèse: our meetings allowed me to identify issues and work on them. Samantha, your great way of teaching is an inspiration, and your excellent course materials made my life teaching grammar so much easier. I would also like to thank Cédric Patin and Mai Tran for allowing me to teach for the first time in my life, in the best possible conditions under dire Covid circumstances. I enjoyed teaching to dedicated speech therapy students a lot, and hope to be able to do it again in the future. I would like to thank the administrative team who helped me so many times to solve issues related to classes: Jean-Baptiste Bonjean Du Manoir, Cora Darem, Kévyn Martin, and Odile Thieffry at the Pont de Bois campus, and Charlotte Abeele for the Speech Therapy department at the Loos campus. To lecturer Keith Dowling, I thank you from the bottom of my heart for allowing me to recruit so many English-speaking participants in my experiment. This gratitude applies to any participant in my experiment, and to anyone who ever helped me recruit participants.

At the STL lab, a special thank you is warranted for Justin Duquesnoy who helped so many of us PhD students. Speaking of which, Covid

showed all of us how crucial it is to see colleagues in-person, in order to interact, bounce back on each other's ideas, encourage each other or just have some plain fun during a break. For this I would like to thank other PhD students at the lab: Adélie Buclier, Rou Chen, Mahamadou Doumbia, Océane Foubert, Huiyun Hu, Oksana Ivchenko, Lucie Latouche, Mégane Lesuisse, Lola Marinato, Diana Oliveira Santos, Christina Piot, Eline Put, Capucine Saulpic, Houssein Taouz, and Alina Tsikulina. A special word is deserved by Océane and Lola, who have been linguistics comrades and friends since even before I started my PhD, and by Capucine, whose conversations often brought light, understanding, and humour even in moments of kafkaesque despair. I would like to say how grateful I am to other researchers I met at conferences, in particular Ian Joo, Melanie Keller, Catherine Laliberté, Robbie Love, Richard McKay, Yuze Sha, Thomas Schlatter, Hans-Jörg Schmid, Richard Stephens, and Diana Wengler. Your enthusiasm, constructive feedback, and conversations on taboo and swear words showed me the value of my work.

To my close friends and family, thank you for your undying support. This thesis would not exist if not for your presence and help. My gratitude towards you is beyond words, and for that reason, I end this text here.

Abstract in English

Tark, *shperlack*, *burfip*, and other alien bad words: Investigating a sound-meaning association in English and French swear words

Swear words, also known as taboo words, profanities, bad words, etc. are words that are socially forbidden because they are considered extremely impolite or gravely insulting. Why are swear words forbidden and not other words? Part of the explanation is that they have meanings related to taboo semantic domains, like religion, sexuality, or body waste (Bergen 2016a: 12–39); but then the next question is: why are their (near-)synonyms not forbidden, e.g., in English, why is *prick* a swear word and not *penis* (Ljung 2011: viii), why is *shit* a swear word and not *excrement* or *stool* (Bergen 2016a: 14–15)? Some researchers hypothesise (Hughes 2006: 343) or propose based on empirical data (Yardy 2010; Bergen 2016a : 52–63; Lev-Ari & McKay 2022; Chiang & Schlatter, ms.) that sounds influence which words are forbidden, because swear words tend to contain some specific phonemes. We can explain this in terms of sound symbolism, the notion that sounds can be associated with meanings (Dingemanse et al. 2015; Haiman 2018: 118–119; Sidhu 2019), or to put it differently, that sounds can be involved in unconscious form-meaning pairings, unconscious constructions in the sense of Construction Grammar (Goldberg 2006).

In this thesis, we review what the literature says on swear words and sound symbolism. We then describe three empirical studies conducted on English and French: one study on swear words of English and French, one on fictional swear words in the same two languages, and one on experimental swear words – we asked native speakers of English and French to

spontaneously invent words from alien, i.e., extra-terrestrial languages as in science-fiction works. Our results suggest the existence of a sound-meaning association between the social and emotional meaning of swear words (Finkelstein 2018: 311, 326) and the least sonorous consonants according to Parker's (2008) sonority scale: plosives (/p/, /t/, /k/, /b/, /d/, and /g/), voiceless fricatives (/f/, /θ/, /s/, /ʃ/, and /h/), and affricates (/tʃ/ and /dʒ/). Our data on French swear words also confirm that they have a more social and emotional meaning, like "violation of hearer's space" or "taboo-breaking".

Finally, we offer a theoretical discussion of what our sound-meaning association and others mean for meaning-making in language. In particular, we propose that even though it is a polemic debate, the classical tenets of linguistics of double articulation (Dingemanse et al. 2015; Martinet 1957) and arbitrariness (Saussure (2005 [1916]) are actually not incompatible with sound symbolism. A reanalysis of Martinet's double articulation and Saussure's arbitrariness suggests that the latter should be redefined more specifically than originally conceived by Saussure, and that arbitrariness thus does not entail double articulation. This reassessment also allows to distinguish between motivated and non-motivated sound-meaning associations. It helps us better understand how speakers can give meaning to sounds, i.e., how sounds can be involved in unconscious sound-meaning pairings that influence language deeply – like the interdiction of swear words – and are not just limited to poetry and other language games.

Keywords: swear words, swearing, bad words, profanities, curses, taboo, English, French, sound symbolism, iconicity, motivation, double articulation, arbitrariness, arbitrariness of the linguistic sign

Résumé en français

Tark, *shperlack*, *burfip*, et autres gros mots aliens : étude d'une association son-sens dans les gros mots de l'anglais et du français

Les gros mots, aussi connus sous le nom de jurons, mots tabous, ou en anglais *profanities*, *bad words*, etc. sont des mots socialement interdits car considérés comme extrêmement impolis ou gravement insultants. Pourquoi les gros mots sont-ils interdits et pas d'autres mots ? Une partie de l'explication vient de leurs sens liés à des domaines sémantiques tabous, comme la religion, la sexualité, ou les déchets du corps humain (Bergen 2016a : 12–39) ; s'ensuit alors la question : pourquoi leurs (quasi-)synonymes ne sont-ils pas interdits, par exemple en anglais, pourquoi *prick* est-il un gros mot et pas *penis* (Ljung 2011 : viii), pourquoi *shit* est-il un gros mot et pas *excrement* ou *stool* (Bergen 2016a : 14–15) ? Certains font l'hypothèse (Hughes 2006: 343) ou proposent à partir de données empiriques (Yardy 2010 ; Bergen 2016a : 52–63 ; Lev-Ari & McKay 2022 ; Chiang & Schlatter, ms.) que les sons influencent quels mots sont interdits, car les gros mots ont tendance à contenir des phonèmes spécifiques. On peut l'expliquer par le phonosymbolisme, l'idée selon laquelle les sons peuvent être associés à des sens (Dingemanse et al. 2015; Haiman 2018: 118–119; Sidhu 2019), ou pour le dire autrement, que les sons peuvent être impliqués dans des associations forme-sens inconscientes, des constructions inconscientes au sens de la Grammaire des Constructions (Goldberg 2006).

Dans cette thèse, nous passons en revue ce que dit la littérature scientifique sur les gros mots et le phonosymbolisme. Nous décrivons ensuite trois études empiriques sur l'anglais et le français : une étude sur les gros mots de l'anglais et du français, une sur les gros mots fictionnels de l'anglais et du

français, et une sur des gros mots expérimentaux – nous avons demandé à des locuteurs et locutrices dont la langue première est l'anglais ou le français d'inventer spontanément des mots aliens, c'est-à-dire de langues extraterrestres comme dans une œuvre de science-fiction. Nos résultats suggèrent l'existence d'une association son-sens entre le sens social et émotionnel des gros mots (Finkelstein 2018 : 311, 326) et les consonnes les moins sonores selon l'échelle de sonorité de Parker (2008) : les occlusives (/p/, /t/, /k/, /b/, /d/, et /g/), les fricatives sourdes (/f/, /θ/, /s/, /ʃ/, et /h/), et les affriquées (/tʃ/ et /dʒ/). Nos données sur les gros mots français confirment également qu'ils ont un sens plus social et émotionnel, comme "intrusion dans l'espace personnel de l'interlocuteur ou interlocutrice", ou "rupture du tabou".

Enfin, nous offrons une discussion théorique de ce que notre association son-sens et d'autres impliquent pour la façon dont la langue transmet du sens. En particulier, nous proposons que, bien qu'il s'agit d'un débat polémique, les principes classiques que sont la double articulation (Dingemanse et al. 2015 ; Martinet 1957) et l'arbitraire du signe (Saussure (2005 [1916]) ne sont en réalité pas incompatible avec le phonosymbolisme. Une réanalyse de la double articulation de Martinet et l'arbitraire de Saussure suggère que ce dernier devrait être redéfini plus spécifiquement qu'originellement conçu par Saussure, et que l'arbitraire n'implique pas logiquement la double articulation. Cette réévaluation permet également de distinguer entre les associations son-sens motivées et non-motivées. Elle nous aide à mieux comprendre comment les locuteurs et locutrices peuvent donner sens aux sons, autrement dit, comment les sons peuvent être impliqués dans des associations sons-sens inconscientes qui influencent profondément la langue – comme dans le cas de l'interdiction des gros mots – et ne sont pas

limitées à la poésie et autres jeux de langage.

Mots-clés : gros mots, jurons, tabou, anglais, français, phonosymbolisme, iconicité, motivation, double articulation, arbitraire du signe

Table of Contents

Chapter 1. Introduction

Few language practices seem as self-evident to speakers as the ban on swear words. To many of us, swear words are bad, shocking, or scandalous. They make us uncomfortable. So they are banned, and people often think they should be. We have an acute awareness that other speakers feel that way about swear words, even when we belong to the more relaxed kind of people. Even among the latter, the most common thing to do with swear words, apparently, is to tolerate them: we may feel that it is okay to swear, among friends typically, or with an amused and complicit audience, but not too much. Otherwise, you will be considered vulgar. You can use swear words, but you use swear words at your own risk.

During our language acquisition years, we learn how to pronounce sequences of sounds that are called "words" and how some of these words are *not* to be used, called "swear words" or "bad words", "curses", "profanities", etc. At the same time, we see that adults are actually allowed to manipulate these risky items.

The inquisitive mind of a naive child may then wonder: why? Why are swear words swear words? If some words exist, why is it forbidden to use them? Why is it more okay for adults to use them than for children? As with many "why" questions from children, the answers from adults may not be entirely satisfactory: some swear words mean really bad things; it is really impolite to use them because they are dirty, shocking, or scandalous and this is how society sees them. The child may wonder why they are dirty, shocking, or scandalous. The answer, again, might not be so satisfactory, *viz.* that this is just

the way it is; one should not use those words, end of story.

Of course, as linguists we cannot be satisfied with such answers. This thesis is an attempt to find an answer to what swear words are, how speakers perceive them, and some of the reasons why they may be banned.

Such interdictions are often called taboos. We use the term here in a rather everyday sense, to mean a social interdiction. Murder, theft, or walking around naked in public, are definitely taboo, and for that reason happen to be illegal too. The understanding of taboo as a social interdiction includes interdictions that are less strict and not enforced legally, for example the relative interdiction to talk about certain subjects like sexuality or body waste; these are considered taboo topics. In the same way, uttering swear words is not illegal – except possibly on specific media – but is certainly discouraged, frowned upon, and considered shocking in many contexts. Swear words can thus be considered a kind of taboo words.

Why are swear words taboo? We will answer that question in more detail in our review of the linguistic literature on that very subject in Chapter 2. By way of introduction, we can say that many swear words have been banned in part because they refer to concepts belonging to taboo semantic domains, i.e., religion, sexuality, or body waste (Stapleton et al. 2022: 2; Bergen 2016a: 12–39). Some, but not all speakers of English (Bergen 2016a: 16) feel like swear words also include another kind of taboo words, namely *slurs*, i.e., bigoted insults against a specific social group, for example the anti-Latin-American slur *spic*, or the anti-gay slur *fag*. For the sake of simplicity, we will use *swear words* in this thesis in the largest possible sense in English, including slurs.

If swear words refer to taboo topics, and this explains why they are banned, then why are their synonyms not banned? To give an example cited in

Ljung (2011: viii), why is *prick* a swear word, but not *penis*? Or as Bergen (2016a: 14–15) points out, why is *shit* a swear word, but not *excrement* or *stool*? As we will discuss further in Chapter 2, there are socio-historical mechanisms at work here, namely social distinction and religious taboos. The process for a word to become a swear word seems to be a long, multifactorial process. Some of this is clearly semantic (or pragmatic), but it has also been proposed that one of the factors at play here is the presence or absence of specific phonemes in a word. In other words, there seems to be a way to sound "swearish", so to speak, an identifiable sound pattern in swear words that makes some words more likely candidates for that complex historical process. More specifically, Bergen (2016a: 49–64) notices that English swear words tend to be closed monosyllabics, and they tend to end not just with any consonants, but with plosives (2016a: 64). Haiman (2018: 209–212) suggests that plosives are iconically appropriate to express "familiarity" or "violation of hearer's space" – a meaning that is relevant for swear words. However, plosives may not be the only consonants that are overrepresented among swear words. Yardy (2010: 12–20, 71–78) argues that cross-linguistically, swear words are likely to contain more of the least sonorous consonants – i.e., least vowel-like consonants, including plosives /p/, /b/, /t/, /d/, /k/, or /g/, the affricates /tʃ/ and /dʒ/, and the voiceless fricatives /f/, /θ/, /s/, /ʃ/, /h/. To put it differently, all obstruents except voiced fricatives are more frequent in swear words according to Yardy (ibid.). Other hypotheses have been proposed by, for example, Lev-Ari & McKay (2022) or Chiang & Schlatter (ms.), but overall these are coherent with Yardy's proposal concerning unsonority. More details on these hypotheses on sound tendencies in swear words are discussed in Chapter 2.

How likely is it that sounds contribute to the taboo status of a word?

Unconscious and probabilistic sound-meaning associations have been found to exist in the lexicon of languages, often following motivated patterns, such as iconic patterns based on resemblance (discussed in more detail in Chapter 2, Section 2.1.3). If such an unconscious sound-meaning association creates a sound tendency in swear words, the question is what that meaning could be. It has to be a meaning general enough to be shared by all swear words. The literature suggests a variety of such general but context-dependent meanings, like familiarity, violation of hearer's space (Haiman 2018: 209–212), hostility or aggression (Yardy 2010: 12–20, 71–78; Chiang & Schlatter ms.), offence (Lev-Ari & McKay 2022: 1106, 1109, 1111), negative emotions or catharsis (Chiang & Schlatter ms.). These insights lead us to our three main research questions:

1. Is there a tendency for swear words to be closed monosyllabics, or to contain more unsonorous consonants?
2. (a) Does the tendency correspond to a cognitively real, unconscious form-meaning pattern for swear words? (b) If so, what is this meaning?
3. If such form-meaning pairings exist, what are the theoretical implications for the (generally accepted) Saussurian idea that the linguistic form (of words) is arbitrary, and for the (generally accepted) idea that phonemes are meaningless?

To answer the first research questions, we have set up three different case studies, where we look at the presence of unsonorous consonants in swear words in English and French based on data collected from dictionary lists, literary works and questionnaires. Our case studies confirm that there is indeed a sound-meaning association for swear words in the lexicon of the two languages (RQ1). Moreover, our case studies show that speakers unconsciously

exploit these patterns when they list or create fictional swear words (RQ2).

Our case studies strongly suggest that speakers exploit the sound-meaning pattern which is thus cognitively real. The question then is what the meaning could be that is associated with the unsonorous consonants in swear words (RQ2b). While this remains a difficult question to answer, we suggest two possibilities: (i) the specific meaning of swear words is a "violation of hearer's space", i.e.., "familiarity" or "aggression" depending on the context, as suggested by Haiman (2018: 209–212) or (ii) it is an emotional and contextual meaning of breaking a taboo, which corresponds to the intrinsic nature of swear words. These two possibilities will be discussed in more detail in Chapter 4.

The existence of this kind of sound-meaning associations runs counter to two classical tenets of linguistics, an issue addressed in our third research question. The first tenet is duality of patterning (Dingemanse 2015: 604), also known as double articulation, initially proposed by Martinet (1957; 2008 [1960]: 37–44). Double articulation posits that sounds do not have meanings. In our discussion in Chapter 4, we will clarify this (apparent) contradiction and consider to what extent the two can still be reconciled. Our argument will rest on the distinction between deterministic form-meaning pairings versus probabilistic sound-meaning associations.

The second classical tenet is the arbitrariness of linguistic signs, proposed by Saussure (2005 [1916]) i.e., the idea that the link between words and their meanings is random, unmotivated, unpredictable (ibid.: 100–102, 155–157). Saussure's arbitrariness and Martinet's double articulation are closely related notions, which is why in the literature the two are not always explicitly distinguished. Against the background of the sound associations found in swear words, we will argue however, that arbitrariness and double articulation

are fundamentally distinct: the latter entails the former, but the former, defined in a more coherent and precise way, does not necessarily entail the latter. Like double articulation, arbitrariness was originally defined with a focus on a specific type of meaning (referential), conveyed in a specific way (compositional, context-independent) which ignores subtle probabilistic sound-meaning associations as those given by our study. As we will see, this conceptual distinction between arbitrariness and double articulation has implications for the theory on sound-meaning associations, how we should use the terminology to refer to them, and for the discussion on the sound-meaning pattern we identify in English and French swear words.

The thesis is structured as follows. In the next chapter (Chapter 2), we first review the existing literature. This allows us to get a first idea of what swear words are, to see what sound tendencies have been proposed for swear words, and to understand how such sound-meaning tendencies function. This allows us to arrive at our first two research questions, i.e., whether there is a tendency for English and French swear words to be closed monosyllabics or to contain unsonorous consonants (RQ1), whether it corresponds to a cognitively real sound-meaning pattern (RQ2a), and what the meaning in that pattern might be (RQ2b). In the second part of the same chapter, we discuss our theoretical framework. Namely, we argue for our own more complete definition of swear words, based on the literature we reviewed and our own reflection. This definition partly informs the methodology of our empirical studies, and allows to determine what kind of meaning swear words have, in relation to one of our research questions (RQ2b).

In Chapter 3, we describe the methodology, data, and results of our three empirical studies, conducted to answer the three aforementioned research

questions. This first study analyses real-life swear words of English and French to determine if they tend to be closed monosyllabics, to contain unsonorous consonants, or neither (RQ1). Swear words tend to contain unsonorous consonants so the next studies test for that tendency only. This part describes the method we follow in order to get an empirically collected list of French swear words. We also give data that is relevant to another research question on the nature and hence the meaning of swear words (RQ2b). These next two studies are concerned with our second research question (RQ2a), i.e., whether the statistical tendency identified in the first study corresponds to a cognitively real sound-meaning pattern. The second study analyses fictional swear words of English and French, while the third study analyses swear words invented experimentally by English speakers and French speakers.

In Chapter 4, we discuss our results and how they answer RQ1 and RQ2a. We elaborate on how our data allows to answer RQ2b on the meaning of swear words. Based on the literature review and our own findings, we also answer RQ3 on the implications that such sound-meaning associations have for the notions of arbitrariness and double articulation.

Finally in Chapter 5, we provide a summary of our findings and contribution, and consider possible avenues for further research.

Chapter 2. Literature review and theoretical framework

This chapter presents the groundwork for our case studies. In the first part, we review the existing literature on swear words, on proposed sound-meaning associations for swear words, and on other empirically observed sound-meaning associations in language. The literature reviewed here allowed us to arrive at our first research questions. In the second part, we provide our own definition of swear words, elaborated from the literature and our own reflection. This more refined definition partly informs the methodology for our case studies, but more importantly, it allows us to define what kind of meaning swear words have, and hence to discuss that meaning later in the discussion.

2.1 Literature review

Two of our research questions require that we define swear words (RQ1: "Is there a tendency for swear words to contain more unsonorous consonants?" RQ2b: "If there is an unconscious form-meaning pattern, what is this meaning?"). We can investigate swear words empirically and propose a common meaning for them, only if we have a precise idea on the nature of swear words. The literature review in Section 2.1.1 will help us find a first tentative definition for swear words. Based on these insights from the literature, we develop our own reflection and find a more refined definition in Section 2.2.

This second definition of swear words informs the method for our case studies described in Chapter 3, and helps us answering our research question RQ2b in Section 4.3 of our discussion.

The literature on swear words shows converging observations and hypotheses on the existence of a sound-meaning pattern in swear words. We review that literature in Section 2.1.2 to arrive at our research question RQ1 ("Is there a tendency for swear words to contain more unsonorous consonants?"), which we answer with our first case study described in Section 3.1.

Sound-meaning associations in language are documented in the literature. These studies on sound-meaning patterns allow us to see how common that sort of association is and how they usually function: what kind of words are involved – i.e., interjections, ideophones, or more regular vocabulary – what kind of meaning is involved – e.g., referential[1] or non-referential – what kind of sounds are involved – i.e., groups of phonemes, submorphemic sequences of phonemes, individual phonemes… – how they are associated – i.e., via probabilistic and contextual mechanisms – their cross-linguistic or language-specific status – i.e., some patterns are found cross-linguistically and thus are probably motivated, others are language-specific – and their cognitive reality – i.e., what the arguments are for their existence in the unconscious part of speakers' linguistic knowledge. This literature, reviewed in Section 2.1.3, allows us to situate our own hypothesised sound-meaning association in swear words among the range of possibilities in sound-meaning associations, and

1 We use the term *referential* to refer to the kind of meaning that *shit* and *excrement* share, or that *penis* and *prick* share, despite their difference in swear word status. This kind of meaning is usually called *denotative*, *referential*, or *truth-conditional*. We use the term *referential* in this thesis but we are not committed a particular stance on the nuances that these terms might otherwise express.

evaluate how likely our hypothesis is.

2.1.1 Swear words: what they do and where they come from

Native speakers can easily identify swearing when they hear it in their own language (Stapleton et al. 2022: 2). To give some examples from English and French, the two languages in our studies, it is certainly obvious to most, if not all native speakers of English that *fuck* is a swear word, and to native speakers of (European) French that *putain* and *merde* are *gros mots* ("swear words"). This intuitive obviousness has the unfortunate result that "[many studies] take swearing for granted as a linguistic, psychological, social or neurological category in its own right" (Ljung 2011: 4) with no need for definition, even though speakers may disagree on what counts as swearing (ibid.; Love 2021: 740–741) because "the human experience of swearing is incredibly subjective" (2021: 740–741). As a consequence, swear words are often not defined explicitly, and when they are, it is with conflicting views and criteria across the literature (Sulpizio et al. 2024: 6–7; Stapleton et al. 2022: 2; Love 2021: 740–743; Ljung 2011: 1–23). Sulpizio et al. (2024) contend that "we know very little of what [swearing] is" (2024: 6) and propose a speaker-informed, cross-linguistic collection of taboo words to contribute to a better definition (2024: 7; 36–39), but do not attempt to provide a synthetic one themselves.

We can make two simple, strong generalisations from the literature. First, swear words are a kind of taboo words, i.e., forbidden words (Ljung

2011: viii, 4–8 ; Bergen 2016a: 12; Love 2021: 741–742; Sulpizio et al. 5–7). Second, swear words express and trigger emotions (Ljung 2011: ix, 4, 21–23; Bergen 2016a: 12, 123–124; Love 2021: 741–743; Sulpizio et al. 2024: 5, 38). Based on the literature reviewed below and our data collection, we will argue for our own more refined definition and tests for identifying swear words in Section 2.2: swear words are socially forbidden words that express and trigger strong emotions, they form a category in speakers" minds often referred to with an everyday term like English *swear words* or French *gros mots*, and for each swear word there is a near-synonymous non-swearing alternative.

As already mentioned in the introduction, a general research question about swear words is: why are swear words taboo, and not other words? What determines the interdiction of those words in particular? A pre-condition seems to be that the meaning of a swear word is – or used to be, as we discuss in Section 2.2.2 – related to taboo semantic domains in a culture, i.e., subjects that speakers are generally forbidden to speak about: the most common domains cross-linguistically are religion, sexuality, and body waste (Stapleton et al. 2022: 2; Bergen 2016a: 12–39). While referring to a taboo topic is a pre-condition to becoming a swear word, this is not enough to qualify as such: *prick* is a swear word, *penis* is not, yet both refer to the taboo domain of sexuality (Ljung 2011: viii); *shit* is a swear word, *excrement* is not, yet both refer to the taboo domain of body waste (Bergen 2016a: 14–15). Actually, the existence of a competing form referring to the same taboo concept seems to be a second pre-condition for turning a neutral word into a swear word: *cunt* was a neutral, publicly acceptable term for female genitalia in 12th-15th century English, even used in medical treaties, before English borrowed a learned alternative from a prestigious language, i.e., *vagina* from Latin (Bergen

2016a: 195; Hughes 2006: 110–111). We will elaborate below on that social process of learned alternatives turning a neutral word into a swear word.

The interdiction of swear words is also related to politeness. The common sense intuition is that swear words are, by default, impolite (Jay 2018; Culpeper 2011: 6; Jumanto & Sulistyorini 2019: 324), but we should nuance and complete that idea: whether swearing is considered impolite is highly dependent on the context, for example on the speaker-listener relationship (Jumanto & Sulistyorini 2019; Culpeper 2011: 142; Jay & Janschewitz 2008). In particular, swear words belong to more informal, relaxed settings (Jumanto & Sulistyorini 2019; Jay & Janschewitz 2008: 273). All those observations suggest a simple idea that has, to our knowledge, not yet been proposed in the literature: that swear words form a specific register, i.e., the lowest, most relaxed, most informal, least valued register possible. It should also be noted that some speaking communities have taboo topics and norms of politeness, but reportedly very few or no swear words, like Japanese (Robertson & Stephens 2017: 268–269; Bergen 2016a: 34–35; Wajnryb 2005: 223–236) or Dalabon, a native language of Australia (Ponsonnet, personal communication). We will come back to the intriguing idea of a language not having any swear words, and what it implies for our research question in Section 2.2.4.

Stapleton et al. (2022) provide a comprehensive overview of studies on the power of swear words, showing that they create "emotional arousal" (2022: 3–7, 9–13). First of all, two medical conditions suggest that swear words, due to their emotional value, are treated in a specific way by the human brain. The first condition is aphasia. In some cases of aphasia, speakers are able to utter only so-called "automatic" speech, i.e., some high frequency words or phrases like counting from one to ten, reciting the alphabet or the days of the

week, greetings, or pause fillers. In other cases, they are able to utter only swear words (Van Lacker & Cummings 1999: 86). The second condition is coprolalia, a common symptom of the Tourette syndrome, where speakers cannot help themselves but utter swear words (Van Lacker & Cummings 1999; Finkelstein 2018) and other "taboo *acts* of talking" (Bergen 2016b), i.e., expressions that are not swear words but taboo in context, e.g., *I love Jesus* during a Jewish religious ceremony (Finkelstein 2018: 321), or words meaning "foolish" or "ugly" in Japanese (Van Lacker & Cummings 1999: 89). Those observations suggest that swear words, like other parts of speech preserved in aphasia and coprolalia, "may be located and processed in different parts of the brain from other speech activities […], [namely] basal ganglia, amygdala and other parts of the limbic system, […] deep structures in the brain that play a central role in processing memory and emotion […] reflect[ing] a more automatic or impulsive mode of processing and expression." (Stapleton et al. 2022: 4)

Moreover, swear words are remembered more easily and consistently than other words, even emotionally-tagged language like reprimands or endearments (ibid.). They are processed more slowly and more accurately compared to other items (ibid.; Sulpizio et al. 2019). They also command more attention and are highly salient cognitively, to the point that they interfere with the processing of other salient experimental stimuli, even more than other types of emotional words (Stapleton et al. 2022: 4–5). For a second language, the effect on memory is also observed but the distraction effect appears only in a reduced form, suggesting that swear words are less powerful to second-language speakers.

Swear words have also been shown to trigger an involuntary body function involved in a "fight or flight" reaction to imminent danger, namely

increased sweating, measured by heightened skin conductance rate (2022: 5–6; Kamp et al. 2024). Increased heart rate, a related involuntary body function, has also been observed in some – but not all – studies (Stapleton et al. 2022: 6). The effect for both reactions is stronger for native speakers (Kamp et al. 2024), and also stronger for swear words compared to their "X-word" euphemisms, e.g., *f-word* for *fuck* (Bowers & Pleydell-Pearce 2011).

Uttering swear words – compared to non-swear words – increases pain tolerance and reduces pain perception, allowing speakers to, for example, withstand immersing their hand in icy water for a longer time, with a stronger effect for people who do not catastrophise pain, or people who do not swear frequently (Stephens and Umland 2011; Stephens et al. 2009; Robertson et al., 2017; Stephens and Robertson, 2020). It also increases muscular power and strength in physical activity, maybe due to the same pain-relief effect, i.e., the effort is made easier because some of the pain is removed (Stephens et al. 2018).

As already mentioned, the most likely explanation for those physical effects of uttering or being exposed to swear words, is that they create emotional arousal (Stapleton et al. 2022: 4–7): some studies assume that swear words are only negatively valenced and always trigger a kind of "fight or flight" response, but they fail to explain the numerous positively valenced effects of swear words, in particular humorous rhetorical effects and interpersonal acts of group affiliation (2022: 7–9). In short, we know that they convey emotional arousal, but not necessarily negative emotions.

If we want to understand more precisely what kind of emotional arousal, and how they convey it, we need to turn to other pieces of evidence. Based on observations on coprolalia, Finkelstein (2018) proposes that swearing is not just a transmission of information, but an emotional act:

> Swearing differs from other forms and usage of language. While words often denote objects and actions, describe events and share ideas, swearing does not. Swearing uses language to express and evoke emotions [...]. Words that name emotions rarely evoke them: *fear*, *disgust*, and *sexual desire* denote emotions but typically do not elicit them. Swearing does. [...] [Swearing] has the power to express emotions and to evoke them, not with semantics but by special use of the socio-cultural attunement […]. *Nigger* is not an *emotion word* in that it does not denote an emotion, but "nigger" is an *emotion act* because of the cultural context attached to it. […] [Swearing] is an emotion act. (Finkelstein 2018: 311, 326; emphasis in original)

Ljung (2011) holds similar views based on a cross-linguistic study of swearing in general, not just coprolalia:

> Semantically, [swear words are] special because […] [they] are not used with their referential or denotative meanings, but function […] as indications of the speaker's state of mind. Consequently swearing does not have meaning in the sense that referential expressions do. Instead it has *emotive meaning*, *viz*. it expresses the speaker's state of mind. (Ljung 2011: ix; emphasis in original)

To put it differently, the referential meanings of swear words, often related to taboo subjects like religion, sexuality, or body waste, are not as important to their nature as their non-referential meaning, which is to express and provoke emotions, in the speaker's and/or addressee's mind. This opinion is coherent with the observations mentioned above that swear words from one's native

language create a stronger reaction, i.e., more distraction, increased sweating, and faster heart-rate, compared to swear words from another language, and compared to their "X-word" euphemisms: in both comparisons, the referential meaning is the same, yet the emotional effect is measurably different.

Further confirmation of this resides in the fact that, when we look at some of the most prototypical swear words of French – i.e., some of the words that strike French speakers as best examples of swear words – their referential taboo meaning has largely if not completely disappeared: they no longer refer to anything related to religion, sexuality, body waste, or any other taboo subject for that matter. We come back to this issue in Section 2.2.2.

The greater importance of their emotional or attitudinal meaning makes swear words close to other linguistic items, namely interjections, e.g., *wow*, *duh*, including semi-natural "pragmatic noise" (Culpeper and Kytö 2010) like *ah*, *oh*, *mmm*, *tut*, *hey*, etc. Wharton (2003) argues, based on their meanings and grammatical constraints on their use, that they are closer to showing than saying on a show-saying continuum, i.e., interjections are felt unconsciously to show emotions and states of mind, rather than tell them (2003: 201– 213). This is not surprising since historically, they seem to have conventionalised from natural vocal reflexes caused by, e.g., surprise, distress, amusement, distraction, thoughtfulness, pain, etc. (Culpeper and Kytö 2010; Wharton 2003: 203–205). Given Finkelstein's (2018: 311, 326) and Ljung's (2011: ix) observations mentioned above, it makes sense to say that swear words show emotions and states of mind rather than tell them, in the same way that interjections do. There is also a more obvious connexion between interjections and swear words, in the fact that many swear words can be used as interjections (e.g., "Damn!", "Fuck!", "God!", "Jesus!", "Shit!", etc.), so much so that a few of them have been

euphemised into new interjections, like *Jesus* becoming *jeez*, or *gee* (Culpeper and Kytö 2010: 305). In Sections 2.1.2 and 2.1.3, we will mention another commonality between swear words and interjections, i.e., they display sound-meaning tendencies.

To sum up, the interdiction of swear words is the product of taboo semantic domains and politeness norms. Those norms and domains gave a strong emotional, attitudinal value to swear words, similar to interjections, which in turn certainly contributes to a kind of self-perpetuating social paradox (Bergen 2016a: 283): we forbid swear words because everyone else does; more precisely, we keep on forbidding swear words because they provoke strong emotions, but they provoke strong emotions because they were forbidden in the first place, i.e., "the remedy is the cause" (ibid.).

All this would explain why there are swear words, and where they get their strong emotional value from, but not why the social interdiction targets those words and not their taboo-related synonyms, e.g., why *shit* and *crap* are swear words but *excrement* and *stool* are not. On this issue, Ljung (2011) is quite pessimistic: "I think we have to conclude that […] the choice of swear words among the words considered to be taboo are to a great extent a matter of chance […]." (2011: 8) McEnery (2006), based on a historical study of discourses on swearing in English, offers a more convincing explanation: the interdiction of swear words is a historical product of social distinction. For English, those words were initially neutral – although taboo-related – and used by the lower social classes but were subsequently, from the 16th century onward, considered problematic because of recurrent moral panics incited by middle and upper classes, especially when those words were used in discourse challenging authority (ibid.: 226–227). If we combine McEnery's insight with the previously

mentioned observation that *cunt* was a neutral word before the borrowing of Latin *vagina* (Bergen 2016a: 195; Hughes 2006: 110–111), we can propose a three-step historical process for the interdiction of swear words: first, higher social classes begin to replace taboo-related neutral words with new learned alternatives from Greek and Latin, then lower social classes continue to use the originally neutral words, and finally upper classes begin to frown upon the continued use of the same words by the lower classes. To take the same *cunt* vs. *vagina* example: first, *vagina* is introduced as a knowledgeable alternative from Latin to *cunt*, then *vagina* is increasingly preferred to *cunt* in higher classes while lower classes continue to use *cunt*, and finally *cunt* is made into a bad word by negative associations with lower classes and moral panics. This three-step model works, but would not apply, or only partly, to swear words related to religion, e.g., *God* and *Jesus Christ* were not replaced by higher class alternatives as they would invariably have been frowned upon in the first place, because of the biblical Third Commandment against taking the Lord's name in vain (Pinker 2007; Hughes 2006:), an exemplary case of word magic (Pinker 2007; Hughes 2006: 201, 462), which is "the belief that words […] have the power to unlock mysterious powers in nature and to affect human beings and their relationships" (Hughes 2006: 512–513). Moreover, social distinction by higher classes alone can hardly account for the interdiction of slurs, i.e., swearing insults expressing bigotry towards a dominated social group, e.g., *nigger, spic,* or *fag* – in all likelihood, the groups targeted by those words also contributed to their interdiction. It is however debatable whether slurs should be categorised as swear words: we will discuss this specific issue in Section 2.2.5.

McEnery's (2006) point on social distinction can explain why some

taboo-related words became swear words and others did not. The remaining parts of the swearing lexicon can be explained by religious taboos, or more recent taboos against bigotry. Yet based on empirical observations, some researchers argue that another element is at play here, making some words a more likely target for social interdiction: the sounds they contain. This is what we cover in the next section.

2.1.2 The form of swear words

Several intuitive observations have been made in the literature about the form of swear words. Wajnryb (2005: 205–210) notices that English swear words are overwhelmingly monosyllabics or trochees, and contain plosives /p/, /b/, /t/, /d/, /k/, and fricatives /f/, /s/, /ʃ/. She gives anecdotal cross-linguistic evidence that this might be a universal pattern, with Finnish *vi**tt**u* ("cunt"), French *mer**d**e* ("shit"), Japanese ***b**a**k**a* ("idiot"), Hebrew ***b**en zonah* ("son of a whore"), *le**ch** **t**iz**d**ayen* ("go fuck yourself"), and ***k**oo**s*** ("cunt") – borrowed from Arabic – and Romanian ***p**iz**d**a* ("cunt"). Hughes (2006: 343) notices that English swear words often start with /b/, /d/, and /f/, and end with /k/, while ethnic slurs tend to be monosyllables (e.g., *chink*, *spic*) or two-syllable diminutives (e.g., *paki*). Pinker (2007) notices too that English swear words are "quick and harsh" as they are monosyllabics or trochees – i.e., a stressed syllable followed by an unstressed syllable – with short vowels and plosives, /k/ and /g/ in particular. Wajnryb (2005), Hughes (2006), and Pinker (2007) do not confirm those somewhat converging intuitions with any hard empirical evidence, so one might easily dismiss those patterns as existing only in the eye of the beholding linguists, a

mere result of chance – plosives and fricatives are very common phonemes after all. We need statistical comparisons with non-swear words, and/or experiments to check whether speakers actually notice and are influenced by such patterns. The studies we review in the next paragraphs, along with our three empirical studies described in Chapter 3, attempt to provide such evidence on swear words.

Yardy (2010) proposes that swear words are more likely to contain a specific group of consonants given a motivated form-meaning pairing. A theory of animal communication called the Affect Induction model (Owren and Rendall 1997, 2001) claims that the *least sonorous* sounds – i.e., signals with a "broadband and harsh" quality (Yardy 2010: 13, 16, 17) – are produced by animals in situation of conflicts, while the *most sonorous* sounds – i.e., signals with a "narrowband and tonal" quality (ibid.: 13) – are produced in situations of submission or harmonious social interaction (ibid.: 12–20; 71–78): "A familiar example ... is the contrast between the growling and barking of an aggressive dog and the yelp or whimper of a submissive one." (ibid.: 13) This model can be applied to the phonemes in human language: consonants are ordered from the most sonorous, i.e., most vowel-like, to the least sonorous, i.e., least vowel-like, so that they can be interpreted unconsciously from most harmonious, to most conflictual: "from an affect induction perspective, sonority could represent a progression from broadband harsh sounding consonants – analogous to animal vocalizations in hostile situations – towards harmonic voiced consonants and vowels, analogous to animal vocalizations in affiliative situations" (2010: 19). Different classifications for sonority are proposed by scholars, but Yardy settles for a hierarchy by Parker (2008) where the scale from most vowel-like to the least vowel-like goes in the following order: glides (e.g., /j/, /w/), rhotic

approximants, flaps, laterals (e.g., /l/), trills, nasals (e.g., /n/, /m/, /ŋ/), voiced fricatives (e.g., /v/, /z/), voiced affricates (e.g., /dʒ/), voiced plosives (e.g., /b/, /d/, /g/), voiceless fricatives (e.g., /f/, /s/, /ʃ/, /θ/, /h/), voiceless affricates (e.g., /tʃ/), and voiceless plosives (e.g., /p/, /t/, /k/) (Yardy 2010: 63; Parker 2008). Yardy (2010) proposes anecdotal evidence that the same (un)sonority-meaning correspondence is also used in interjections:

> A real world example of how these psychoacoustic properties of phonemes are pragmatically utilized is the use of the obstruent "shhh!" [/ʃ/] in an effort to quiet another individual or the use of the sonorant nasal "mmm"[/m/] in the affiliative situation of eating food in a social context. Indeed, the most basic phonetic distinction between vowels and consonants may have its basis in this Affect Induction model of animal communication. (Yardy 2010: 19)

With this theoretical model in mind, Yardy compares corpus data from English songs – namely, lullabies, carols, and heavy metal songs – with data on swear words. He finds significant tendencies where, compared to words from songs, swear words contain less of the most sonorous consonants, like /j/, /w/, /r/, /l/, /m/, /n/ or /ŋ/, voiced fricatives, and the voiced affricate, and contain more of the least sonorous consonants, namely plosives, voiceless fricatives, and the voiceless affricate /tʃ/ (Yardy 2010: 52–64).

There is a criticism to be made to Yardy's study however, namely that his dataset is unreliable for several reasons. First of all, Yardy compares isolated swear words with all words used in a corpus of songs – words from carols, lullabies, and heavy metal songs (ibid.: 50): he compares the lexicon (swear words) with language use in a context (words in songs), which is

problematic in itself – a high frequency in the lexicon does not correspond to a high frequency in language or vice versa. Second, his collection of swear words is not systematic. He uses swear words from the following sources: "an internet video of an unofficial list of profane and offensive words that are optionally censored in the comment section of YouTube; a swearing/profanity web site; a profanity list used by a "politifier" program designed to replace swear words with euphemisms; and a profanity list meant for webmasters and forum administrators." (2010: 50). These informal lists have all been collected through the explicit judgment of a handful of speakers: they may not be selective enough and not be representative of what English speakers consider to be swear words. This concern is all the more justified since Yardy's final list of swear words contains a sizeable amount of 437 words (2010: 51): the makers of these online lists have probably collected so many words for the mere sake of exhaustivity with no regards for frequency or salience – compare with our 71 swear words for English and 78 swear words for French given in Section 3.1. Third, this results in Yardy's list containing numerous homophones, for example *phuck* and *fuck* are both included, and numerous word derivations, for example *fuck* is contained in 64 words and *shit* in 29 words (ibid.: 51, 56). Consequently, one may object that the tendency to contain more of the least sonorous consonants is simply due to an overwhelming presence of homophones, derivations or cognates of a few swear words who happen to contain those consonants, like ***fuck*** or ***shit***. Finally, Yardy compares words from songs and swear words, but never compares any of those subgroups to a representative sample of the English lexicon. Thus one could logically interpret his results in the following way: the proportion of unsonorous consonants in English swear words may be representative of the English lexicon, so that there

is nothing particular about swear words, and what is noticeable instead is that words from English songs tend to deviate from the that and feature fewer unsonorous consonants.

Although his results are based on an unreliable dataset, Yardy's hypothesis is well worth investigating, because it makes sense in a systematic way of the piecemeal observations by Wajnryb (2005), Hughes (2006), and Pinker (2007) mentioned earlier, and because it is coherent with many other observations in the literature on swear words (see below). For these reasons, Yardy's (2010) unsonority hypothesis serves as the basis for our RQ1 and methodology for our case studies described in Chapter 3.

A major inspiration for our research is Bergen's (2016a: 49-64) suggestion that for speakers of English, a specific form feels intuitively appropriate for a swear word: a closed monosyllabic word ending with a consonant, like *fu**ck***, *shi**t***, *di**ck***, or *cun**t***. Bergen's claim is based on the observation that English swear words tend to be closed monosyllabics, compared to the rest of the English lexicon, more specifically closed monosyllabics ending with a plosive (2016a: 52–53). He asked native speakers to rate whether specific nonce-words sounded like they could be swear words of English: the results seem to confirm an unconscious association between the word being a closed monosyllabic and being a swear word (2016a: 55–56). He offers anecdotal evidence that this closed monosyllabic pattern might be true in English swear words from fiction (2016a: 54). Bergen considers different origins for this phenomenon. First, he gives a functionalist explanation, i.e., this form may be more practical for speakers when they want to swear. Second, he considers an explanation in terms of indirect symbolism, i.e., it may be symbolically opposed to the open and repeated syllables pronounced by

infants. Finally, he offers an explanation in terms of cultural specificity, i.e., it may be a self-reinforcing idiosyncrasy of English (2016a: 56–63). He remarks that only a comparison with swear words in other languages can help finding the right explanation. In this thesis, we take up his invitation (2016a: 60, 63) to test for this form-meaning association in other languages, by looking at French (see Chapter 3).

Based on psycholinguistic evaluation by English speakers of existing swear words and other taboo-related words (e.g., *cancer*), Reilly et al. (2020) found that obstruence, i.e., the presence of plosives and fricatives, is a "statistically significant word form/phonological predictor of tabooness"(2020: 142). They find that syllable structure and word length are not significant predictors (ibid.), which runs counter to Bergen's (2016a: 49-64) suggestion of a closed monosyllabic tendency for swear words. However, in a second study, when asked to rate the plausibility of invented swearing compounds (e.g., *ass-rocket*), "participants judged shorter words with more stop [i.e., plosive] consonants as better candidates for novel taboo terms" (Reilly et al. 2020: 144). Those results are coherent with Bergen's (2016a: 52–53) side remark on the presence of plosives at the end of monosyllabic swear words, and even more with Yardy's (2010) unsonority hypothesis: plosives and/or fricatives are comparatively less sonorous than other phonemes, so they should be more frequent in swear words.

Lev-Ari & McKay (2022), on the other hand, argue that cross-linguistically, swear words do not have a tendency to contain any specific category of sounds, contrary to previous claims, but instead have a tendency to avoid sounds like /l/ /r/ /w/ and /j/ – which they classify as "approximants". Their dataset is more reliable than Yardy's (2010). While Yardy compares

lexicon with specific usage, they compare the swearing lexicon with a representative list of the non-swearing lexicon, i.e., the 100-item Swadesh list of words for basic concepts, often used in studies involving cross-linguistic comparisons (Lev-Ari & McKay 2022: 1105). Instead of relying on informal lists compiled by individual speakers aiming for maximal exhaustivity, they elicit swear words from speakers via a questionnaire, and select only the items provided by at least two different participants (ibid.). They also avoid the presence of homophones or derivations completely by filtering out any variant, and "kept the variant that was rated as more offensive" (ibid.). Their dataset is still unreliable however, for the following reasons: the definition of swear words in their instructions is quite open to interpretation and cross-cultural misunderstandings, and has probably focused the respondents" attention too much on insults: "the most vulgar words that are used in [your language] when someone gets hurt or frustrated and the most offensive words that are used to curse someone (i.e., to disparage or insult them)" (ibid.). Such vague and insult-focused instructions may have skewed the respondents" answers and led to the presence of noise, or on the contrary, absence of relevant items: because of the vagueness, non-swear words might end up in the list, and because of the focus on insults, non-swearing insults might end up in the list while other existing swear words escape the respondents" memory. We find confirmation of those concerns at least in their French data, which clash with our intuitions as a native speaker, and more importantly with the speaker-informed data collected for this thesis (see Section 3.1). For example, we would argue that noise leads them to include *bouffon* ("buffoon") and *trou de balle* ("anus" or "butthole", literally "bullet hole") in their list of French swear words, while these are insults but not swear words. Their list includes only 22 items, while

ours includes 78. Strikingly, their respondents provided the insult *sous-merde* (literally "sub-shit") but not the much more frequent and prototypical swear word it is derived from, i.e., *merde* ("shit"). Their respondents returned the Quebecois French *tabarnak* (literally "tabernacle") but none of the other well-known examples of Quebecois French swear words, e.g., *crisse* (literally "Christ"), *calice* (literally "chalice"), or *hostie* (literally "Host"). These shortcomings undermine their claim that swear words do not tend to contain more plosives (2022: 1106) or only in "a handful of related Indo-European languages" (2022: 1104). However, the tendency they do identify is coherent with Yardy's (2010) unsonority claim: if the more sonorous consonants are less frequent in swear words, then that would include sounds like /l/ /r/ /w/ or /j/, which they found to be less frequent in English, Hebrew, Hindi, Hungarian, Korean, and Russian swear words (Lev-Ari & McKay 2022: 1106). We can take this as another partial confirmation of the Yardy's (2010) unsonority hypothesis.

Finally, Chiang & Schlatter (ms.) collect swear words across 31 languages from 10 language families, and observe that swear words are often plosives, fricatives, bilabial, velar, palatal, and voiceless, while the vowels are often back and open. They offer an explanation which is both functional and iconic: that those tendencies are suitable for catharsis and the expression of aggression and other negatively valenced emotions, because language has evolved "from a device used solely for the multimodal expression of emotions to the complex systems we find today", and "that the ancestor of modern day cathartic swearing is found in mammalian calls", e.g., cries and facial expressions among primates. In other words, producing those sounds and the resulting facial expressions is what we do – along with other primates – when we express strong negative emotions; so words with those sounds feel, and

physically are, better suited than other candidates, resulting in cross-linguistic tendencies in swear words. Similar to our comments on the study by Lev-Ari and McKay (2022), we would argue that the instructions they gave their respondents are overly vague and focused on insults, so that their resulting swear word data may not be exhaustive and/or contain noise. This is again visible for their French data which include items clashing with our native speaker intuitions, and absent from our speaker-informed data: *mince alors* and *zut,* both translatable as "blast!" or "drat!", i.e., non-swearing interjections; and *punaise* ("darn!", literally "stink bug" or "pin"), a common non-swearing interjection and euphemism for *putain* ("fuck!"). However, and again just like Levi-Ari and McKay (2022), the tendencies they identify for swear words are largely coherent with Yardy's (2010) unsonority proposition: plosives, fricatives, and voiceless sounds are comparatively lower on the scale of sonority and hence should indeed be more frequent in swear words. Moreover, their explanation is quite similar and not incompatible with Yardy's (2010): both mechanisms – an all-species unsonority tendency, along with primate-specific, multi-modal expression tendencies – can combine to create the sound patterns observed in swear words.

Such studies suggest that there is indeed a way to sound swearish, i.e., that the presence or absence of specific sounds in words makes them more or less likely candidates for swear words, due to an unconscious sound-meaning association. Yet as we have seen, there are conflicting views as to what sounds exactly are involved and what meaning exactly is involved, and there are different explanations for this pairing, most of them hypothesising some form of cross-linguistic motivation. We take Yardy's (2010) unsonority hypothesis to be the most likely: of all the proposed explanations, it is the one

that accounts for most of the observations found in the literature. Hence our first research questions on English and French: "Is there a tendency for swear words to contain more unsonorous consonants?" (RQ1) "Does that tendency correspond to a cognitively real, unconscious form-meaning pattern for swear words?" (RQ2a).

Before delving further into the particulars of swear words however, we should evaluate how likely that kind of unconscious sound-meaning association is in language, not in the least because they run against one of the classical tenets of linguistics, known as the duality of patterning (Dingemanse 2015: 604) or double articulation (Haiman 2018: 117–118; Martinet 1957). Also, to figure out exactly what is going on in swear words, it is useful to understand how that sort of association usually works, i.e., what kind of meanings are associated with what kind of sounds, in what kind of words.

2.1.3 Sound-meaning tendencies in and across languages

Duality of patterning (Dingemanse 2015: 604) or double articulation (Haiman 2018: 117–118; Martinet 1957) is the notion that sounds in language, i.e., phonemes, are meaningless. Functionally, we need sounds to be meaningless units so that we can combine them into an infinite number of sequences, i.e., possible words or morphemes, corresponding to an infinite number of possible meanings. Conversely, if our limited set of sounds in language had meanings, then we could combine them into sequences with only a limited number of possible meanings, derived from the limited number of meanings attributed to sounds. A direct consequence of double articulation is that it is impossible to

predict the meaning of a word or morpheme, given the sounds it contains; or the other way around, that it is impossible to predict what sounds a word or morpheme contains, given its meaning – we will come back to duality of patterning in more details in the discussion in Sections 4.4 and 4.5.

There is no doubt that duality of patterning is a relevant notion to describe language and that sounds are definitely not paired with meanings in the same way as words and morphemes. However, the existing literature on sound-meaning associations nuances that position by showing that submorphemic sound sequences, individual sounds, or even categories of sounds, can actually be associated unconsciously with meanings. Before we go any further on sound-meaning associations, we must also argue that duality of patterning is related to, but fundamentally distinct from Saussure's arbitrariness (2005 [1916]). This distinction is not necessarily made explicit in the literature, and is clearer if we redefine arbitrariness in a way that is more specific than originally discussed by Saussure. Some of the sound-meaning associations discussed below run counter to arbitrariness (for example, they are iconic sound-meaning associations, based on resemblance and found across non-cognate languages) and some do not, according to our definition of arbitrariness (i.e., the association between this specific sound and that specific meaning is, as far as we know, arbitrary and language-specific). On the contrary, all of them run counter to double articulation by definition, because they are sound-meaning associations. We will develop further on these theoretical issues and distinctions in the discussion in Section 4.5.

A well-known type of sound-meaning association in language is onomatopoeia, i.e., words that imitate natural sounds (Dingemanse 2015: 604; Sidhu 2019: 128; Ibarretxe-Antuñano 2024: 490). Onomatopoeic words imitate

natural sounds with phonemes, so that their phonemes correspond to their meanings and can be partly predicted, given their meanings. For example, it is easy to predict that an onomatopoeia for explosion will feature an initial plosive phoneme imitating the initial burst of an explosion, like initial /b/ in English *boom* or *bang*, or initial /d/ in Japanese *don* or *dokan* (Kakehi et al. 1996: 261–262; 269) or in Basque *danba* (Ibarretxe-Antuñano 2024: 490).

Onomatopoeic words belong to a larger category of words well-known for displaying sound-meaning associations: ideophones (Ibarretxe-Antuñano 2024: 490, 498; 2023: 324–326; Haiman 2018: 79–88). These hardly definable, expressive, performance-like items could be informally characterized as linguistic sound effects: speakers – and linguists alike – feel that ideophones show what they mean instead of telling what they mean the way regular words would (Haiman 2018: 76–88; Ibarretxe-Antuñano 2023: 314) – again, just like swear words and interjections, as discussed in Section 2.1.1. They can convey non-auditory, even very abstract meanings: Japanese ideophone *run-run* conveys "the state of being in a happy, light-hearted mood" and *kuyo-kuyo* conveys "the manner of fretting; the state of being fretful" (Kakehi et al. 1996: 726; 1058). Ideophones are rare in most European languages, but frequent and often conventionalized in others, especially in African and Asian languages (Voeltz & Kilian-Hatz 2001: 4). There are specialised ideophone dictionaries for, for example, Japanese (Ruiz Martínez 2019; Kakehi et al. 1996) and Basque (Ibarretxe-Antuñano 2023: 313). Ideophones display sound-meaning regularities irrespective of onomatopoeia: Haiman (2018: 117–139) provides an extensive review of the existing literature where he reports the following main findings:

- in ideophones from a variety of languages, high tone and/or high front vowels and/or voiceless consonants convey a meaning of small size, and their respective opposites big size. The use of high frequency to mean "small" is an application of Ohala's (1994, 1984) frequency code, a type of sound-meaning association we will cover further below.
- on the contrary in Korean ideophones, high vowels correspond to augmentatives, and low vowels to diminutives
- high front vs. open back vowels correspond to pleasant vs. unpleasant, and slippery vs. not slippery meanings in Siwu ideophones
- lax vs. aspirated vs. tense plosives correspond respectively to slow vs. faster vs. fastest meanings in Rengao ideophones
- in Japanese ideophones, voiced vs. voiceless sounds correspond respectively to smooth, light, fine, vs. big, heavy, coarse meanings, while palatalized consonants are associated with childishness, excessive energy, instability, unreliability, lack of elegance, and continuant vs. not continuant sounds are associated respectively with continuous movement, shapelessness vs. abrupt movement or surface, and the vowel /e/ corresponds to negative affect.

For ideophones of Emai, Egbokhare (2001: 92–93) describes a particularly extensive, coherent – and according to him, iconic – system where entire categories of phonemes correspond to geometric shapes and densities: vowels correspond to meanings of compactness, e.g., /i/ is used in ideophones for something compact and dense, while /a/ for something very flat and diffuse; nasal consonants for something abnormal and crooked; approximants for material cohesion, e.g., /w/ for something loose and chunky, /h/ for something

light; fricatives for tactile dimensions, e.g., /z/ to something creasy, /ʃ/ to something scruffy or rough; and each pair of plosives corresponds to a geometric dimension, either width with /b/ for broad vs. /p/ for narrow, proportionality with /d/ for proportional vs. /t/ for disproportional, length with /g/ for long or high vs. /k/ for short or low, and size with /g͡b/ for big vs. /k͡p/ for small.

In the previous section, we quoted Yardy's (2010: 19) remark on the possible iconic motivation for "shh!" to silence someone and "mmm" when eating. Other interjections display many other systematic sound-meaning correspondences. Ward (2003) identifies tendencies for American English, based on corpus data: /m/ expresses thoughtfulness, /oʊ/ (/əʊ/ in British English) expresses surprise, a creaky voice expresses a claim for authority, clicks (for example, noises usually written *tut* or *tsk*) express dissatisfaction, and /h/ or breathiness express concern (2003: 21–36). Culpeper and Kytö (2020: 284–286), based on a corpus of Early Modern English written dialogue, i.e., Renaissance theatre plays, confirm those patterns for /m/ and thoughtfulness in *hem*, *um*, *hum*, and /oː/ – the Renaissance form of the diphthong /oʊ/ or /əʊ/ in present-day English – for surprise in *oh*. Sound-meaning patterns are present in English interjections, but also across languages, probably because they "have arisen as exaggerations or developments of entirely natural responses" (Wharton 2003: 204). For example, "despite the fact that interjections that express pain are language specific – English *ouch*, French *aïe*, Spanish *ay* – they do all begin with the same mid-front vowel" (Wharton 2003: 205) that Darwin (1872 [1998]: 97) "describes as being naturally expressive of pain." (Wharton 2003: 205).

First names can also be subject to interpretation following

unconscious sound-meaning pairings: speakers of English unconsciously associate sonorous consonants (e.g., /l/ /n/ and /m/ in *Luna* or *Milo),* with high emotionality, agreeableness, and conscientiousness, whereas they associate plosives (e.g., /k/ and /t/ in *Jack* or *Kate*) with high extraversion (Sidhu, Deschamps, Bourdage, & Pexman. 2019: 1610).

This kind of association between sounds and meanings is not confined to the apparent margins of language like onomatopoeia, ideophones, interjections, or proper names, and are often not confined to one language at a time.

A first type of such meaningful sound-meaning associations is phonaesthemes. These are sound-meaning pairings that are statistically significant, present at the level below words and morphemes, observable in small parts of a lexicon, and unconsciously present in speakers" minds. For example, Bergen (2004: 293) mentions the *gl-* onset associated with "light" or "vision" in English, where 39% of words starting with *gl-* have such meaning, like *glitter*, *glisten*, *glow*, *glance*, *glare*, etc. Another example is that 28% of *sn*-words like *snack*, *snore*, or *sneeze* have a meaning related to "mouth" or "nose". Contrary to morphemes, phonaesthemes are non-compositional, e.g., *gl-* means "light or vision" in *glitter*, but *-itter* does not necessarily mean anything. Moreover, only a certain percentage of words with a given sequence corresponding to a given meaning. This does not make for a form-meaning relationship as reliable and systematic as a morpheme or a word. Yet it makes the meaning predictable to a certain extent – 39% or 28% is more than you would expect by chance – and speakers do notice: phonaesthemes are not just present in the lexicon, but also unconsciously in the mind of speakers, so that they have priming effects on word recognition comparable to morphemes

(Bergen 2004: 299–303). Statistical research on the English lexicon suggests that phonaesthemes are more common in language than previously assumed (Gutiérrez et al. 2016: 2379).

Ohala (1994, 1984) notices that across languages, words expressing the meaning "small" tend to use "high front vowels like [i ɪ y e]" or high tones, while words expressing the meaning "high" tend to use "low back vowels like [ɑ ʌ ɔ o]" or low tones (1994). He sees it as part of a larger frequency code using resemblance between these sounds and their meaning, i.e., using iconicity: high frequency sounds are pronounced with a comparatively smaller oral cavity, while a low frequency sounds with a larger one (1984). The rationale for investigating cross-linguistic sound-meaning tendencies is the following: "If we find similar sounds associated with similar meanings in a sufficiently diverse number of languages, especially those not genetically or areally connected, then we might with some confidence consider whether a non-arbitrary sound-meaning relation exists and then seek an explanation for it" (1994). To put it differently, if unrelated languages tend to express a given meaning with a specific sound or type of sounds, then the link between form and meaning is not as random as posited by Saussure's tenet of arbitrariness, and might be the result of some motivation instead. This rationale for motivation in cross-linguistic sound-meaning tendencies underlies the other studies that we cover below. One might disagree with specific motivation hypotheses proposed by researchers to account for these cross-linguistic sound-meaning tendencies, but they contradict arbitrariness by definition and have to be explained in some way.

Blasi et al. (2016) analyse 100 basic vocabulary items in around 4000 languages, from 359 lineages, and identify sound-meaning tendencies found

across unrelated and geographically distant languages. These associations concern mostly property words, e.g., /i/, /y/ and /tʃ/ for "small", varieties of r-sounds like /r/ for "round", or /p/ and /b/ for "full", and body parts, e.g., /l/, /e/, /ɛ/ and /œ/ for "tongue", /n/, /u/, and /ɯ/ for "nose", /m/, /u/, and /ɯ/ for "breasts" (2016: 10820). Some of them can be explained intuitively by iconicity. Their results confirm partly Ohala's (1994, 1984) frequency code: high front vowels /i/ and /y/ are iconically appropriate for "small" because they are pronounced with a comparatively smaller oral cavity, just like the ideophones with high front vowels for small-sized concepts that we mentioned earlier (Haiman 2018: 117). R-sounds are probably imitations of a rolling object. The nasal consonant /n/ is probably used for "nose" because the air flow goes through the nose, and /m/ for "breasts" because of "the mouth configuration of suckling babies" and/or "sounds feeding babies produce" (Blasi et al. 2016: 10820). There is an eye-popping iconic association in their results, which for some reason they do not comment on, maybe because they would not interpret it themselves in terms of iconicity: words for "sand" tend to contain voiceless alveolar fricatives like /s/ (ibid.). We would interpret this as an imitation of the sound produced by moving sand.

Based on experiments with German speakers, Aryani et al. (2018) propose that short vowels, voiceless consonants, plosive consonants, and hissing sibilants – i.e., a subtype of fricatives and affricates, e.g., /s/ /z/ /ʃ/ – feel more arousing and negative. This applies well to swear words, which provoke arousal and are often perceived as negatively valenced. Aryani et al. (ibid.) are not interested in swear words proper, but they use the example of English *piss* – felt more rude – vs. *pee* – felt more childish or polite – to illustrate their point. This is also coherent with Yardy's (2010) unsonority hypothesis for swear

words, as all those categories of sounds – voiceless consonants, plosive consonants, and hissing sibilants – are comparatively less sonorous. Moreover, vowels are by definition the most sonorous phonemes, but one could see a short vowel as comparatively less sonorous than a long vowel: this vowel length parameter could have contributed to the interdiction of swear words with short vowels, ultimately resulting in a few minimal pairs in English with short vowels in swear words and long vowels in near-identical non-swear words, like *shit* (short vowel) vs. *sheet* or *shoot* (long vowels), or *bitch* (short vowel) vs. *beach* (long vowel). Aryani et al. (2018) do not provide cross-linguistic data as they tested only German speakers, but they interpret their results in terms of motivation nonetheless.

Haiman (2018: 209–212) observes that in 10 out of the 15 language families of the world, according to Greenberg's (1987: 337) typology, the most archaic forms of personal pronouns obey the following two-fold rule: (1) to express first person singular ("me"), use a nasal consonant (such as /m/ or /n/), do not use any plosive consonant; (2) to express second person singular (singular "you"), use either a plosive or a nasal consonant (depending on the language family).[2] Haiman (2018) interprets this historical cross-linguistic fact in terms of the (non-)violation of the interlocutor's space:

> Pointing, staring, the speech acts of hailing and address, and the use of the second person pronoun are equivalent familiarities, insofar as they are all violations of the hearer's space. All are sublimations of physical intimacies such as over-proximity, touching, shoving, jostling, grabbing, caressing, penetration – surely welcomed on

2 The 10 families following this rule are Amerind, Eurasiatic, Afro-Asiatic, Hmong-Mien, Indo-Pacific, Yenisian, Niger-Kordofanian, Sino-Tibetan, Austroasiatic, and Australian.

> occasion, but from friends and intimates alone. [...] If pointing is best approximated by the mouth gesture of [expulsion] of air in direction of alter, then 2nd person would be expressed by explosive stops (less so: sibilants, fricatives). The opposite of the second person (1sg.) might then be the opposite of pointing. [...] The avoidance of violating the interlocutor's space may be best expressed *by nasals*: here air is expelled [...] but gently, and through the nose, the exhalation being thus not only filtered, but directed at the ground, rather than being directed vigorously, through the mouth, at the hearer. [...] It seems possible that the 2sg. nasal [in some of the 10 families] may have arisen as a kind of euphemism [...]. Rather than point to alter, the speaker may pretend that he is pointing to someone else [...], or not pointing at all, and *using a non-pointing nasal instead of an aggressive explosive stop.* (Haiman 2018: 197, 202, 209–210; emphasis in original)

In short, plosive consonants are iconically appropriate for expressing a "violation of the hearer's space" and hence for second person singular pronouns, because of their manner of articulation; nasal consonants are iconically appropriate to express the opposite idea, because their manner of articulation is somewhat opposite. This is of particular interest for the study of swear words: they are also a sign of familiarity with friends but potential aggression with strangers. And although Haiman (2018) has a different interpretation, his observations are also, again, coherent with Yardy's (2010) unsonority hypothesis for swear words: plosives are comparatively less sonorous than nasals, and vice versa.

Joo (2023) observes that a fourth of the words for staple food – e.g., "bread", or "rice" – in languages from 66 different families, start with labial consonant followed by a low vowel, resulting in onsets like /pa/, /ma/, /fa/, /wa/, etc., for example /pan/ ("bread") in Spanish, /fan/ ("cooked grain") in Mandarin Chinese, or /pɑːx/ ("bread") in Navajo. He attributes this cross-linguistic tendency to the iconic association between opening the mouth and the concept of eating.

While those studies deal with existing languages and words, iconic sound-meaning correspondences have also been observed for nonce-words. This is known in the literature as the *bouba-kiki* effect*:* in a *bouba-kiki* experiment, speakers are presented with a round shape and a spiky, angular shape, and asked to name one *bouba* and the other *kiki*; respondents are more likely to call the round one *bouba* and the angular one *kiki*. The effect is observable across languages and also for non-literate children, ruling out the possibility that the shapes of letters influence the results (Reilly, Biun, Cowles & Peelle 2008: 5; Yardy 2010: 6–12; Sidhu, Deschamps, Bourdage & Pexman 2019: 1595–1596). It could be argued that the interpretation of nonce-words is irrelevant to how sounds are interpreted in authentic words, but the *bouba-kiki* phenomenon shows at the very least how easy it is for speakers to agree on the attribution of the same sound to the same meaning, both across individuals, and across cultures, following unconscious correspondences. This goes some way in explaining why we can find so many cross-linguistic sound-meaning associations. Moreover, a bouba-kiki experiment ask speakers to attribute nonce-words (*bouba* or *kiki*) to meanings (the round shape or the angular shape). Based on this insight, we can imagine that speakers might be able to do so with their own creative input, i.e., to create new words from scratch to attribute to

given meanings. This is what we do in our third case study, where we ask speakers to playfully invent alien, i.e., extraterrestrial words for given meanings – see Section 3.3.

To summarise what the literature tells us, swear words are a kind of socially forbidden words who express and trigger powerful emotions, to the point that they are treated in the more emotional parts of the brain and provoke measurable physical reactions from speakers. They are often historically related to taboo subjects like religion, sexuality, or body waste, but their specific effect is separate from this taboo referential meaning, at least because the physical reactions are not as strong for L2 speakers or for "X-word" euphemisms, and because they are often not used with such referential meanings. Actually, their specific meaning compared to non-swear words is best understood as an emotional kind of meaning. This emotional meaning is not necessarily negative, even though swear words are usually seen as impolite, suggesting that they form a subregister, i.e., the extreme end of the informal register. Many swear words became swear words due to a historical process of social distinction, but not all of them.

It has been suggested that the sounds contained in swear words might have contributed to their interdiction. They seem to follow a sound pattern, a way to sound swearish that biased the interdiction process towards words with specific sounds. There are piecemeal observations and diverging interpretations, but the hypothesis that accounts most coherently for those observations is that unsonority, i.e., plosives, voiceless fricatives, and affricates are unconsciously associated with swearing. This is the basis for our two research questions (RQ1 and RQ2a) that our three case studies described in

Chapter 3 try to address.

Empirical research on sound-meaning tendencies shows that such unconscious sound-meaning associations can exist in language, although they run counter to classical tenets of arbitrariness and double articulation. They are found pervasively in interjections, onomatopoeia, and ideophones. These are categories of words with a more emotional meaning, closer to showing than telling, just like swear words. That sort of tendencies can also influence the interpretation of proper names. They are also detected in more run-of-the-mill items of the lexicon. Some associations, like phonaesthemes, seem to be language specific and have been proved to be cognitively real, but more attention has been devoted in previous research to cross-linguistic tendencies associating a specific sound of group of sounds with a specific meaning. If form-meaning pairings in language were completely unmotivated, then any given meaning could be expressed by any sounds in unrelated existing languages, but cross-linguistic tendencies show the opposite. Consequently, they are interpreted in terms of motivation, most often as the result of an unconscious sense of resemblance, i.e., iconicity, between the meanings and the corresponding sounds. Whether the sound-meaning association we hypothesise in English and French swear words is language-specific, or cross-linguistic and motivated, is beyond the scope of this thesis. In any case, the existence of such sound-meaning associations gives credence to the idea that this kind of sound-meaning association could happen in English and French swear words, i.e., that a group of sounds – unsonorous consonants – could be associated with an emotional meaning – the meaning of swear words.

If there is such an unconscious sound-meaning association in swear words, the next question to ask is what the meaning is, which is another one of

our research questions (RQ2b). To ascertain what specific emotional meaning might be shared by all swear words, we have to analyse the nature of swear words more deeply. By elaborating our own reflection based on the existing literature, and on some elements anticipating on our results on real-life swear words, we can propose a more refined definition of swear words. This is what we aim to do in the next section. Two arguments we give in this section also inform some of the methodological choices for our three case studies described in Chapter 3: namely, swear words form a prototype category, and this category is referred to by an everyday term like English *swear words* or French *gros mots* ("big words").

2.2 Swear words revisited

2.2.1 Defining swear words: basic issues and perspectives

The definition of swear words that we can derive from the existing literature (see Section 2.1.1) comes down to the following. Swear words are a kind of taboo, i.e., socially forbidden words that convey strong emotions, often with measurable physical reactions. This emotional arousal is not necessarily negative. In fact, the contextualized emotion they express and provoke is what distinguishes them from non-swear words, not their referential meanings. In that regard, they are similar to interjections or ideophones, closer to showing

than saying on a showing-saying meaning continuum. In the case of swear words, this strong emotional effect is, at least in part, a product of social norms, sometimes via a process of social distinction initiated by upper classes. Their strong emotional effect probably contributes to the self-perpetuating interdiction of swear words, i.e., we forbid swear words because they trigger strong emotions, and they trigger strong emotions because we forbid them.

The existing literature also suggests that there is an unconscious sound-meaning association between unsonority on the one hand and some meaning shared by swear words on the other hand (see Section 2.1.2). The next question then is what the meaning is (RQ2b). Answering that question requires to define even more precisely what swear words are.

First, it is important to make clear what kind of definition we are aiming for. We aim to define swear words in a way that is consistent with speakers" linguistic knowledge, i.e., their intuitive judgment of what counts as a swear word and what does not. We are thus following what Drummond (2020) calls a "common-sense approach" to swearing definition (2020: 3). This approach differs from other definitions like Ljung's (2011: ix) who restrict what counts as swearing to "non-literal" uses like interjections and intensifiers (Love 2021: 740–743). On the contrary in a common-sense approach, we aim to account for the fact that "someone saying "I fucked him", at a family meal for example, would most likely be seen [intuitively by speakers] as swearing" (Drummond 2020: 3). This example illustrates how in English, *swearing* refers to any use of specific lexical items called *swear words*, regardless of whether they are used as "non-literally" as interjections, intensifiers, or "literally" as nouns, verbs, adjectives, etc. As a native speaker, we consider that the categorisation expressed by the equivalent terms in French (*gros mots* for *swear words*, *dire des*

gros mots for *swearing*) functions in the same way.

A common-sense perspective like this also has the advantage that it can be more empirically based. To define what words should be categorised as swear words, we can collect data elicited from speakers. It has been pointed out in Section 2.1.1 that speakers may disagree on what counts as a swear word. That is only partially true, because for many items, speakers seem to largely agree on what counts and does not count as swear word: to take the same examples as before, it is certainly obvious to most native speakers of English that *fuck* counts as a swear word, and to native speakers of (European) French that *putain* and *merde* are *gros mots* ("swear words"). The presence of clear-cut, consensual examples vs. non-consensual examples suggests that swear words form a prototype category. Some words are considered by the overwhelming majority of speakers as examples swear words, and thus are the best examples of the category. Others items are not as consensual and might be considered as swear words by a smaller majority, or even only a minority of speakers. If we ask speakers to give examples of swear words, their most frequent answers will be the most consensual examples, closer to the prototype of the category, showing collective agreement. The least frequent answers will be the non-consensual examples. To get a list of French swear words, this is what we asked native speakers of French to do, i.e., to give us a few examples of swear words in our first case study described in Section 3.1. The elicited French data from our first case study confirm that swear words form a prototype category. Some items were spontaneously given by an overwhelming majority of respondents, and other items by fewer people, following a prototypicality cline. The question then from a methodological point of view is to choose a threshold: how frequent does a word have to be in the respondents" answers, that we can

reasonably categorise it as a swear word? We settle for two in our case study, i.e., any word given by at least two different speakers made it into our list of French swear words. This is not a controversial choice: Lev-Ari and McKay (2022) also "only kept words and phrases that were provided by at least two participants" (2022: 1105). It should be pointed out that this elicitation of data from speakers is made easier by the existence of an everyday term to refer to the category, like English *swear words, curses, profanity,* etc. or French *gros mots* ("big words"). Consequently, we simply use those terms to collect data in our first, second, and third case studies described in Chapter 3.

We use the linguistic knowledge of speakers on this prototype category of swear words, on their intuitive knowledge of what counts and does not count as a swear word. We try to account for the nature of that category in a coherent way, for what speakers collectively mean when they use a term like English *swear words* or French *gros mots* ("big words"). However, this does not mean that we make use of their meta-linguistic knowledge. In other words, we do not take into account what speakers would answer to a question like: "What are swear words?" Speakers' metalinguistic knowledge is notoriously unreliable and in our case, it seems that when speakers discuss swear words meta-linguistically, they can easily confuse them with the related overlapping categories of interjections and insults, as we will discuss further below.

Moreover, we aim to account for what speakers fundamentally mean when they use a term like *swear word*, but not necessarily in a specific culture or group of cultures. There are cultural differences, even between the two languages we investigated in our case studies, but we aim for a definition as cross-linguistic as possible, based on our English and French data, and on the cross-linguistic observations in the literature already mentioned, precisely

because swearing has been described as a cross-linguistic phenomenon. Regardless of cultural differences in how the phenomenon plays out, it can be accounted for cross-linguistically, and this is what we aim to contribute to here. Whenever the phenomenon of swear words exists in a language, our criteria should apply.

2.2.2 The taboo on words vs. the taboo on referential meanings

The referential meaning of swear words is less important than their contextual, emotional effect on speakers and listeners. In many of their uses, swear words do not refer to their referential meaning, for example someone interjecting *Shit!* is not referring, not even metaphorically, to any excrement; someone saying *That's fucking great!* is not referring, not even metaphorically, to sexual intercourse. Based on our intuitions as a native speaker of French, we would go even further and argue that actually, from a synchronic point of view, the referential meaning is fundamentally irrelevant to defining and identifying swear words. The cross-linguistic fact that swear words usually belong to semantic taboo domains like religion, sexuality, and body waste is an important but diachronic fact: it is not necessarily observable for all swear words at a given time. We can illustrate this with many French items from our first case study, anticipating on the results described in Section 3.1.2 and discussed in Section 4.2. For many prototypical swear words of French – i.e., words that strike French speakers as best examples of swear words, found among the most

frequently elicited answers of our list in Section 3.1.2 – their referential taboo meaning related to religion, sexuality, or body waste, has largely if not completely disappeared. Below we give a list of the six most clear-cut cases from our study (*putain* "Fuck!"/"fucking…", *con* "jerk"/"dumb", *bâtard* "bastard", *bordel* "mess", *enfoiré* "scumbag", *foutre* "to do"/"to make"/"to put") with detailed comments, of that semantic change phenomenon. The list is ordered by the frequency of elicitation in our French questionnaire described in Section 3.1. The historical, referential taboo meanings, along with the more recent meanings, can be found in the *Trésor de la langue française informatisé* (*TLFi* for short)[3] or the Larousse[4], two of the largest online French dictionaries.

The first word that illustrates the semantic shift away from a referential taboo domain is *putain* (loosely equivalent to the interjection *Fuck!* or the intensifier *fucking* (*X*)). It used to mean "whore", but is used mostly nowadays as a swearing interjection – *Putain!* meaning "Fuck!" – or swearing intensifier – for example, the novel title *Une putain d'histoire* (Minier 2015) roughly meaning "One hell of a story" or "A fucking story". We would argue that the "whore" meaning, precisely because it is more ancient, actually gives it a more formal connotation so that paradoxically, this meaning survives only in more formal contexts, e.g., the history book title *Le Soldat et la Putain* ("The Soldier and the Whore") (Benoit 2013), the French title for a documentary called *Whores' Glory* (Glowagger 2011) translated as *La Gloire des Putains* ("The Glory of Whores"), or fixed phrases like *la maman et la putain* ("the mum and the whore"), used to describe and criticise binary gender roles for women, probably made popular by the title of 1973 French New Wave film *La Maman et la Putain*

3 Trésor de la langue française informatisé: http://atilf.atilf.fr/

4 Larousse: https://www.larousse.fr/dictionnaires/francais

(Eustache 1973) – for an example of a journalistic use of this phrase, see Chollet (2012). This intuitive claim would need to be further confirmed by corpus data in addition to our questionnaire data, but *bordel* ("fuck!"/"brothel") and *gueule* ("mouth"/"face") (see below) are other arguable examples of this discrepancy between an ancient, rarer meaning in more formal settings vs. a new and more frequent meaning in less formal settings.

Another example is *con* (loosely equivalent to *jerk*, *(fucking) idiot*, *stupid*, *dumb*). It used to mean "vagina", so that it was an equivalent of English *pussy* or *cunt*, but is now used as a swearing noun meaning something like "(fucking) jerk" or "(fucking) idiot", or a swearing adjective meaning something like "(goddamn) stupid". Based on our elicitation data aligning with our own intuitions, we can safely state that only the swear word *chatte* ("pussy"/"female cat") has that sexual meaning in today's French, and is today's equivalent of English *pussy* or *cunt*.

A third example is *bâtard* ("bastard"). Both *bâtard* and its English equivalent *bastard* have become a swearing insult devoid of precise referential meaning, least of all their past sex-related meaning of "male child born out of wedlock" which survives only in accounts on history, or historically inspired fiction like the television series *Game of Thrones* (Benioff & Weiss 2011–2019), where protagonist Jon Snow often reminds other people that he is "a bastard", i.e., he was born out of wedlock. The loss of its referential character is nicely illustrated by the humorous effect of Jon Snow saying "I'm a bastard" at a present-day dinner in a late night show skit (Late Night with Seth Meyers 2015). The French equivalent term *bâtard* was used to translate this joke in French online news (Première 2015) showing that the meaning has evolved in the exact same way in English and French.

A fourth example of the loss of referential taboo meaning is *bordel* (loosely equivalent to the interjection *Fuck!* or to *mess*). It used to mean "brothel", "prostitution house", but is now mostly used as a swearing interjection (*Bordel!* translatable as *Fuck!* or *Damn!*), or as a noun meaning "(fucking) mess" as in the fixed phrase *foutre le bordel* ("to make a fucking mess"). The legal interdiction of brothels in France certainly contributed to the loss of its prostitution-related meaning, which survives typically in journalistic accounts of prostitution in the past – see for example Valeix (2023) or Radio France (2023) – or in foreign countries – see for example Philip (2018) about Thailand.

Another word from our dataset that illustrates the irrelevance of the historical taboo meaning is *enfoiré* (loosely equivalent to insults like *scumbag, bastard, asshole*). It allegedly had a homophobic and body waste meaning, namely "homosexual man, covered with feces after anal sex" (Rouayrenc 1996: 43, 107). Yet this claim about the meaning of *enfoiré* seems to originate from a 1931 slang dictionary written by a non-linguist (Chautard 2013 [1931]), and to our knowledge, this usage has not been confirmed by corpus data or contemporary testimonies. What is certain from our synchronic elicited data and again, in line with our own native speaker intuitions, is that even if this word may have had a referential taboo meaning, it has now been completely lost. *Enfoiré* is now simply a swearing insult, quite similarly to *bâtard* ("bastard") or its English equivalent *bastard*.

Another word from our dataset that lost the historical taboo meaning is *foutre* (loosely equivalent to *(fucking) do*, *(fucking) make*, *(fucking) put*) and its derived forms. *Foutre* used to mean "to fuck", but is now a swear word meaning "to do", "to make" or "to put". We can illustrate by some examples.

First, in the American film *Pulp Fiction* (Tarantino 1994), when a character crashes a car into a house's garden, *Have you lost your fucking mind?* was translated in the French dub as *Mais qu'est-ce que tu fous, merde ?* (which translates literally as "Just what are you fucking doing, shit?" or more idiomatically "Shit, what the fuck are you doing?"). Later in the same film *Where the fuck is it?* uttered by a character looking for a lost watch, was translated in the French dub as *Où est-ce que tu l'as foutue ?* (possible reverse translation "Where did you fucking put it?" or more idiomatically "Where the hell did you put it?"). Conversely, in the French film *Asterix and Obelix take on Caesar* (Zidi 1999) the sentence *Qu'est-ce qu'ils foutent ?* ("What are they fucking doing?" or "What the hell are they doing?") was translated in the English dub as *Where the hell are they?* In none of these examples is there any reference to the more ancient meaning of that swear word.

We can see further evidence for this semantic shift of *foutre* in the usage of its past participle adjective form, which is *foutu* (masculine form) or *foutue* (feminine form, same pronunciation). It often means "broken" or "over, finished", for example *C'est foutu!* could be translated as *It's over, goddamnit!* It can also mean "made", for example the fixed phrases *bien foutu* and *mal foutu* could be translated respectively as *well (fucking) made* or *badly (fucking) made*. *Foutu/foutue* is also used, just like *putain de*, as an intensifier, for example in the same Pulp Fiction scene (Tarantino 1994), *Of all the fucking things she could fucking forget, she forgets my father's watch!* was translated in the French dub as *Il y avait une chose à laquelle il fallait faire gaffe, c'est cette foutue tocante !* (literally translates as *There was one thing to pay attention to, it was this goddamn watch!*).

Finally, many fixed expressions with *foutre* exist where the original meaning is equally lost. We already mentioned the fixed phrase *foutre le bordel*

("to make a fucking mess"), but there are other (partly) fixed items, like *s'en foutre* ("to not give a fuck/damn"), i.e., a swearing equivalent of *s'en moquer* ("to not care"), or *foutre le camp* (literally "to (fucking) do the camp", meaning "to leave" and translatable as *to get the fuck/hell out*). Despite the loss of this referential meaning, there is one specific and rare context in which *foutre* retains a sexual meaning, i.e., the context where it is used as a noun, meaning "semen" and hence equivalent to the English noun *cum*.

As is clear from the above discussion, six swear words in our elicitation experiment (*putain*, *con*, *bâtard*, *bordel*, *enfoiré,* and *foutre)* have largely if not completely lost any referential meaning related to taboo semantic domains. Their high position in the elicitation list clearly suggests that they are considered as belonging to the most prototypical swear words of French by contemporary French speakers. In other words, when a swear word loses its taboo referential meaning, it does not lose its swear word status at all. This process is not limited to the most prototypical swear words, but also concerns less salient, derivated swear words, such as *emmerder* ("to annoy"). Despite being derived from *merde* ("shit"), which can refer to excrement just like the English equivalent *shit*, and despite its historical meaning "to cover in shit", *emmerder* is never used in today's French with a meaning related, literally or metaphorically, to excrement. Its only possible referential meaning is "to annoy", but it has nevertheless retained its status as a swear word.

So swear words can lose any referential meaning referring to taboo topics like religion, sexuality, or body waste, and yet clearly remain swear words, expressing and triggering strong emotional reactions. *Bastard*, already mentioned above, is a possible example for English, suggesting that such semantic loss of taboo domains is not a French-specific phenomenon, even if it

seems more advanced and pervasive in French swear words. A definition of *swear words* including reference to taboo semantic domains as a necessary criterion would fail to describe the present situation in French, and potentially in other languages where the same historical dynamic has played.

Even from a diachronic perspective, we find arguable exceptions: *gueule* ("(damn) mouth"/ "(damn) face"), often used in the fixed phrase (*ferme*) *ta gueule* ("shut your damn mouth"/ "shut the fuck up"), does not refer and never referred to a taboo semantic domain like religion, sexuality, or body waste. It is probably related to a contextual, not referential taboo: *gueule* also means "mouth (of an animal)", in which case it is not considered a swear word, as in the fixed phrase *se jeter dans la gueule du loup* (literally "to throw oneself into the wolf's mouth", equivalent to English *to walk into the lion's den*). *Gueule* must have gained its taboo status from the shocking act of comparing human beings to animals, but animals can hardly be considered a taboo semantic domain in the same way as religion, sexuality, and body waste. Comparing humans with animals is reportedly a prolific source of swearing insults in Korean (Bergen 2016a: 28).

Referential meanings related to taboo topics are likely for swear words, especially from a diachronic perspective, but not an obligatory criterion, especially from a synchronic perspective. We should therefore abandon reference to taboo semantic domains as a relevant synchronic criterion for defining and identifying swear words. They should be defined and identifiable irrespective of their historical origins. Other historical dynamics that contributed to the interdiction of swear words, like McEnery's (2006: 226–227) process of social distinction, should also not be used as a necessary criterion, for the same reason it may not concern the present-day value of those words in the

speakers' minds.

2.2.3 A word on interjections and insults

As mentioned above, swear words are easily conflated with interjections and insults – see for example Wajnryb (2005: 223–228) – because they are overlapping categories. The following examples illustrate the different origins of these confusions and aim to dispel them.

While many swear words are interjections (e.g., *Fuck!, Shit!, God!, Damn!*) many interjections are not swear words, for example *Wow!* is an interjection, but it is certainly not a swear word. Moreover, many swear words cannot be used as interjections: it is safe to assume that speakers of English would never or rarely use *Ass!*, *Piss!*, *Cock!* or *Cunt!* as interjections.

Secondly, many swear words are insults, e.g., *bastard*, *fucker*, *wanker*, but many are not, e.g., *goddamn*, *crap*, or *tits*. There is another metonymical link between swearing and insults: swear words are considered by default as impolite, and being impolite can be seen by default as insulting. Therefore, swear words can be intuitively felt by speakers as by default insulting, and count as insults. Here are however examples that we hope can show the contradiction in that rationale. An utterance like *Damn, that was fucking great!* contains two swear words, but it can hardly count as insulting. On the contrary, *idiot* or *crook* can be grave insults, and yet are certainly not swear words. Interjecting (e.g., *Fuck!*, *Damn!*) is a common function of many swear words, but they are not insulting by definition (e.g., *Wow, that is great!*).

To illustrate these two points, a sentence like *Wow, you are such an*

idiot! contains an interjection (*Wow*) and an insult (*idiot*), but no swear word. The confusion is also understandable because swear words, due to their emotional value, intensify the effects of interjections and insults, so that *Fuck, you are such a dipshit!* is more powerful than *Wow, you are such an idiot!* although the referential meaning is equivalent.

2.2.4 Defining swear words with criteria

In this section, we propose four criteria to define swear words. None of these is sufficient in and by itself. The more criteria are met by a word, the more likely it is to be a swear word. The first two criteria and tests are mostly based on the literature reviewed in Section 2.1.1, while the last two are mostly based on our own reflection.

The first criterion is that swear words are socially forbidden words. It is a strong interdiction that is however context-dependent loosens in a few contexts, typically between friends or colleagues. That interdiction is often not usually made explicit, but sometimes it is made explicit and even legal, for example swear words on American television and radio are legally forbidden and hence censored, typically with loud bleeps (Bergen 2016a: 235–238; Pinker 2008: 205, 215, 230). We also propose that the context where swear words are most strictly forbidden, is when uttered by children, or by adults in the presence of children. The first criterion to be considered is thus whether the word is censored in numerous social contexts, such as on television or in front of children.

The second criterion is that swear words express and trigger a strong

emotional arousal, either negative or positive depending on context, that can be physically measured. That emotional arousal makes them likely to be pronounced by some people with aphasia or by people with coprolalia, and to create unease or laughter, especially when used by or in front of children.

The third criterion to determine whether a word is a swear word is whether native speakers list it when asked for *swear words* or any other equivalent everyday term in their language, like *gros mots* ("big words") in French. This is the criterion that we used for our case studies and turns out to be of great help when aiming for a cognitively real categorisation and conceptualisation of swear words. The more native speakers think spontaneously of a word when asked to give S, or agree with categorising it as an S – where S is the proposed everyday term for swear words in that language – the more justified it is to consider that this word is a swear word of the language, because it is closer to the prototype of the category.

The fourth criterion to determine whether a word is a swear word is that it is itself forbidden and not for its referential meaning, so there has to be a referentially equivalent, non-swearing alternative. It serves to distinguish between words that are taboo merely because they refer to a taboo subject, and swear words. It accounts for how *penis* and *excrement* are not swear words, while *prick* and *shit* are. Similarly, words referring to shocking, emotional, potentially triggering taboo subjects, like *cancer* or *rape*, are not easily used because of their reference to unpleasant things, and therefore can be considered by some speakers to be taboo words – see for example Sulpizio et al. (2024) – but they are unlikely to be seen as swear words, as there is no equally widespread social interdiction requiring more acceptable alternatives.

This points to a core issue: the taboo on a swear word is a social

taboo concerning the word itself, not its referential meaning, regardless of how that meaning may have contributed to the interdiction in the past. Historically, the interdiction may originate in part from their belonging to taboo domains like religion, sexuality, or body waste, but over time the taboo has become autonomous from that meaning and now applies to the word itself. Even the interdiction of swear words that still refer to taboo domains can be seen as an intermediate situation between two extreme alternatives: not talking about a taboo topic at all, or talking about a sensitive topic with no restriction whatsoever. Therefore, the interdiction of swear words appears to be a social compromise to deal with taboo topics. Some speech communities are close to the first extreme, i.e., speakers talk so little about a sensitive topic that there are no neutral, acceptable words to refer to those taboo concepts. For example, in the Australian language Dalabon, the words for taboo concepts tend to be figurative, and are only used in very controlled and restricted contexts, with no acceptable alternative words or phrases (Ponsonnet, personal communication). The absence of a neutral alternative clearly suggests that we are dealing with a very strong taboo on the referential meaning, not with an autonomous taboo on the word itself. The words in question should not be considered as "swear words" of Dalabon, but as words referring to a taboo subject.

The first and fourth criteria require some level of intuitive common sense. If we apply them too extensively, any word or phrase that belongs to an informal register would qualify as a swear word, e.g., *cop* would count as a swear word because it is socially frowned upon in some contexts and replaced by *policeman*, or *don't* because it is sometimes socially forbidden and replaced by *do not*. These are requirements of register. The interdiction of swear words and requirement for alternatives is a rule that is much more widespread, much

clearer and more intuitive to speakers. It applies more strictly and to more contexts, not to one or a handful of specific contexts. Swear words belong to the informal register, so it makes sense that those criteria somehow apply to informal words too. Yet as already mentioned in Section 2.1.1, swear words are the extreme end of the informal register, so there is a clear difference of degree in the severity of the interdiction and the number of contexts where it applies. This is the nuance we try to encapsulate when we mention a "strong" and "widespread" interdiction loosening "only in a few contexts" when defining the first and fourth criteria.

Another useful test to distinguish swear words from the rest of the informal register is their number. Based on our dataset of elicited swear words (71 for English, 78 for French), we can hypothesise that in a given language, swear words should consist of at most a few dozen words. Why this is so remains at this point an open question. One possible explanation may reside in the higher cognitive effort to remember a strong social interdiction on several hundreds of items.

The third criterion is the most immediate, easily applicable and reliable one: the more respondents report a word as S (= the everyday term in the language for swear words), the more reliably it can be classified as a swear word. The meaning of this everyday term can vary slightly from language to language though. We address this issue for English and French and what it means for a cross-linguistic conceptualisation of swear words, and for our methodological choices for our case studies, in the next section.

2.2.5 English *swear words* and French *gros mots*

So far, we have used the English term *swear words* and its French equivalent *gros mots* ("big words") interchangeably. There is, however, a difference in categorisation between English and French, which we accounted for in our protocols. It is therefore warranted that we briefly discuss this issue here.

According to Bergen (2016a: 16) and Finkelstein (2018: 326), the English *swear words* category includes slurs, i.e., bigoted insults against a dominated group of people, like the anti-gay slur *fag*. Bergen (2016a) however notices that "not everyone agrees that slurs are profanity [RV: swear words] — for some people, *nigger* is a swear word, whereas for others it falls into a distinct category of taboo word" (2016a: 16). He himself includes slurs in the category of swear words, but also acknowledges that "slurs have the greatest potential to cause harm and therefore demand different treatment." (ibid.)

The situation is markedly different, and more consensual across speakers, when it comes to the categorisation of slurs and *gros mots* ("big words") in French. French has no equivalent everyday term for *slur*. More precisely, it only has terms for subcategories of slurs: racial slurs are called *insultes racistes* ("racist insults"), homophobic slurs are called *insultes homophobes* ("homophobic insults"), etc.

We propose that the category *gros mots* ("big words") includes slurs related to sexuality such as *pédé* ("fag") or *pute* ("whore"), but it excludes those that are not related to sexuality, such as racial slurs like *nègre* ("nigger") or *bougnoule* (an anti-Arab slur). For the sake of brevity, we call them *sexual* vs. *non-sexual slurs*. This distinction comes from our intuitions as a native speaker of French. Anticipating again on our results, we can say that this distinction is

borne out by the data from our Real-life swear words study described in Section 3.1. We asked native speakers of French to list the *gros mots* ("big words") they knew via seven thematic questions, and some insults they knew in two other questions. Our intuitive categorisation explains our collected French data much more efficiently than alternative categorisations, as we discuss in Section 4.2. We illustrate this French categorisation with Figure 1. Since *gros* ("big") in *gros mots* ("big words") comes historically from *grossier* ("rude") (Rouayrenc 1996: 3–6), we propose that the *gros mots* ("big words") French term might be best translated as *rude swear words*, and this categorisation can be applied to English as we do in Figure 2 below. It can also partly explain Bergen's observation mentioned above that some English speakers would not count racial slurs as swear words: maybe they categorise swear words in the same way as French speakers, and would apply the term *swear words* only to the category we call *rude swear words* in Figure 2.

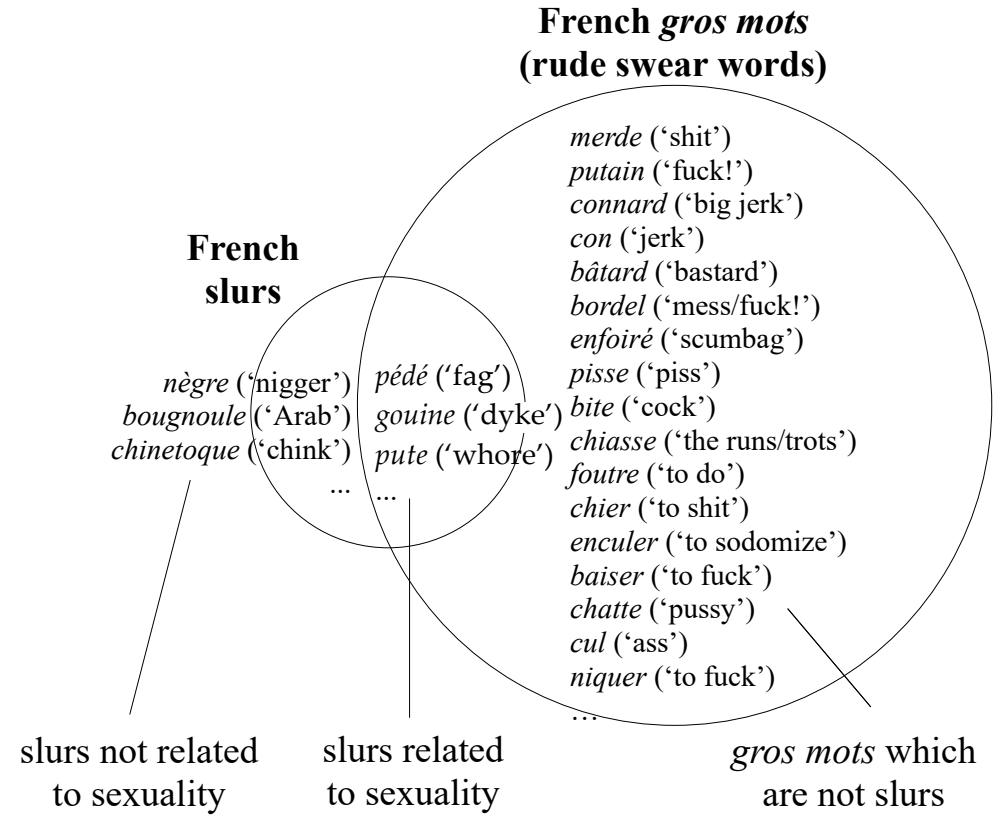

Figure 1: Categorisation of French swear words and slurs

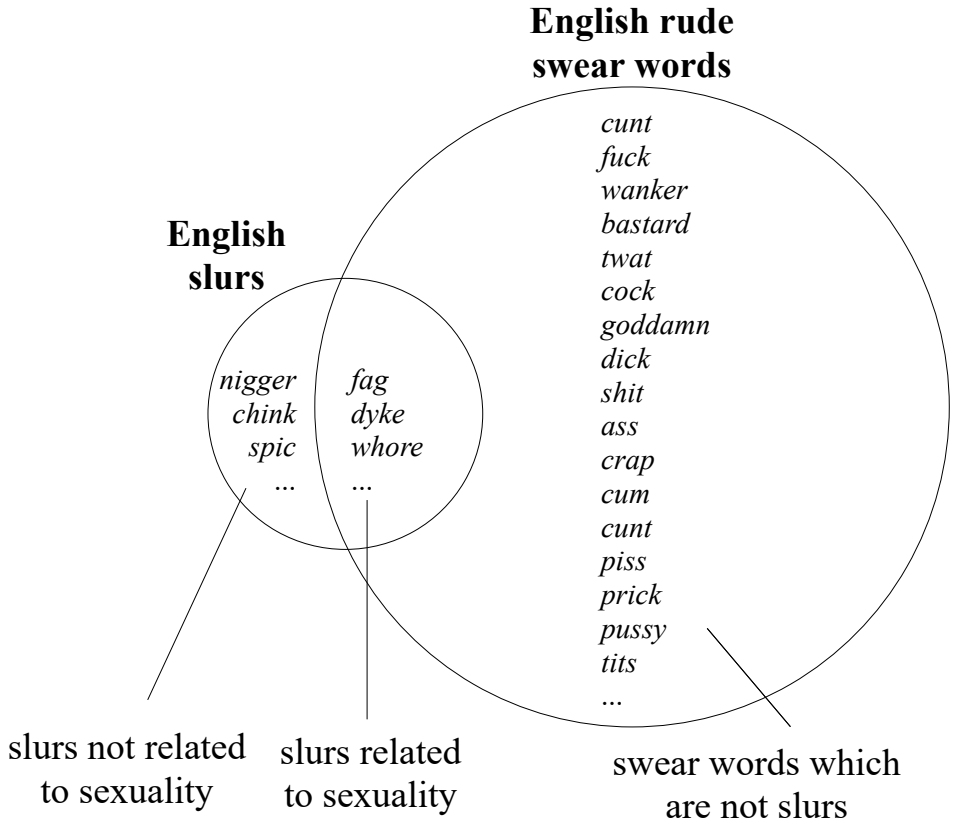

Figure 2: Categorisation of English swear words and slurs

Some further observations can be made with regard to the fact that the *gros mots* ("big words") category overlaps with, but does not include all slurs. First of all, it suggests that, even if the term *swear words* includes both categories according to some speakers of English, we are actually dealing with two different kinds of taboos. Rude swear words are forbidden by politeness, slurs are forbidden by anti-bigotry, egalitarian ethics. Yet there is an overlap, namely sexual slurs like *fag* or *whore*: they are forbidden by both politeness conventions as well as anti-bigotry ethics. This complicates the situation and creates ambiguity, and poses new research questions for further avenues that we consider in our conclusion in Chapter 5.

Despite this slight difference in categorisation, our four criteria apply to some slurs in French – only the third criterion does not apply to non-sexual slurs – and they apply to all slurs in English according to some speakers. For that reason, and if only for the sake of exhaustivity, we collect any slurs regardless of their sexual or non-sexual meanings, in our three empirical studies described in Chapter 3.

This shows however how the everyday S term for swear words can reflect slightly different categorisations in different languages. We take that into account when exploring the cross-linguistic viability of our definition in the next section.

2.2.6 The swear words category across languages

We proposed a cross-linguistic definition for swear words in Section 2.2.4. We use that definition to investigate the meaning of swear words – discussed in 4.3 – present in our sound-meaning association. This section offers a first exploration of the cross-linguistic viability of our definition, based on other studies and insights from fellow linguists who happen to be native speakers of languages with swear words. This is a first sketch that still needs to be expanded systematically, but the first observations are reassuring. We also mention a specific case where researchers disagree on the existence of swear words in a given language, namely in Japanese, and how that case illustrates the need for a common cross-linguistic definition.

Insights from a fellow linguist and native speaker of Quebecois French suggest that at least three of four of our criteria apply: a few taboo

words, maybe six or seven, are not pronounced in front of children (Laliberté, personal communication), there is an everyday term to refer to them which is *sacres*, and they are forbidden regardless of their referential meaning so that there are non-swearing alternatives. There is however a marked difference in categorisation compared to English and European French: the word *sacres* refers only to the Quebecois words that are historically related to religion, e.g., *tabarnak* ("tabernacle"), *ostie* ("Host"), *câlice* ("chalice"), or *crisse* ("Christ"). The label is easily recognisable as religious – *sacres* is a cognate of *sacré* ("sacred") – and does not apply to non-religious swear words like *fuck* (borrowed from English) which Quebecois French speakers would categorise as *gros mots* ("big words" "swear words") or *jurons* ("swear words") instead. Moreover, applying the label *sacre* to non-Quebecois religious swear words – for example saying in Quebecois French that *Oh my God!* or *goddamn* are English *sacres* – would be somewhat acceptable only as a figure of speech that requires some interpretation that the speaker is actually talking about non-Quebecois equivalents of *sacres* (ibid.). This shows that linguistic communities can conceptualise categories of swear words that are, compared to English *swear words* or European French *gros mots* (rude swear words), more restrictive and culturally specific, even if they otherwise correspond to our definition.

Italian has swear words that satisfy our four criteria. They are referred to as *parolaccia* ("bad word", where *parola* means "word" and *-accia* is an evaluative suffix for something negative or bad). Similarly to French, this category would not include non-sexual slurs (Mairano, personal communication).

Mandarin Chinese has rude swear words that satisfy our criteria: they are strongly forbidden and censored, they express and trigger strong

emotions like laughter for native speakers, there is an everyday Mandarin Chinese phrase to refer to them, i.e., *cū huà* ("big words", "vulgar words"), and they require alternatives, since speakers replace them with euphemistic plays on words based on tone – i.e., the euphemism is identical to the swear word it replaces, except for a difference in tone (Hu, personal communication; Sha, personal communication). From a pilot study replicating our French swear word collection for Mandarin Chinese, we can say that the taboo on these words is so strong that compared to French, fewer Mandarin Chinese speakers are comfortable with listing the swear words they know, even in meta-linguistic use for a research questionnaire (Hu, personal communication). It makes the empirical collection of Mandarin Chinese *cū huà* ("big words", "vulgar words") harder, and combined with Bergen's (2016a: 31–32) opinion that the taboo on swear words is looser in European French than in American English, this suggests that the taboo on European French swear words is on the looser scale of the spectrum than Mandarin Chinese and American English. Finally, Mandarin Chinese speakers do not provide any slurs when asked for them, not even sexual slurs, suggesting that the *slurs* category has no actual equivalent in Chinese culture (ibid.).

In Russian, there is also an everyday term, *mat,* to refer to socially forbidden words – related to the domain of sexuality (Bergen 2016a: 34) – that provoke strong emotions like unease when speakers overuse them (Tsikulina, personal communication). Similarly to Mandarin Chinese, Russian does not include bigoted insults in the *mat* category at all; such insults are not discussed and made taboo, so that the *slurs* category has not appeared in Russian culture either (ibid.).

Japanese is a controversial example that illustrates the need for a

clear cross-linguistic definition of swear words. Bergen (2016a: 33) claims, quoting a native speaker in a news interview, that Japanese has no swear words: "We don't really have curse words in Japanese, so I like the fact that the Western languages allow me to say things that I otherwise can't." Kosugi (2012) takes a similar point of view, calling Japanese a "swearless" language. However, a completely opposite view is taken by Wajnryb (2005) who argues at length about how sceptical she is that Japanese, or any language for that matter, can be truly devoid of swear words:

> It has been said that some cultures don't swear at all. The list usually consists of the Japanese, Eskimos, Malayans, Polynesians, and Native Americans. "Don't swear at all" I take to mean as lacking native swear words, a concept that I find quite baffling, even counterintuitive. […] Take the case of the Japanese, whom a number of writers on swearing have claimed don't swear, in a language with no swearwords. Certainly, many Japanese are loath to admit that they swear or that their language has a swearing vocabulary. If pushed, they may admit to the existence of *warui kotoba* ("a bad word"), then they'll urge you never to use it. And this reticence is widespread. One of my informants, an Englishman married to a Japanese woman, asked his wife the questions I was using to elicit data about Japanese. She told him she couldn't help because she didn't know any Japanese swear words. This she said, mind you, in wide-eyed innocence to a husband who was fully aware, as she was aware that he was, from firsthand experience of her skills in that department. (2005: 218, 223)

Pinker (2007) agrees with her and develops further:

> Claims that profanity is lacking altogether in a particular language have to be taken with a grain of salt. It's true that in many places if you ask speakers to list their profanities, they may demur. But swearing and hypocrisy go hand in hand, to the extent that some personality questionnaires include items like "I sometimes swear" as a check for lying. (2007: 207)

Robertson et al. (2017) have an intermediate opinion: they report that Japanese is often described as devoid of swearing, but consider one word to be a swear word: *kuso*, often used as an interjection and translated as "shit!" That description is coherent with Van Lacker & Cummings (1999: 89) who report words produced by Japanese speakers with coprolalia: we find uses of *kuso* ("shit!") or derivatives, but other words seem to have rather benign meanings, such as "ugly" or "fat", i.e., it seems that they are used in taboo acts of talking, but the words themselves are not taboo (Bergen 2016b). It is possible that Japanese has few or no swear words because it is in a situation similar to Dalabon described above in Section 2.2.4, i.e., taboo meanings are so taboo that no word is truly acceptable to express them, so no autonomous taboo on the words themselves has developed, and hence there is no true *swear words* category.

It is beyond the scope of this thesis to determine whether Japanese has swear words and how many, but it would certainly be a relevant practical case to investigate by applying our proposed criteria and tests. We know at least part of the answer to our second criterion. Robertson et al. (2017) used *kuso* to investigate whether swear words have the same pain relief effect for Japanese

speakers as English speakers. Their results show that *kuso* does seem to have the same pain relief effect. This suggests that Japanese has at least one swear word, but as long as the four criteria are not answered with the help of native speakers for every proposed Japanese swear word, the jury is still out as to whether and how many swear words exist in that language.

We can give the following general insights based on those first tentative cross-linguistic comparisons. The four criteria we propose apply at least to English, European French, Quebecois French, Mandarin Chinese, and Russian. *Swear words* is a cross-linguistic, but not universal conceptual category, some languages are deprived of it (Dalabon), or at least seem largely deprived of it (Japanese). The exact boundaries of the category speakers referred to with the equivalent everyday term for *swear words* are not the same across languages, for example English *swear words* include all slurs according to some speakers, and Quebecois *sacres* include only Quebecois swear words historically related to religion. The taboo on swear words is a relatively strong taboo, but it can be more or less strong depending on the language, for example the taboo in European French is weaker than in American English or Mandarin Chinese taboo. The *slurs* category is not universal: a taboo on bigoted insults has not necessarily developed in every culture that otherwise has swear words. A lack of clear definition and criteria for swear words can lead to confusion and contradictory claims about their existence in a given language, as in the case of Japanese.

We now have a more refined conceptualisation of what swear words are, that seems to be reliably applicable across languages. To recapitulate, these words are the target of a strong social interdiction, applying most strictly to or in front

of children. They express and trigger emotional arousal. They form a prototype category referred to with an everyday term like *swear words.* The taboo is on the word itself instead of its referential meaning, so that they are replaced with non-swearing alternatives. As already mentioned, this conceptualisation allows us to discuss the meaning of swear words (RQ2b), involved in our hypothesised sound-meaning association. We discuss that meaning in Section 4.3. This definition also informs some of methodological choices for our three case studies, described in the next chapter: we use the everyday term to elicit real-life French swear words and collect any word given spontaneously by at least two respondents, in order to answer our research question (RQ1) on the statistical existence of a sound tendency in swear words. Then we use the everyday term again to collect fictional swear words and elicit experimental swear words, in order to answer our second research question (RQ2a) about the existence of an unconscious sound-meaning association in speakers' minds.

Chapter 3. Case studies on swear words

We conducted three studies to answer whether there is a sound tendency in English and French swear words compared to the regular lexicon (RQ1), and whether it corresponds to a cognitively real pattern in the minds of speakers (RQ2a). The first study consisted in trying to replicate Bergen's claim for a sound tendency in English swear words and check if it applied equally to French swear words. It did not prove conclusive for French, so in the same study we checked for an alternative sound tendency in swear words that is common to both languages. The goal of our second and third studies was to find evidence that this statistical tendency corresponds to a cognitively real pattern: if it is real, then speakers would follow the pattern unconsciously when they select existing words to serve as swear words in fiction, or when they create fictional swear words from scratch. Our second study consisted in checking whether such fictional swear words from fiction in English or French followed the tendency. Our third study consisted in asking native speakers of English and French to spontaneously invent swear words and non-swear words of an alien, i.e., extra-terrestrial language during an experiment, and then check if those experimental swear words followed the same tendency, compared to the experimental non-swear words.

3.1 Case study 1: Real-life swear words

In this first study, we investigate Bergen's three claims (1) that compared to regular words, English swear words are more often monosyllabics, (2) that compared to monosyllabic regular words, English monosyllabic swear words end more often with a consonant, and (3) that compared to regular words, English swear words are more often closed monosyllabics. We then investigate the equivalent claims for French swear words, and our alternative hypothesis about a sound tendency in English and French swear words, i.e., that they tend to contain more of the least sonorous consonants.

3.1.1 Method for Real-life swear words

In this first study, we used swear word lists to compare English swear words to the regular English lexicon, and French swear words to the regular French lexicon, drawing on existing lists for English while building our own for French.

Bergen (2016a: 15–25, 52–7) claims English swear words tend to be closed monosyllabics. He bases that claim on a list of English swear words compiled from separate studies on American English (Janschewitz 2008; Bergen 2016a: 21–22), British English (Millwood-Hargrave 2000), and New Zealand English (Broadcasting Standards Authority 2010). In those studies, researchers list potential swear words, then ask speakers to evaluate them, checking whether they consider them to be swear words. We combined the top results reproduced in Bergen (2016a: 15–25) to get a list of 71 English swear words,

given below in Table 1.

English swear word	IPA	Is a slur or not	Frequency (MRC)
anal	/ˈeɪnəl/	-	0
arse	/ˈɑːs/	-	NA
arsehole	/ˈɑːshəʊl/	-	0
ass	/ˈæs/	-	2
asshole	/ˈɑːshəʊl/	-	0
balls	/ˈbɔːlz/	-	0
ballsack	/ˈbɔːlsæk/	-	NA
bastard	/ˈbɑːstəd/	-	3
bitch	/ˈbɪtʃ/	slur	0
bloody	/ˈblʌdi/	-	42
blowjob	/ˈbləʊdʒɒb/	-	NA
bollocks	/ˈbɒləks/	-	1
bugger	/ˈbʌgə/	-	3
bullshit	/ˈbʊlʃɪt/	-	0
buttfuck	/ˈbʌtfʌk/	-	NA
chink	/ˈtʃɪŋk/	slur	0
clit	/ˈklɪt/	-	NA
cock	/ˈkɒk/	-	3
cocksucker	/ˈkɒkˌsʌkə/	slur	NA
crap	/ˈkræp/	-	0
cum	/ˈkʌm/	-	NA
cunt	/ˈkʌnt/	-	0
dick	/ˈdɪk/	-	0

English swear word	IPA	Is a slur or not	Frequency (MRC)
dickhead	/ˈdɪkhed/	-	NA
dumb	/ˈdʌm/	slur	0
dyke	/ˈdaɪk/	slur	0
fag	/ˈfæg/	slur	0
faggot	/ˈfægət/	slur	0
fuck	/ˈfʌk/	-	1
fuck off	/ˈfʌk ɒf/	-	NA
gay	/ˈgeɪ/	slur	1
God	/ˈgɒd/	-	73
goddamn	/ˈgɒdæm/	-	NA
gook	/ˈguːk/	slur	NA
homo	/ˈhəʊməʊ/	slur	0
hooker	/ˈhʊkə/	slur	0
Jesus	/ˈdʒiːzəs/	-	5
Jew	/ˈdʒuː/	slur	0
kike	/ˈkaɪk/	slur	NA
lesbo	/ˈlezbəʊ/	slur	NA
loser	/ˈluːzə/	-	1
moron	/ˈmɔːrɒn/	slur	0
motherfucker	/ˈmʌðəˌfʌkə/	-	NA
nigger	/ˈnɪgə/	slur	0
nutsack	/ˈnʌtsæk/	-	NA
paki	/ˈpæki/	slur	NA
piss	/ˈpɪs/	-	0
piss off	/ˈpɪs ɒf/	-	NA

English swear word	IPA	Is a slur or not	Frequency (MRC)
pissed off	/ˈpɪst ɒf/	-	NA
prick	/ˈprɪk/	-	1
pussy	/ˈpʊsi/	-	0
queer	/ˈkwɪə/	slur	0
retard	/ˈriːtɑːd/	slur	0
rimjob	/ˈrɪmdʒɒb/	-	NA
scum	/ˈskʌm/	-	1
shag	/ˈʃæg/	slur	0
shit	/ˈʃɪt/	-	3
shithead	/ˈʃɪthed/	-	NA
skank	/ˈskæŋk/	slur	NA
slag	/ˈslæg/	slur	0
slut	/ˈslʌt/	slur	0
sodding	/ˈsɒdɪŋ/	-	0
sodomize	/ˈsɒdəmaɪz/	-	NA
spastic	/ˈspæstɪk/	slur	3
spic	/ˈspɪk/	slur	NA
tit	/ˈtɪt/	-	0
tits	/ˈtɪts/	-	0
twat	/ˈtwɒt/	-	NA
wank	/ˈwæŋk/	-	0
wanker	/ˈwæŋkə/	-	NA
whore	/ˈhɔː/	slur	4

Table 1: List of 71 English swear words by alphabetical order.

For French, we had to build an unbiased list of swear words from scratch. We

did so by asking native speakers of French to list the swear words they knew, via an anonymous online questionnaire with nine thematic questions.

Two methodological points deserve to be mentioned regarding French. First, there is no everyday French term exactly equivalent to *swear words* as defined earlier. The closest equivalents are *jurons* ("swear words"), which is quite old-fashioned in contemporary French, and *gros mots*, which historically meant "rude words" (Rouayrenc 1996: 3–6) but can be literally translated as "big words". We used the term *gros mots* in the questionnaire because it is more commonly used than *jurons*. Contrary to English *swear words*, the French category called *jurons* or *gros mots* clearly excludes slurs unrelated to sexuality, like racial slurs, as discussed in Section 2.2.5 above. For that reason, we asked for slurs in a separate question. A second methodological point, related to the same question, is that French has no exact equivalent for the term *slurs*. This is why in Question 8 of our questionnaire, we asked respondents for slurs by using a paraphrase ("insults that must not be said because they target a group of people").

The nine questions were the following, listed below. For this thesis, we added some explanatory comments. The term *gros mots* is left untranslated to remind the reader of the difference of categorization between swear words in English and the more restrictive *gros mots* in French.

1. *Listez des gros mots, les premiers qui vous viennent à l'esprit.* ("List *gros mots*, the first ones that come to mind.")

   Justification: this was meant to elicit the most prototypical of all *gros mots*, without even mentioning any function or semantic domains (as the next questions do).

2. *Listez des gros mots qu'on dit lorsqu'on est surpris ou énervé.* ("List *gros mots* that people say when they are surprised or angry.")
   Justification: this aimed at collecting interjections without giving a technical or cumbersome explanation of what an interjection is. Interjecting is a frequent function of swear words, so this was an efficient way to elicit many of them at once.
3. *Listez des gros mots qui ont un rapport avec la sexualité.* ("List *gros mots* that are related to sexuality.")
   Justification: this was meant to elicit swear words from the domain of sexuality.[5]
4. *Listez des gros mots qui ont un rapport avec les déchets du corps humain.* ("List *gros mots* that are related to human body waste.")
   Justification: this was meant to elicit swear words from the domain of body waste.
5. *Listez des gros mots qui ont un rapport avec la religion.* ("List *gros mots* that are related to religion.")
   Justification: this was meant to elicit swear words from the domain of religion.
6. *Listez des gros mots qui sont des insultes.* ("List *gros mots* that are insults.")
   Justification: insulting is a frequent function of swear words, so this was an efficient way to elicit many of them at once.
7. *Listez des insultes qui NE sont PAS des gros mots.* ("List insults that are NOT *gros mots*.")

5 Bergen (2016a: 12–39) argues that cross-linguistically swear words are usually drawn from the semantic domains of religion, sexuality, and body waste, hence questions 3, 4, and 5.

Justification: this aimed at shifting the attention of respondents to insults in general, before focusing on a particular subcategory of insults in the next question.

8. *Listez des insultes (gros mots ou pas gros mots) qu'il ne faut pas prononcer car elles visent un groupe de personnes.* ("List insults (*gros mots* or not), that must not be said because they target a group of people.")
   Justification: this aimed at collecting slurs.
9. *Si vous en connaissez d'autres, listez des gros mots que vous n'avez donnés dans aucune réponse précédente.* ("If you know any others, list *gros mots* that you have not given in any previous answer.")
   Justification: this aimed at collecting any remaining words that would be less salient but still considered as *gros mots* by speakers.

Respondents were asked to list a maximum of 10 items for each question. The series of questions was meant to be broad enough to elicit all swear words that would spontaneously come to mind to native speakers of French. Respondents were able to go back and forth between questions and change their answers. Fifty-six native speakers solicited via Facebook contacts, who in turn solicited their own Facebook contacts, participated in the questionnaire from April 2018 to May 2018. We explicitly asked for native speakers of French – we would not have collected the answers of an anonymous non-native speaker, if any had participated. Before the questionnaire proper, on the first page, respondents gave basic information about their profile by answering the following five questions:

- *Quel est ton âge ?* ("How old are you?")
- *Quel est ton genre ?* ("What is your gender?")

- *Quel est ton métier ou occupation ?* ("What is your job or occupation?")
- *Quel est le diplôme le plus avancé que tu as obtenu ?* ("What is the highest diploma you got?")
- *Est-ce que le français est ta langue maternelle, et quel est ton pays de naissance ?* ("Is French your native language, and in what country were you born?")

The last question about native language aimed to confirm that the respondent was a native speaker of French, and collect what national variety of French they were a native speaker of, e.g., French of France vs. Belgian French vs. Quebecois French[6]. Other questions meant to give us a basic profile of our respondents. All respondents were born in France except one (1.8%) born in Poland. They were aged from 17 to 62, but 47 (83.9%) were in their twenties. Forty (71.4%) were women and the rest were men. Thirty-two (57.1%) were students. Their diplomas ranged from French middle school level (*Brevet des collèges*) to Master's degree.

Respondent	Country of birth	Age	Gender	Job/occupation	Diploma
1	France	23	M	Student	2- or 3-year degree
2	France	22	F	Student	2- or 3-year degree
3	France	22	F	Student	2- or 3-year degree
4	France	34	F	Accountant	2- or 3-year degree

6 We only considered national varieties because, based on our native speaker intuition, French swear words do not vary across regions of France, or only minimally so. Among the dozens of French swear words we collected, only one (*brin* "shit") seems typical of a specific region, namely of northern France.

Respondent	Country of birth	Age	Gender	Job/occupation	Diploma
5	France	23	F	Student	2- or 3-year degree
6	France	25	M	Student	Brevet des collèges
7	France	62	F	Housewife	2- or 3-year degree
8	France	51	F	Teacher	2- or 3-year degree
9	France	21	F	Student	Brevet des collèges
10	France	17	M	High schooler	2- or 3-year degree
11	France	22	F	Student	2- or 3-year degree
12	France	24	M	Student	Baccalauréat
13	France	45	M	Policeman	2- or 3-year degree
14	France	22	F	Student	2- or 3-year degree
15	France	50	F	Accountant	2- or 3-year degree
16	France	27	F	Server manager	2- or 3-year degree
17	France	24	F	Student	2- or 3-year degree
18	France	21	M	Student	Master
19	France	25	F	Export manager	Master
20	France	22	F	Student	Master
21	France	28	M	Teacher	2- or 3-year degree
22	France	25	F	-	Baccalauréat
23	France	22	F	Student	2- or 3-year degree
24	France	22	F	Student	2- or 3-year degree
25	France	21	F	Student	2- or 3-year degree
26	France	23	F	Student	2- or 3-year degree
27	France	22	M	Student	Master
28	France	26	F	Unemployed	2- or 3-year degree
29	France	22	F	Student	Master

Respondent	Country of birth	Age	Gender	Job/occupation	Diploma
30	France	26	M	Lawyer	2- or 3-year degree
31	France	23	F	Architect	Brevet des collèges
32	France	45	M	Office employee	2- or 3-year degree
33	France	24	F	Shop manager	2- or 3-year degree
34	France	22	F	Student	2- or 3-year degree
35	France	23	F	Student	2- or 3-year degree
36	France	22	F	Student	2- or 3-year degree
37	France	22	F	Student	Master
38	France	25	M	Game designer	2- or 3-year degree
39	France	21	F	Student	2- or 3-year degree
40	France	22	F	Student	2- or 3-year degree
41	France	20	F	Student	Brevet des collèges
42	France	23	M	Artist	2- or 3-year degree
43	France	22	F	Travel agent	2- or 3-year degree
44	France	23	M	Landscape gardener	2- or 3-year degree
45	France	22	F	Student	Baccalauréat
46	France	20	F	Student	2- or 3-year degree
47	France	22	F	Teaching assistant	2- or 3-year degree
48	France	21	F	Student	2- or 3-year degree
49	France	24	F	Employee	2- or 3-year degree
50	France	23	M	Student	Baccalauréat
51	France	20	F	-	Brevet des collèges
52	France	17	M	Student	2- or 3-year degree
53	France	20	F	Student	Baccalauréat
54	Poland	-	F	Restaurant owner	Master

Respondent	Country of birth	Age	Gender	Job/occupation	Diploma
55	France	27	M	Company manager	2- or 3-year degree
56	France	21	F	Student	2- or 3-year degree

Only swear words given by at least two respondents were included in the list. If a word is given spontaneously by two speakers as a swear word, it is probable that at least some proportion of the population does consider it to be a swear word. On the contrary, if a word is given by only one out of 56 respondents, then it is less likely that other speakers consider it to be a swear word. Most of these one-off answers (148 out of 249 items) seem indeed unreliable and have most likely resulted from an over- or misinterpretation of the instructions. For example, when asked to give swear words related to religion, one respondent answered *sang-de-bourbe*, which is the French translation for the fictional insult *mudblood* from the Harry Potter book and film series. We apply that minimum threshold of two participants separately for *gros mots* and for slurs, i.e., we do not include any word that was given by only one participant as a *gros mot* (Questions 1 to 6, and 9), and by only one other participant as a slur (Question 8).

Moreover, statistical comparisons were only possible with one-word items of the lexicon because the databases on English and French words that we used (see below) do not include multi-word expressions. As a result, 23 multi-word items were not included in the list, like *fils de pute*, equivalent to English *son of a bitch*. Such multi-word fixed expressions are obviously considered taboo because they include an individual swear word, which respondents listed anyway. For example, *fils de pute* ("son of a bitch") contains

pute ("whore/bitch") which was also found in the answers. Thus, removing multi-word expressions does not remove any important items. The resulting list of French swear words contains 78 items, given below in Table 2.

French swear word	IPA	Gloss	Freq. as *gros mot* / as slur	Frequency (Lexique)
merde	/mɛʁd/	shit	54 / 0	206.68
putain	/pytɛ̃/	fuck! / fucking [intensifier]	54 / 1	287.83
connard	/konaʁ/	big (male) jerk	46 / 0	40.7
salope	/salɔp/	dirty woman [slut]	45 / 2	62.54
enculé	/ɑ̃kyle/	sodomized man [scumbag]	43 / 8	22.73
pute	/pyt/	whore	41 / 6	87.91
connasse	/konas/	big (female) jerk	25 / 0	5.1
con	/kɔ̃/	(male) jerk	25 / 0	93.43
bâtard	/bataʁ/	bastard [scumbag]	21 / 0	9.89
pédé	/pede/	fag	20 / 34	25.64
bordel	/bɔʁdɛl/	mess / [interjection] / brothel	19 / 0	97.84
enfoiré	/ɑ̃fwaʁe/	[scumbag]	18 / 0	30.94
salaud	/salo/	dirty man [scumbag]	15 / 0	66.74
conne	/kɔ̃n/	(female) jerk	9 / 0	8.57
pisse	/pis/	piss	9 / 0	5.98
bite	/bit/	cock / dick	7 / 0	22.93
chiasse	/ʃjas/	the runs / the trots [diarrhea]	5 / 0	0.81
cul-bénit	/kybeni/	[overly devout person]	5 / 0	0.02
foutre	/futʁ/	to make / to do / to put / cum	5 / 0	97.99
pétasse	/petas/	tart	5 / 0	6.91
tapette	/tapɛt/	fag	5 / 5	4.77

French swear word	IPA	Gloss	Freq. as *gros mot* / as slur	Frequency (Lexique)
brin	/bʁɛ̃/	shit	4 / 0	4.95
chienne	/ʃjɛn/	bitch	4 / 0	12.84
chier	/ʃje/	to shit	4 / 0	53.3
enculer	/ɑ̃kyle/	to sodomize	4 / 0	5.88
pédale	/pedal/	fag	4 / 3	3.38
suceuse	/sysøz/	(female) (cock)sucker	4 / 0	0.55
baiser	/beze/	to fuck	3 / 0	42.25
chatte	/ʃat/	pussy	3 / 0	16.54
crotte	/kʁɔt/	shit / turd	3 / 0	3.46
cul	/ky/	ass	3 / 0	145.85
feuj	/føʒ/	Jew	3 / 1	0.04
niquer	/nike/	to fuck / to screw	3 / 0	3.27
pouffiasse	/pufjas/	tart	3 / 0	2.73
sacrebleu	/sakʁəblø/	[interjection]	3 / 0	0.94
saloperie	/salɔpəʁi/	filth	3 / 0	19.38
tarlouze	/taʁluz/	fag	3 / 3	NA
abruti	/abʁyti/	idiot	2 / 0	19.13
branleur	/bʁɑ̃lœʁ/	wanker	2 / 0	2.9
couilles	/kuj/	balls [testicles]	2 / 0	35.24
couillon	/kujɔ̃/	idiot	2 / 0	4.04
ducon	/dykɔ̃/	idiot	2 / 0	6.86
enculée	/ɑ̃kyle/	sodomized woman [scumbag]	2 / 0	0.04
fumier	/fymje/	manure [scumbag]	2 / 0	15.66
gouine	/gwin/	dyke	2 / 5	2.41
gourgandine	/guʁgɑ̃din/	trollop	2 / 0	0.06

French swear word	IPA	Gloss	Freq. as *gros mot* / as slur	Frequency (Lexique)
grognasse	/gʁoɲas/	[grumpy woman]	2 / 0	0.14
gueule	/gœl/	mouth / face	2 / 0	118.45
merdeux	/mɛʁdø/	shit-covered boy [brat]	2 / 0	5.22
nique	/nik/	fuck ... / screw ...	2 / 0	1.56
pédophile	/pedofil/	pedophile	2 / 0	0.95
pourriture	/puʁityʁ/	rot	2 / 0	3.93
puceau	/pyso/	(male) virgin	2 / 0	2.15
queutard	/køtaʁ/	sex-obsessed man	2 / 0	0.19
sainte-nitouche	/sɛ̃tnituʃ/	[sanctimonious woman]	2 / 0	0.35
saligaud	/saligo/	dirty man [scumbag]	2 / 0	2.46
sheitan	/ʃetan/	demon	2 / 0	NA
suce-boules	/sysbul/	ball-sucker	2 / 0	NA
tchoin	/tʃwɛ̃/	whore	2 / 0	NA
bougnoule	/buɲul/	[anti-Arab slur]	1 / 14	0.12
trisomique	/tʁizɔmik/	person with Down's syndrome	1 / 4	0.32
youpin	/jupɛ̃/	kike [antisemitic slur]	1 / 2	1.05
nègre	/nɛgʁ/	nigger	0 / 9	11.26
négro	/negʁo/	nigger	0 / 5	4.64
chinetoque	/ʃinətɔk/	chink	0 / 4	1.06
jaune	/ʒon/	yellowman	0 / 4	6.43
triso	/tʁizɔ/	person with Down's syndrome	0 / 4	NA
bridé	/bʁide/	slant-eye	0 / 3	0.63
handicapé	/ɑ̃dikape/	handicapped	0 / 3	1.08
Juif	/ʒɥif/	Jew	0 / 3	14.59
Arabe	/aʁab/	Arab	0 / 2	6.71

French swear word	IPA	Gloss	Freq. as *gros mot* / as slur	Frequency (Lexique)
attardé	/ataʁde/	retarded	0 / 2	0.74
bicot	/biko/	[anti-Arab slur]	0 / 2	0.11
boukak	/bukak/	[anti-Arab slur]	0 / 2	NA
débile	/debil/	moron / idiot	0 / 2	8.32
éclopé	/eklope/	cripple	0 / 2	0.17
Mongol	/mɔ̃gɔl/	Mongol [Down's syndrome]	0 / 2	0.21
Noir	/nwaʁ/	black	0 / 2	54.22

Table 2: List of 78 French swear words by order of frequency in the answers of the questionnaire.

We give below the answers given for every question below in Table 3, Table 4, Table 5, Table 6, Table 7, Table 8, Table 9, Table 10, and Table 11. These data on each question are useful in order to evaluate two claims we made on French swear words. In Section 2.2.2, we argued that some swear words of French lost their meaning related to taboo subject. In Section 2.2.5, we argued that the French gros mots category does not include non-sexual slurs. We discuss this part of our data and the implications for the definition and meaning of swear words (RQ2b) in Section 4.2. Here in Table 3, Table 4, Table 5, Table 6, Table 7, Table 8, and Table 11 we include only answers who made it into the final list in Table 2, i.e., the tables do not include multi-word answers or any *gros mot* given by only one respondent throughout the whole questionnaire. However for Question 8 on slurs in Table 10, we include answers given by only one respondent and multi-word answers, because they are relevant to our discussion on slurs. We also include here the one-word answers given by at least two respondents to Question 7 ("List insults that are NOT *gros mots*.") because they are relevant to the same discussion.

1. "List *gros mots*, the first ones that come to mind." (Answer rate: 100%)	**Number of respondents who gave that word**	**In %**
putain ("fuck!" / "fucking" [intensifier])	47	83.9%
connard ("big (male) jerk")	40	71.4%
merde ("shit")	37	66.1%
enculé ("sodomized man" [scumbag])	24	42.9%
salope ("dirty woman" [slut])	22	39.3%
bâtard ("bastard" [scumbag])	13	23.2%
bordel ("mess" / [interjection] / "brothel")	13	23.2%
con ("(male) jerk")	13	23.2%
pute ("whore")	11	19.6%
connasse ("big (female) jerk")	10	17.9%
enfoiré ("scumbag")	7	12.5%
salaud ("dirty man" [scumbag])	7	12.5%
conne ("(female) jerk")	3	5.4%
pédé ("fag")	3	5.4%
bite ("cock" / "dick")	2	3.6%
chier ("to shit")	2	3.6%
gueule ("mouth" / "face")	2	3.6%
pétasse ("tart")	2	3.6%
saloperie ("filth")	2	3.6%
chiasse ("the runs" / "the trots" [diarrhea])	1	1.8%
chienne ("bitch")	1	1.8%
couilles ("balls" [testicles])	1	1.8%
cul ("ass")	1	1.8%

1. "List *gros mots*, the first ones that come to mind." (Answer rate: 100%)	**Number of respondents who gave that word**	**In %**
enculée ("sodomized woman" [scumbag])	1	1.8%
enculer ("to sodomise")	1	1.8%
foutre ("to make" / "to do" / "to put" / "cum")	1	1.8%
fumier ("manure" [scumbag])	1	1.8%
nique ("fuck ..." / "screw …")	1	1.8%
pédale ("fag")	1	1.8%
saligaud ("dirty man" [scumbag])	1	1.8%

Table 3: Answers to Question 1 of our French swear word questionnaire

2. "List *gros mots* that people say when they are surprised or angry." (Answer rate: 100%)	**Number of respondents who gave that word**	**In %**
putain ("fuck!" / "fucking" [intensifier])	52	92.9%
merde ("shit")	36	64.3%
bordel ("mess" / [interjection] / "brothel")	14	25.0%
connard ("big (male) jerk")	7	12.5%
enfoiré ("scumbag")	5	8.9%
enculé ("sodomized man" [scumbag])	3	5.4%
nique ("fuck ..." / "screw …")	1	1.8%
pédé ("fag")	1	1.8%
salope ("dirty woman" [slut])	1	1.8%
saloperie ("filth")	1	1.8%

Table 4: Answers to Question 2 of our French swear word questionnaire

3. "List *gros mots* that are related to sexuality." (Answer rate: 96.4%)	**Number of respondents who gave that word**	**In %**
salope ("dirty woman" [slut])	30	53.6%
pute ("whore")	28	50.0%
enculé ("sodomized man" [scumbag])	26	46.4%
pédé ("fag")	17	30.4%
putain ("fuck!" / "fucking" [intensifier])	10	17.9%
bite ("cock" / "dick")	6	10.7%
chienne ("bitch")	4	7.1%
tapette ("fag")	4	7.1%
suceuse ("(female) (cock)sucker")	4	7.1%
baiser ("to fuck")	3	5.4%
chatte ("pussy")	3	5.4%
con ("(male) jerk")	3	5.4%
enculer ("to sodomise")	3	5.4%
pédale ("fag")	3	5.4%
salaud ("dirty man" [scumbag])	3	5.4%
tarlouze ("fag")	3	5.4%
bordel ("mess" / [interjection] / "brothel")	2	3.6%
couilles ("balls" [testicles])	2	3.6%
cul ("ass")	2	3.6%
foutre ("to make" / "to do" / "to put" / "cum")	2	3.6%
gouine ("dyke")	2	3.6%
puceau ("(male) virgin")	2	3.6%

3. "List *gros mots* that are related to sexuality." (Answer rate: 96.4%)	**Number of respondents who gave that word**	**In %**
queutard ("sex-obsessed man")	2	3.6%
suce-boules ("ball-sucker")	2	3.6%
bâtard ("bastard" [scumbag])	1	1.8%
branleur ("wanker")	1	1.8%
enculée ("sodomized woman" [scumbag])	1	1.8%
gourgandine ("trollop")	1	1.8%
merde ("shit")	1	1.8%
nique ("fuck ..." / "screw …")	1	1.8%
niquer ("to fuck" / "to screw")	1	1.8%
pétasse ("tart")	1	1.8%
tchoin ("whore")	1	1.8%

Table 5: Answers to Question 3 of our French swear word questionnaire

4. "List *gros mots* that are related to human body waste." (Answer rate: 94.6%)	**Number of respondents who gave that word**	**In %**
merde ("shit")	44	78.6%
pisse ("piss")	9	16.1%
chiasse ("the runs" / "the trots" [diarrhea])	5	8.9%
brin ("shit")	4	7.1%
chier ("to shit")	3	5.4%
crotte ("shit" / "turd")	3	5.4%
foutre ("to make" / "to do" / "to put" / "cum")	2	3.6%
pourriture ("rot")	2	3.6%
bite ("cock" / "dick")	1	1.8%
merdeux ("shit-covered boy" [brat])	1	1.8%
saloperie ("filth")	1	1.8%

Table 6: Answers to Question 4 of our French swear word questionnaire

5. "List *gros mots* that are related to religion." (Answer rate: 44.6%)	**Number of respondents who gave that word**	**In %**
cul-bénit ([overly devout person])	5	8.9%
feuj ("Jew")	3	5.4%
sacrebleu ([interjection])	3	5.4%
pédophile ("pedophile")	2	3.6%
sainte-nitouche ([sanctimonious woman])	2	3.6%
sheitan ("demon")	2	3.6%

Table 7: Answers to Question 3 of our French swear word questionnaire

6. "List *gros mots* that are insults." (Answer rate: 98.2%)	**Number of respondents who gave that word**	**In %**
connard ("big (male) jerk")	37	66.1%
enculé ("sodomized man" [scumbag])	29	51.8%
salope ("dirty woman" [slut])	29	51.8%
pute ("whore")	23	41.1%
connasse ("big (female) jerk")	19	33.9%
bâtard ("bastard" [scumbag])	17	30.4%
con ("(male) jerk")	13	23.2%
enfoiré ("scumbag")	12	21.4%
salaud ("dirty man" [scumbag])	11	19.6%
pédé ("fag")	9	16.1%
conne ("(female) jerk")	7	12.5%
putain ("fuck!" / "fucking" [intensifier])	5	8.9%
merde ("shit")	3	5.4%
pétasse ("tart")	3	5.4%
abruti ("(male) idiot")	2	3.6%
pouffiasse ("tart")	2	3.6%
bordel ("mess" / [interjection] / "brothel")	1	1.8%
chienne ("bitch")	1	1.8%
ducon ("(male) idiot")	1	1.8%
fumier ("manure" [scumbag])	1	1.8%
gouine ("dyke")	1	1.8%
grognasse ([grumpy woman])	1	1.8%
tarlouze ("fag")	1	1.8%

6. "List *gros mots* that are insults." (Answer rate: 98.2%)	**Number of respondents who gave that word**	**In %**
tapette ("fag")	1	1.8%
tchoin ("whore")	1	1.8%

Table 8: Answers to Question 6 of our French swear word questionnaire

7. "List insults that are NOT *gros mots*." (Answer rate: 87.5%)	**Number of respondents who gave that word**	**In %**
idiot ("(male) idiot")	13	23.2%
débile ("moron" / "idiot")	12	21.4%
imbécile ("imbecile")	7	12.5%
stupide ("stupid")	7	12.5%
abruti ("(male) idiot")	4	7.1%
idiote ("(female) idiot")	4	7.1%
attardé ("retard")	3	5.4%
blaireau ("badger" [jerk])	3	5.4%
clochard ("tramp" / "bum" / "hobo")	3	5.4%
enfoiré ("scumbag")	3	5.4%
merde ("shit")	3	5.4%
bougnoule ([anti-Arab slur])	2	3.6%
chienne ("bitch")	2	3.6%
crétin ("idiot")	2	3.6%
déchet ("human waste" / "waste of space")	2	3.6%
feignant ("idler" / "lazy")	2	3.6%
gourgandine ("trollop")	2	3.6%

7. "List insults that are NOT *gros mots*." (Answer rate: 87.5%)	**Number of respondents who gave that word**	**In %**
lâche ("coward")	2	3.6%
menteur ("liar")	2	3.6%
mongol ("Mongol" [person with Down's syndrome])	2	3.6%
mouton ("sheep" / "sheeple")	2	3.6%
pédale ("fag")	2	3.6%
pédophile ("pedophile")	2	3.6%
radin ("cheapstake" / "skinflint")	2	3.6%
sorcière ("witch")	2	3.6%
vendu ("sellout")	2	3.6%

Table 9: Answers to Question 7 of our French swear word questionnaire

8. "List insults (*gros mots* or not), that must not be said because they target a group of people." [i.e., slurs] (Answer rate: 80.4%)	**Number of respondents who gave that word**	**In %**
pédé ("fag")	34	60.7%
bougnoule ([anti-Arab slur])	14	25.0%
nègre ("nigger")	9	16.1%
enculé ("sodomized man" [scumbag])	8	14.3%
pute ("whore")	6	10.7%
gouine ("dyke")	5	8.9%
négro ("nigger")	5	8.9%
tapette ("fag")	5	8.9%
chinetoque ("chink")	4	7.1%
jaune ("yellowman")	4	7.1%

8. "List insults (*gros mots* or not), that must not be said because they target a group of people." [i.e., slurs] (Answer rate: 80.4%)	**Number of respondents who gave that word**	**In %**
triso ("person with Down's syndrome")	4	7.1%
trisomique ("person with Down's syndrome")	4	7.1%
bridé ("slant-eye")	3	5.4%
handicapé ("handicapped")	3	5.4%
Juif ("Jew")	3	5.4%
pédale ("fag")	3	5.4%
tarlouze ("fag")	3	5.4%
Arabe ("Arab")	2	3.6%
attardé ("retard")	2	3.6%
bicot ([anti-Arab slur])	2	3.6%
boukak ([anti-Arab slur])	2	3.6%
débile ("moron" / "idiot")	2	3.6%
éclopé ("cripple")	2	3.6%
gros tas ("big heap")	2	3.6%
Mongol ("Mongol" [person with Down's syndrome])	2	3.6%
Noir ("black")	2	3.6%
salope ("dirty woman" [slut])	2	3.6%
youpin ("kike" [antisemitic slur])	2	3.6%
autiste ("autistic person")	1	1.8%
baleine ("whale")	1	1.8%
bamboula ([anti-black slur])	1	1.8%
bouffeur de chiens ("dog eater")	1	1.8%
catho ("Catholic")	1	1.8%

8. "List insults (*gros mots* or not), that must not be said because they target a group of people." [i.e., slurs] (Answer rate: 80.4%)	**Number of respondents who gave that word**	**In %**
cotorep ([slur for a disabled person])	1	1.8%
crouille ([anti-Arab slur])	1	1.8%
face de citron ("lemon face")	1	1.8%
femmelette ("sissy")	1	1.8%
feuj ("Jew")	1	1.8%
fiotte ("sissy")	1	1.8%
gonzesse ("chick" [woman])	1	1.8%
gros feuj ("big Jew")	1	1.8%
gros sac ("big bag")	1	1.8%
grosse tarte ("big pie")	1	1.8%
has been ("has been")	1	1.8%
nain ("dwarf")	1	1.8%
nazi ("nazi")	1	1.8%
niakoué ([anti-Asian slur])	1	1.8%
nique sa race ("fuck his/her/their/its race")	1	1.8%
polack ("Polack")	1	1.8%
portos ([slur for a Portuguese person])	1	1.8%
petit gris ("little grey" [anti-Arab slur])	1	1.8%
putain ("fuck!" / "fucking... [intensifier]")	1	1.8%
rebeu ("Arab")	1	1.8%
sale Arabe ("dirty Arab")	1	1.8%
sale blanc ("dirty white")	1	1.8%
sale Juif ("dirty Jew")	1	1.8%

8. "List insults (*gros mots* or not), that must not be said because they target a group of people." [i.e., slurs] (Answer rate: 80.4%)	**Number of respondents who gave that word**	**In %**
sale nègre ("dirty nigger")	1	1.8%
sale pute ("dirty whore")	1	1.8%
sale trans ("dirty transgender")	1	1.8%
schizo ("schizophrenic")	1	1.8%
taré ("defective")	1	1.8%
terroriste ("terrorist")	1	1.8%
travelo ("tranny")	1	1.8%

Table 10: Answers to Question 8 of our French swear word questionnaire

9. "If you know any others, list *gros mots* that you have not given in any previous answer." (Answer rate: 32.1%)	**Number of respondents who gave that word**	**In %**
couillon ("idiot")	2	3.6%
niquer ("to fuck" / "to screw")	2	3.6%
pétasse ("tart")	2	3.6%
salope ("dirty woman" [slut])	2	3.6%
branleur ("wanker")	1	1.8%
connasse ("big (female) jerk")	1	1.8%
con ("(male) jerk")	1	1.8%
conne ("(female) jerk")	1	1.8%
ducon ("idiot")	1	1.8%
grognasse ([grumpy woman])	1	1.8%
gourgandine ("trollop")	1	1.8%
merde ("shit-covered boy" [brat])	1	1.8%

9. "If you know any others, list *gros mots* that you have not given in any previous answer." (Answer rate: 32.1%)	**Number of respondents who gave that word**	**In %**
pédale ("fag")	1	1.8%
pouffiasse ("tart")	1	1.8%
pouilleux ("flea-ridden")	1	1.8%
pute ("whore")	1	1.8%
saligaud ("dirty man" [scumbag])	1	1.8%
tapette ("fag")	1	1.8%

Table 11: Answers to Question 9 of our French swear word questionnaire

Bergen (2016a: 52) compared his list of swear words to the 10% most frequent monosyllabics in the MRC Psycholinguistic Database (University of Western Australia, School of Psychological Science 1997). The same database was used here to replicate his study and compare swear words to the top 10% most frequent words (in spoken English) of the English lexicon.[7] The list of French swear words was compared to the equivalent group in the French lexicon in the Lexique database (New & Pallier 2019). Swear words which were among those top 10% were obviously removed from the list of regular words used for the comparison, to avoid comparing swear words with swear words.

Two methodological clarifications are in order concerning the MRC Psycholinguistic Database website. First, a user's request yields a mere list, so the number of words was counted semi-automatically using a spreadsheet file.

7 The restriction to the 10% most frequent words is justified as they are more representative of the lexicon (actively) used by speakers; moreover, it allows us to compare our results to Bergen's (2016a), as he used the same restriction. Function words, which presumably would not show any sound symbolic quality were not excluded from this list, but as they represent only a very small portion (less than 1.5%) their impact can be ignored.

Second, the website displays only the first 5,000 results for a given request. Whenever our request returned more than 5,000 results, we ran several more specific requests and simply summed the different numbers. For example, in order to find the number of words with plosives (more than 5,000), we first asked for nouns with plosives, then for verbs with plosives, then for words with plosives that are neither nouns nor verbs, and added up the three numbers.

Each list of swear words was then compared to the regular lexicon via Fisher's exact tests performed in the R software environment.[8] Table 12 gives an example of an input table for English swear words.

	English swear words	**English regular words**
monosyllabic	35	1,625
not monosyllabic	36	8,499

Table 12: Input table for Fisher's exact test

The Fisher test returns a p-value below 0.001, which means that the observed difference in proportions is due to chance. We use the conventional threshold of 5 percent ($p < 0.05$) to consider that a tendency is significant. We opted for Fisher's exact test, a standard significance test for such comparisons, because it allows us to compare our findings to Bergen's (2016a) who also uses that test.[9] All the subsequent statistical tests in this thesis are also Fisher exact tests.

A first series of tests was conducted on the list of English swear words to replicate Bergen's (2016a) study and thus verify if his observations still hold. Bergen makes three related claims, (1) that compared to regular words, English swear words are more often monosyllabics, (2) that compared to

8 https://www.r-project.org/. Last accessed on 31 May, 2019.

9 One could consider other methods, but this would prevent us from comparing our findings with Bergen's (2016a).

monosyllabic regular words, English monosyllabic swear words end more often with a consonant, and (3) that compared to regular words, English swear words are more often closed monosyllabics.[10] We tested every claim for English, then for French.

As we will cover in the next section, the results did not prove conclusive for French. Consequently, we looked for an alternative sound tendency in swear words that would apply to both languages. Bergen notices that English closed monosyllabic swear words tend to end not just with any consonants, but with plosives (2016a: 64). Plosives, Haiman (2018: 209–212) suggests, are iconically appropriate to express "familiarity" and/or "violation of hearer's space", a meaning that is relevant for swear words. Yet plosives may not be the only consonants that are overrepresented among swear words. Yardy (2010: 12–20, 71–78) argues that cross-linguistically, swear words contain more of the least sonorous consonants, i.e., least vowel-like consonants, and conversely fewer of the most sonorous consonants. He supports his claim with empirical data on English swear words (2010: 52–56). Although his results are based on an unreliable dataset (see Vallery 2019), his hypothesis is worth investigating on the basis of our data. The sonority scale used by Yardy consists in the following order with sonority decreasing: glides (e.g., /j/, /w/), rhotic approximants, flaps, laterals (e.g., /l/), trills, nasals (e.g., /n/, /m/, /ŋ/), voiced fricatives (e.g., /v/, /z/), voiced affricates (e.g., /dʒ/), voiced plosives (e.g., /b/, /d/, /g/), voiceless fricatives (e.g., /f/, /s/, /ʃ/, /θ/, /h/), voiceless affricates (e.g., /tʃ/), and voiceless plosives (e.g., /p/, /t/, /k/) (Yardy 2010: 63; Parker 2008) We represent that sonority scale in Figure 3.

10 Bergen did not conduct statistical tests about claims (1) and (3) although such tests are relevant to the main argument which is why we decided to conduct them.

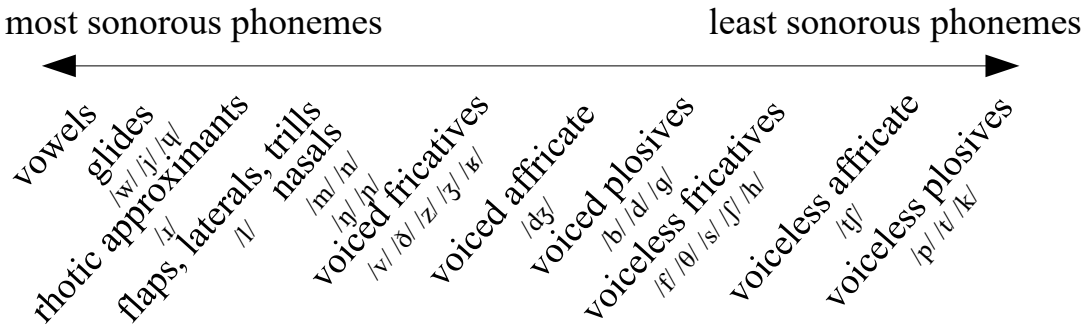

Figure 3: Sonority scale according to Parker (2008) used by Yardy (2010: 63)

The consonants that were more frequent in swear words according to Yardy (2010) were voiced plosives, voiceless fricatives, the voiceless affricate, and voiceless plosives (2010: 63–64). One could imagine that actually, the tendency might apply only to a smaller group of consonants, for example only voiceless obstruents. Or on the contrary, one could imagine that the tendency applies to a larger group of consonants, for example all obstruents, i.e., plosives, fricatives, and affricates, which are all relatively more unsonorous compared to sonorants, i.e., all the other more sonorous consonants from nasals all the way to glides. In order to account for that, we tested for which individual unsonorous consonants are more frequent in swear words – again, using the Fisher exact test on our data on regular vs. swear words in English and French. Our results suggest that the consonants more frequent in swear words are the five least sonorous categories, i.e., the voiced affricate, the voiced plosives, the voiceless fricatives, the voiceless affricate, and the voiceless plosive. Consequently, we grouped those five least sonorous categories together in our subsequent test: we tested for a tendency for swear words to contain at least one of these unsonorous consonants, i.e., at least one /p/, /t/, /k/, /tʃ/, /f/, /θ/, /s/, /ʃ/, /h/, /b/,

/d/, /g/, or /dʒ/. We used the same grouping for subsequent tests and studies. In this thesis, we will call this supercategory *the least sonorous consonants* or *unsonorous consonants* for short.[11]

A potential bias with that method of counting is arguably the length of swear words. The longer words are, the more likely they are to contain at least one plosive or at least one unsonorous consonant. A different measure is to compare the number of unsonorous consonants to the number of phonemes in a given word (token frequencies). We will call this measure the *unsonority density* of words. For example, the swear word *dickhead*, pronounced /ˈdɪkhed/, contains six phonemes, four of which are unsonorous consonants (/d/, /k/, /h/, and /d/). Its unsonority density is thus four out of six or 0.67. The higher the ratio, the higher the unsonorous character of the word. We computed the ratio semi-automatically for all regular words in the English and the French databases, by subtracting the number of words with a given (minimum) number of unsonorous consonants, from the larger number of words with (minimum) one less unsonorous consonant. For example, we queried the MRC Psycholinguistic Database website for how many English words contain exactly three phonemes and at least one unsonorous consonant (a minimal density of 0.33), which gave us 606 words. Subsequently, we queried how many of these 606 words contain at least two unsonorous consonants, which yielded 220 words. From this, we can deduce that there are 606-220=386 words in the database which contain exactly three phonemes and exactly one unsonorous consonant (exact density of 0.33). We repeated those requests and calculations

11 We are aware that the five least sonorous categories of consonants are not equally (un)sonorous: the voiced affricate is relatively more sonorous than voiced plosives, voiced plosives are relatively more sonorous than voiceless fricatives, etc.

until we arrived at a complete list of all observed densities and how many words there are for each.

The list of observed densities for regular English and French words, from the highest (1, i.e., all phonemes are unsonorous consonants) to the lowest (0, i.e., no phoneme is an unsonorous consonant) is given below in Table 13. The number of English regular words is much lower than in our previous tests (4,849 instead of 14,449). This is because only some of them are annotated for their number of phonemes on the MRC Psycholinguistic Database.

Unsonority density	**English words with that density**	**In %**	**French words with that density**	**In %**
1	3	0.1%	15	0.1%
0.8	2	0.0%	3	0.0%
0.75	68	1.4%	45	0.3%
0.71	6	0.1%	4	0.0%
0.7	1	0.0%	0	0.0%
0.67	250	5.2%	440	2.9%
0.63	13	0.3%	12	0.1%
0.6	123	2.5%	172	1.2%
0.58	1	0.0%	0	0.0%
0.57	61	1.3%	88	0.6%
0.56	14	0.3%	12	0.1%
0.55	4	0.1%	1	0.0%
0.5	768	15.8%	1938	13.0%
0.46	3	0.1%	8	0.1%
0.46	22	0.5%	11	0.1%

Unsonority density	English words with that density	In %	French words with that density	In %
0.44	67	1.4%	91	0.6%
0.43	165	3.4%	383	2.6%
0.42	25	0.5%	7	0.0%
0.4	398	8.2%	1293	8.7%
0.39	2	0.0%	2	0.0%
0.38	144	3.0%	242	1.6%
0.36	35	0.7%	13	0.1%
0.36	2	0.0%	0	0.0%
0.33	755	15.6%	2158	14.5%
0.31	7	0.1%	7	0.0%
0.3	74	1.5%	88	0.6%
0.29	158	3.3%	748	5.0%
0.27	28	0.6%	40	0.3%
0.27	1	0.0%	0	0.0%
0.25	445	9.2%	1797	12.0%
0.23	4	0.1%	1	0.0%
0.22	58	1.2%	137	0.9%
0.21	2	0.0%	1	0.0%
0.2	280	5.8%	1470	9.8%
0.18	11	0.2%	13	0.1%
0.17	136	2.8%	810	5.4%
0.14	91	1.9%	408	2.7%
0.13	42	0.9%	192	1.3%
0.11	21	0.4%	61	0.4%

Unsonority density	English words with that density	In %	French words with that density	In %
0.1	6	0.1%	22	0.1%
0.09	1	0.0%	4	0.0%
0.08	3	0.1%	2	0.0%
0.08	0	0.0%	1	0.0%
0	549	11.3%	2185	14.6%
Total	4849	100%	14925	100%

Table 13: Observed densities of unsonorous consonants for regular English and French words

The comparison of the density measures was done by using a Fisher exact test, setting off how many swear words vs. how many regular words have a minimal density of 0.33 (at least one out of three phonemes is an unsonorous consonant). The minimum threshold of 0.33 is a somewhat arbitrary cut-off point, but it is methodologically justified for French fictional swear words used in our second case study (see Section 3.2.1). In order to make the results fully comparable, we use the same threshold throughout our studies. Again, swear words which happened to be among the top 10% were removed from the list of regular words used for the comparison.

3.1.2 Results for Real-life swear words

The results for each of Bergen's claims are given respectively in Table 14, Table 15 and Table 16, where the percentage of specific words within the set of regular words (RW) is compared to that of swear words (SW). The second test

(Table 15) compares monosyllabic regular words to monosyllabic swear words, consequently the subset for regular words is smaller, comprising the top 10% most frequent words among English monosyllabics (instead of the top 10% of all words).[12] These results are presented visually in Figure 5.

	English SW	English RW	English RW	English SW
monosyllabic	35	1,625	**16.1%**	**49.3%**
not monosyllabic	36	8,499	83.9%	50.7%
total	71	10,214	100%	100%

Table 14: Our data on Bergen's (2016a) *first claim on English swear words (**$p = 9.652^{-11}$**)*

	English RW (most frequent monosyllabics only)	**English SW (monosyllabics only)**	**English RW (idem)**	**English SW (idem)**
closed	787	31	**63.4%**	**88.6%**
not closed	455	4	36.6%	11.4%
total	1,242	35	100%	100%

*Table 15: Our data on Bergen's (2016a) second claim on English swear words (**$p = 0.001912$**)*

12 We limit this comparison to the subset of monosyllabic regular words only since Bergen uses this method too, so we can better approximate his method. He does not explain this methodological choice but we assume it makes the comparison stronger.

	English RW	English SW	English RW	English SW
closed monosyllabics	1,280	32	**12.6%**	**45.1%**
not closed monosyllabics	8,844	39	87.4%	54.9%
total	10,124	71	100%	100%

*Table 16: Our data on Bergen's (2016a) third claim on English swear words (**$p = 2.275^{-11}$**)*

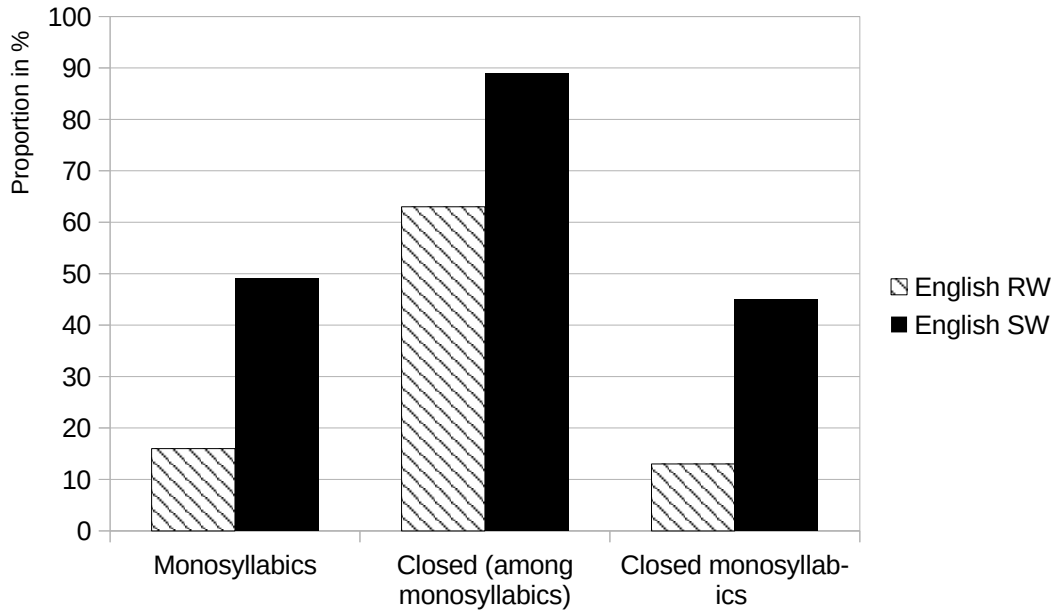

Figure 4: Tendencies corresponding to Bergen's claims on English swear words (p<0.01 for all three).

Bergen's (2016a) findings for English have all been confirmed. However, for French swear words, the same tests turn out to be non-significant, as shown in Table 17, Table 18, and Table 19. In line with what we did for English, two different sets are used: the top 10% most frequent monosyllabic regular words in the second test and the top 10% most frequent of the entire set of regular words in the first and third tests. The results are presented visually in Figure 7.

	French RW	French SW	French RW	French SW
monosyllabic	3,252	24	**21.9%**	**30.8%**
not monosyllabic	11,565	54	78.1%	69.2%
total	14,817	78	100%	100%

*Table 17: French swear words are not significantly more often monosyllabics (**p = 0.07359**)*

	French RW (most frequent monosyllabics only)	French SW (monosyllabics only)	French RW (idem)	French SW (idem)
closed	560	19	**59.0%**	**79.2%**
not closed	389	5	41.0%	20.8%
total	949	24	100%	100%

*Table 18: French monosyllabics swear words do not significantly end with a consonant (**p = 0.05731**)*

	French RW	French SW	French RW	French SW
closed monosyllabics	3,165	15	**21.4%**	**19.2%**
not closed monosyllabics	11,652	63	78.6%	80.8%
total	14,817	78	100%	100%

*Table 19: French swear words are not more often closed monosyllabics (**p = 0.7817**)*

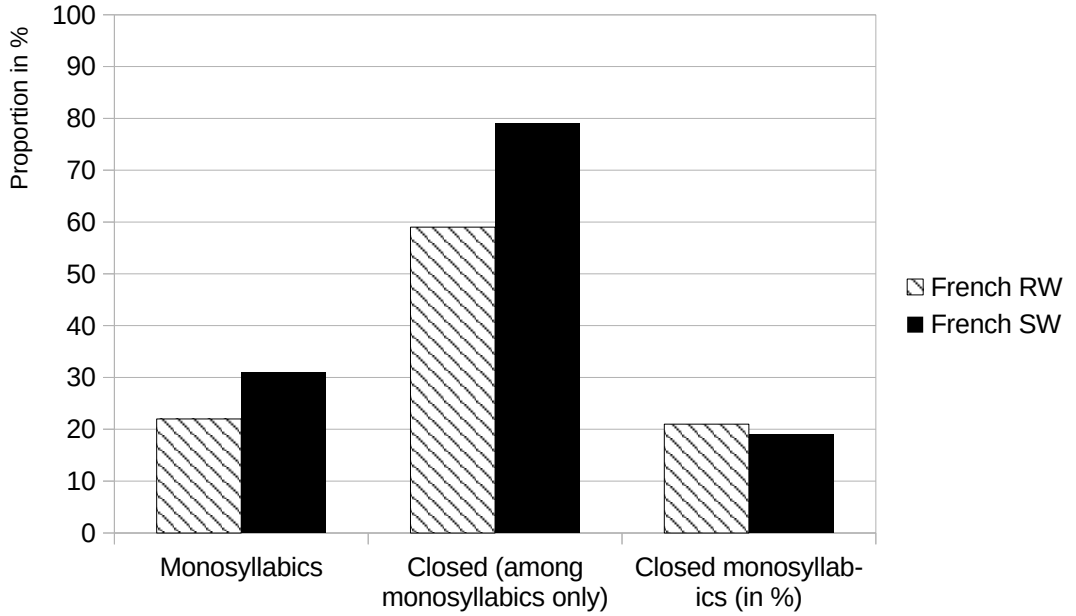

Figure 5: Tendencies corresponding to Bergen's claims for French swear words (p>0.05 for all three).

The first two tendencies – swear words being monosyllabic, monosyllabic swear words ending with a consonant – are smaller for French and do not reach significance and the third one – swear words being closed monosyllabics – is not observable at all and even opposite: slightly fewer French swear words are closed monosyllabics. That last observation suggests that the properties of English swear words are not cross-linguistic, contrary to what Bergen (2016a: 60–63) hypothesised.

We checked Yardy's (2010) proposition that swear words contain more unsonorous categories of consonants, namely plosives, voiceless fricatives, and the voiceless affricate, because they are iconically more appropriate for expressing aggression – see Section 2.1.2. We looked at individual consonants rather than at a whole group to test whether the categories that are more frequent in swear words are indeed these ones and only these ones – see Section 3.1.1. We tested regardless of where this

individual consonant occurs in the word. The consonants that are significantly more present in swearwords are some of the least sonorous, and conversely, the ones that are significantly less present are some of the most sonorous. We represent those results in below. There are only a few exceptions to that general tendency: /ŋ/ ($p<0.01$) and for English and /n/ ($p<0.05$) for French – two out of fifteen significant results.

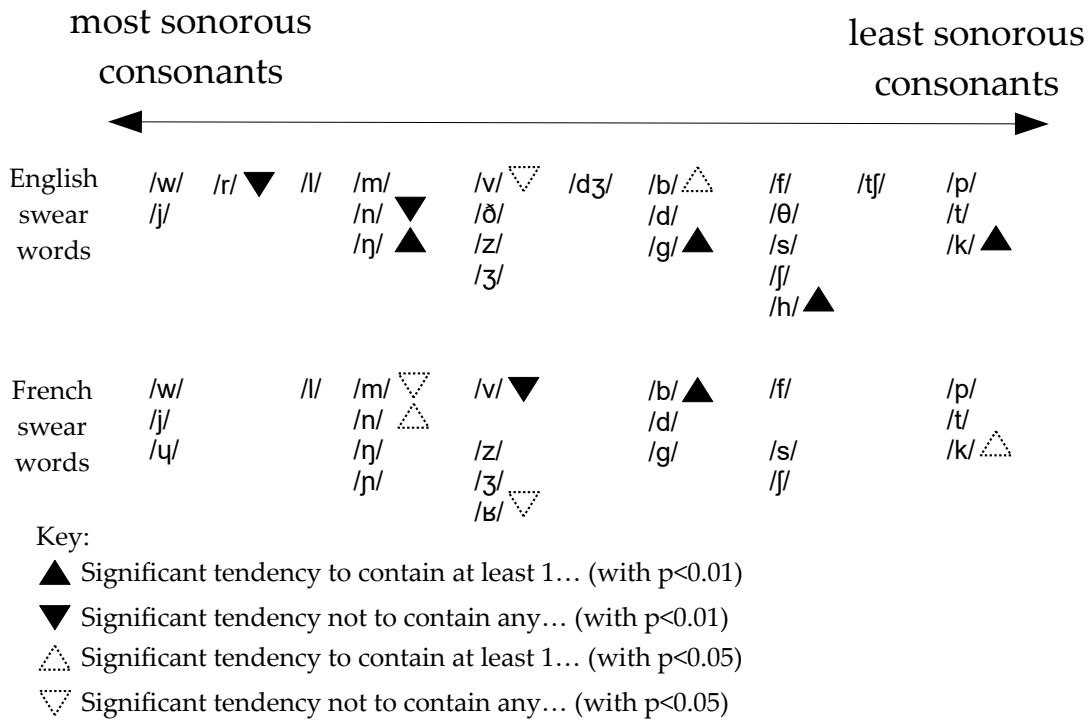

Figure 6: Tendencies of swear words for specific consonants

There is a caveat here. In the case of multiple comparisons, the probability of getting false positives increases. In the literature, p-value adjustments are done using different methods, but we decided to set the alpha value to 0.01, i.e., to first consider only the results where $p<0.01$ instead of $p<0.05$. Also, we test the same hypothesis, whereas such p-value adjustments are required to avoid false rejections of null hypotheses.

In terms of results where $p<0.01$, significantly fewer English swear words contain at least one approximant /ɹ/ (see Table 20), fewer at least one nasal /n/

(Table 21), more at least one nasal /ŋ/ (Table 22), more at least one voiced plosive /g/ (Table 23), more at least one voiceless fricative /h/ (Table 24), and more at least one voiceless plosive /k/ (Table 25). These results are presented visually in Figure 7.

	English RW	**English SW**	**English RW**	**English SW**
contain a /r/	3,058	5	**21.2%**	**7.0%**
contain no /r/	11,391	66	78.8%	93.0%
total	14,449	71	100%	100%

*Table 20: significantly fewer English swear words contain at least one /r/ (**$p = 0.001971$**)*

	English RW	**English SW**	**English RW**	**English SW**
contain a /n/	3,803	5	**26.3%**	**7.0%**
contain no /n/	10,646	66	73.7%	93.0%
total	14,449	71	100%	100%

*Table 21: significantly fewer English swear words contain at least one /n/ (**$p = 6.298^{-05}$**)*

	English RW	**English SW**	**English RW**	**English SW**
contain a /ŋ/	226	5	**1.6%**	**7.0%**
contain no /ŋ/	14,223	66	98.4%	93.0%
total	14,449	71	100%	100%

*Table 22: significantly more English swear words contain at least one /ŋ/ (**$p = 0.005421$**)*

	English RW	**English SW**	**English RW**	**English SW**

contain a /g/	792	10	**5.5%**	**14.1%**
contain no /g/	13,657	61	94.5%	85.9%
total	14,449	71	100%	100%

*Table 23: significantly more English swear words contain at least one /g/ (**p = 0.005414**)*

	English RW	**English SW**	**English RW**	**English SW**
contain a /h/	449	7	**3.1%**	**9.9%**
contain no /h/	14,000	64	96.9%	90.1%
total	14,449	71	100%	100%

*Table 24: significantly more English swear words contain at least one /h/ (**p = 0.006777**)*

	English RW	**English SW**	**English RW**	**English SW**
contain a /k/	3,087	29	**21.4%**	**40.8%**
contain no /k/	11,362	42	78.6%	59.2%
total	14,449	71	100%	100%

*Table 25: significantly more English swear words contain at least one /k/ (**p = 0.0002177**)*

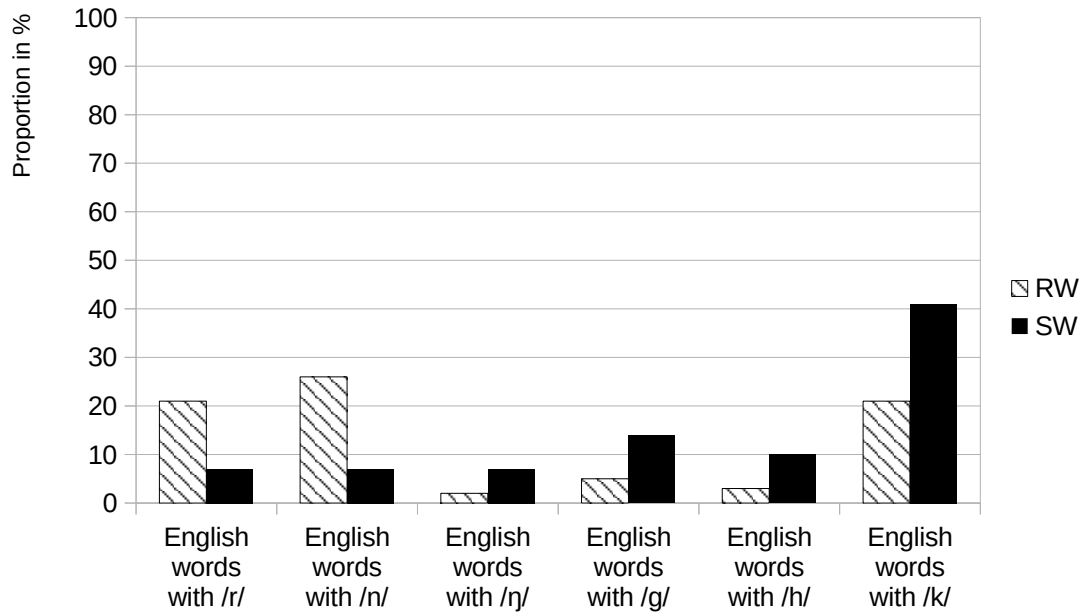

Figure 7: Tendencies for English swear words to contain at least one /ŋ/, /g/, /h/, or /k/, and to contain no /r/ or /n/ ($p<0.01$ for all).

French swear words seem to follow the same overall tendency, as shown in the following results where $p<0.01$. Significantly fewer French swear words contain at least one voiced fricative /v/ (see Table 26), and more at least one voiced plosive /b/ (Table 27). These results are presented visually in Figure 8. We conducted tests on all consonants, but are only presenting here the cases where there was a significant result ($p<0.01$ or $p<0.05$). For this reason, the results concern different phonemes in English vs. in French. This does not invalidate the overall claim that unsonorous consonants tend to be more present in swear words and sonorous consonants to be less present, as seen in Figure 6 above. The same tendency can play out slightly differently in different languages, and other individual tendencies may exist but be too subtle to be detected by such tests.

	French RW	French SW	French RW	French SW
contain a /v/	1,661	0	**11.2%**	**0%**
contain no /v/	13,156	78	88.8%	100%
total	14,817	78	100%	100%

*Table 26: significantly fewer French swear words contain at least one /v/ (**p = 0.0001852**)*

	French RW	French SW	French RW	French SW
contain a /b/	1,424	16	**9.6%**	**21%**
contain no /b/	13,393	62	90.4%	79%
total	14,817	78	100%	100%

*Table 27: significantly more French swear words contain at least one /b/ (**p = 0.003208**)*

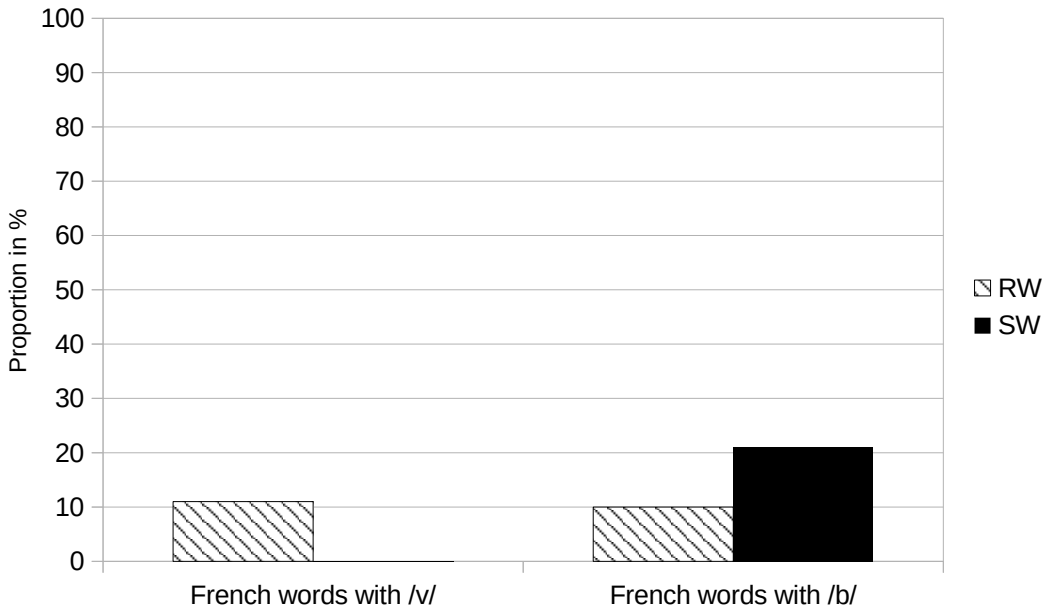

Figure 8: Tendencies for French swear words to contain at least one /b/ and to contain no /v/ (p<0.01 for both)

For the sake of exhaustivity, we can still have a look at the results for English

where p<0.05 (instead of p<0.01). Significantly fewer English swear words contain at least one voiced fricative /v/ (see Table 28), and more at least one voiced plosive /b/ (Table 29). The results are presented visually in Figure 9.

	English RW	**English SW**	**English RW**	**English SW**
contain a /v/	963	0	**6.7%**	**0.0%**
contain no /v/	13,486	71	93.3%	100.0%
total	14,449	71	100%	100%

*Table 28: significantly fewer English swear words contain at least one /v/ (**p = 0.01433**)*

	English RW	**English SW**	**English RW**	**English SW**
contain a /b/	1,289	12	**8.9%**	**16.9%**
contain no /b/	13,160	59	91.1%	83.1%
total	14,449	71	100%	100%

*Table 29: significantly more English swear words contain at least one /b/ (**p = 0.0331**)*

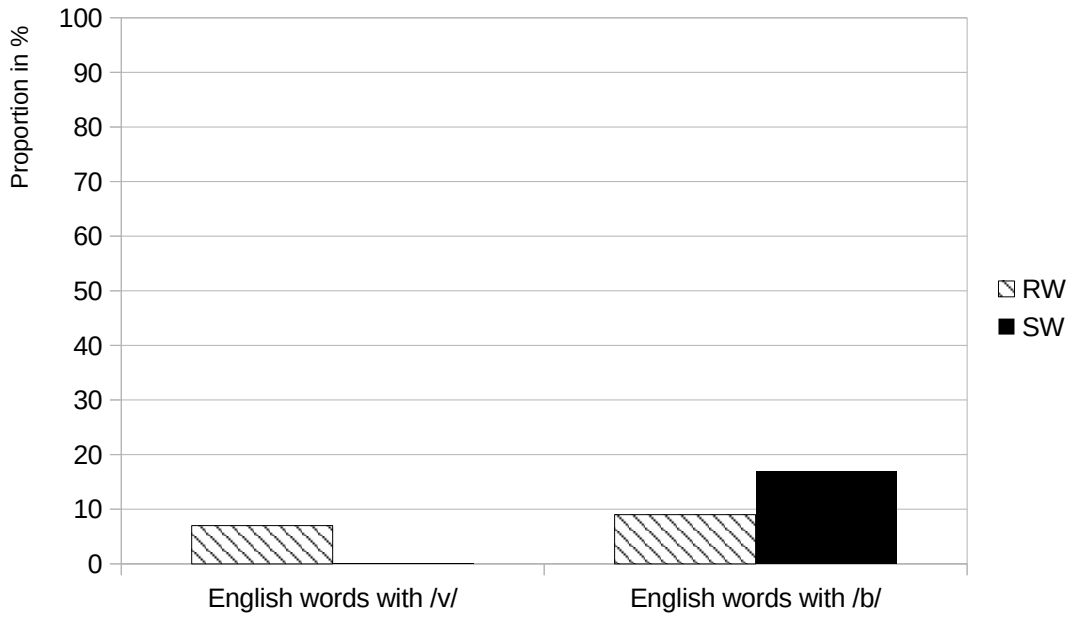

Figure 9: Tendencies for English swear words to contain at least one /b/ and to contain no /v/ (p<0.05 for both).

The equivalent results for French, where p<0.05 (instead of p<0.01) are the following. Significantly fewer French swear words contain at least one nasal /m/ (see Table 30), more at least one nasal /n/ (Table 31), fewer at least one voiced fricative /ʁ/ (Table 32), more at least one voiceless plosive /k/ (Table 33).

These results are presented visually in Figure 10.

	French RW	**French SW**	**French RW**	**French SW**
contain a /m/	2,374	5	**16.0%**	**6%**
contain no /m/	12,443	73	84.0%	94%
total	14,817	78	100%	100%

*Table 30: significantly fewer French swear words contain at least one /m/ (**p = 0.01902**)*

	French RW	**French SW**	**French RW**	**French SW**
contain a /n/	1,831	16	**12.4%**	**21%**

contain no /n/	12,986	62	87.6%	79%
total	14,817	78	100%	100%

*Table 31: significantly more French swear words contain at least one /n/ (**p = 0.03749**)*

	French RW	**French SW**	**French RW**	**French SW**
contain a /ʁ/	7,051	26	**47.6%**	**33%**
contain no /ʁ/	7,766	52	52.4%	67%
total	14,817	78	100%	100%

*Table 32: significantly fewer French swear words contain at least one /ʁ/ (**p = 0.01229**)*

	French RW	**French SW**	**French RW**	**French SW**
contain a /k/	3,001	23	**20.3%**	**29%**
contain no /k/	11,816	55	79.7%	71%
total	14,817	78	100%	100%

*Table 33: significantly more French swear words contain at least one /k/ (**p = 0.04826**)*

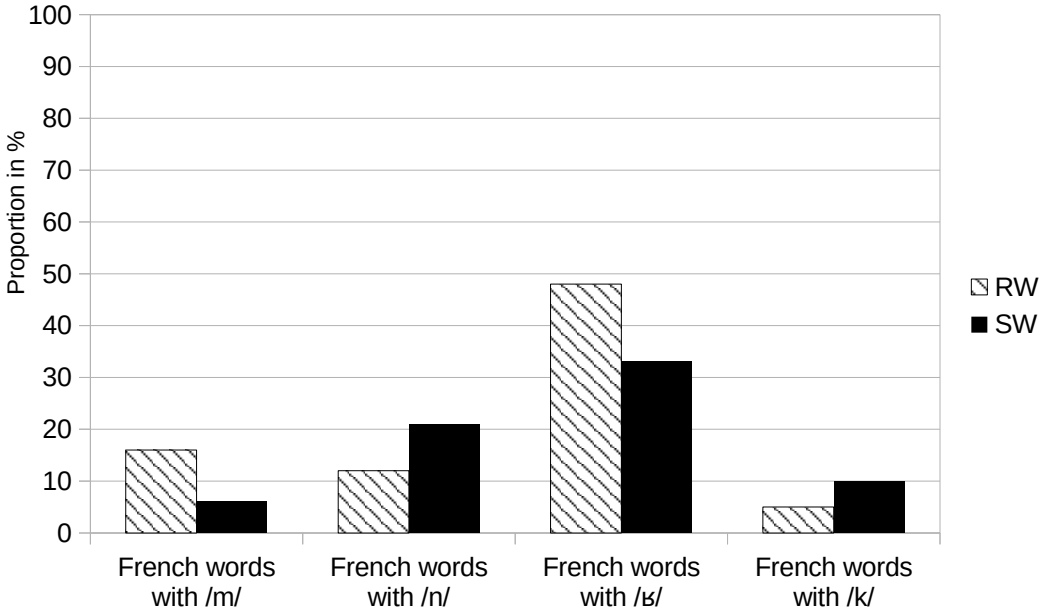

Figure 10: Tendencies for French swear words to contain at least one /n/, or /k/, and to contain no /m/, or /ʁ/ ($p<0.05$ for all)

All the observed tendencies with either $p<0.01$ or $p<0.05$ are recapitulated in Figure 6. Again, despite a few exceptions to the general tendency, the consonants that are more present belong to the least sonorous categories, and conversely, the ones less present belong to the other most sonorous categories. The tests show that sonority is a relevant variable, and that our observed tendency involves the same grouping of consonants as the one proposed by Yardy (2010: 63–64), which includes plosives, voiceless fricatives, and affricates. As far as we can tell, no other categories are involved in the observed tendency: if we look at the slightly more sonorous voiced fricatives, /v/ and /ʁ/ are significantly less present already. It is coherent to include all the least sonorous ones starting from the next, i.e., starting from voiced plosives, since the voiced plosives /b/ and /g/ are significantly more present already.

To evaluate this unsonority tendency, we tested the presence in swear words of all the least sonorous consonants – i.e., plosives, voiceless

fricative, and affricates – taken together. A comparison based on our data on regular vs. swear words in English and French (again, using the Fisher exact test) confirms that when considering the least sonorous categories together, there is indeed a significant tendency for swear words to contain them. More English swear words contain at least one of the least sonorous consonants /p/, /t/, /k/, /tʃ/, /f/, /θ/, /s/, /ʃ/, /h/, /b/, /d/, /g/, or /dʒ/ (see Table 34), while more French swear words contain at least one of the consonants /p/, /t/, /k/, /f/, /s/, /ʃ/, /b/, /d/, or /g/ (the consonants /tʃ/, /θ/, /h/, and /dʒ/ do not exist in French) (see Table 35).

	English RW	**English SW**	**English RW**	**English SW**
contain an unsonorous cons.	11,276	68	**78.0%**	**95.8%**
contain no unsonorous cons.	3,173	3	22.0%	4.2%
total	14,449	71	100%	100%

*Table 34: Significantly more English swear words contain at least one of the least sonorous consonants (**$p = 6.741^{-05}$**)*

	French RW	**French SW**	**French RW**	**French SW**
contain an unsonorous cons.	12,692	76	**85.7%**	**97.4%**
contain no unsonorous cons.	2,125	2	14.3%	2.6%
total	14,817	78	100%	100%

*Table 35: Significantly more French swear words contain at least one of the least sonorous consonants (**$p = 0.0009721$**)*

These results are presented visually in Figure 11.

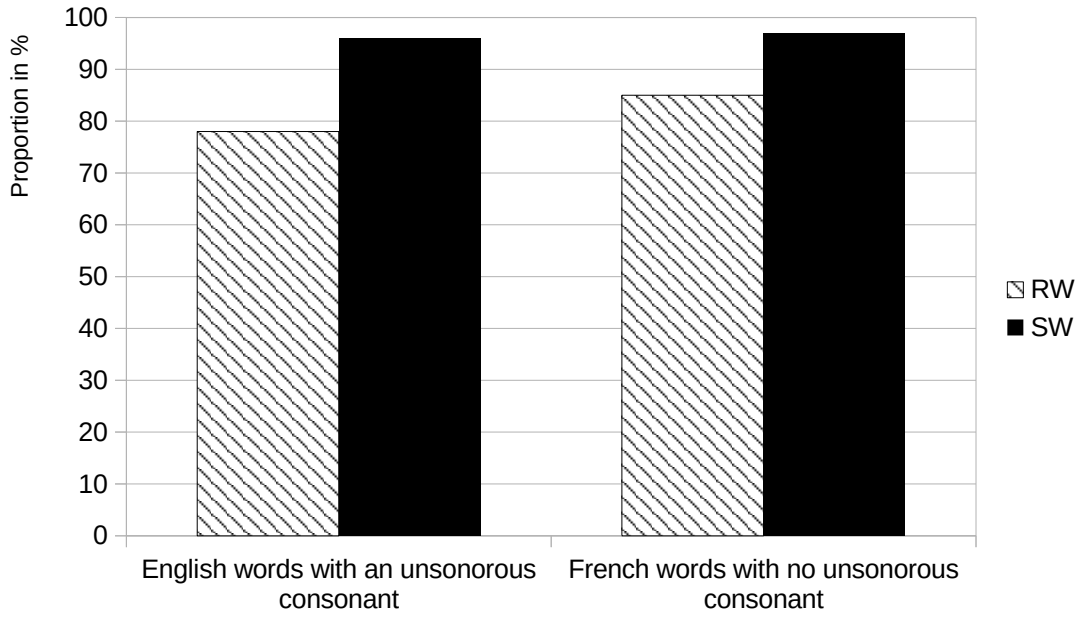

Figure 11: Tendencies for English and French swear words to contain at least one unsonorous consonant ($p<0.01$ for both)

As a further correction to the above result, we need to control for word length, since the longer a word is, the more likely it is to contain at least one plosive or unsonorous consonant. In order to control for length in an reliable way, we counted for the unsonority density of words, i.e., the proportion of unsonorous consonants in one word – see the methodology in Section 3.1.1, where the list of observed densities for English and for French is given.

The results show a statistically significant tendency for swear words to contain a minimal proportion (0.33 or more) of these consonants irrespective of length. Significantly more English swear words have an unsonority density of 0.33 or more, which means that at least one phoneme out of three is one of the least sonorous consonants (see Table 36), and the same is true for French (Table 37). These results are presented visually in Figure 12. Again, the number

of English regular words is much lower than in our previous tests (4,849 instead of 14,449), because only some of them are annotated for their number of phonemes in the database we use.

	English RW	**English SW**	**English RW**	**English SW**
have a 0.33 density or higher	2,917	60	**60.4%**	**84.5%**
have a density lower than 0.33	1,916	11	39.6%	15.5%
total	4,833	71	100%	100%

*Table 36: significantly more English swear words have a 0.33 unsonority density or more (**$p = 2.133^{-05}$**)*

	French RW	**French SW**	**French RW**	**French SW**
have a 0.33 density or higher	6,910	49	**46.5%**	**62.8%**
have a density lower than 0.33	7,965	29	53.5%	37.2%
total	14,875	78	100%	100%

*Table 37: significantly more French swear words have a 0.33 unsonority density or more (**$p = 0.0009721$**)*

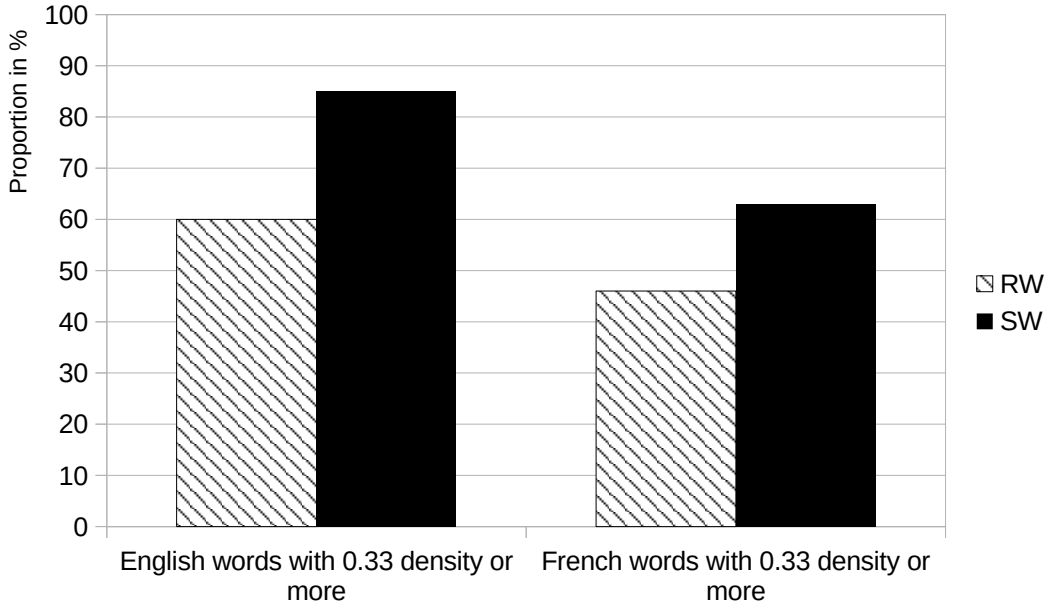

Figure 12: Tendencies for English and French swear words to contain at least one unsonorous consonant out of three phonemes (0.33 density) ($p<0.01$ for both).

Our first case study thus answered our first research question (RQ1). English and French swear words do tend to contain more unsonorous consonants. It remains to be seen if this tendency corresponds to a cognitively real pattern (RQ2b). This is is the aim of our second and third case studies which investigate swear words from fiction, and experimental swear words.

3.2 Case Study 2: Fictional swear words

Our second case study presents a counterpart to the first one by looking at swear words from fiction. We look at swear words created and reused by English-speaking and French-speaking authors of fiction to serve as swear words in their fictional universes. If they follow the same tendency, then it is an

indication that this sound-meaning association is not merely statistical, but also exists in the authors' minds. Our results turn out to be significant for English words, and the non-significant for French, even though the tendency is observable still.

3.2.1 Method for Fictional swear words

When authors invent names for characters, objects, or places in their fictional worlds, they follow their creative intuition. Such intuition also applies to what could be (or rather, sounds like) a convincing swear word when creating a swear word from scratch (invented swear words) or when they select an existing word to serve as a swear word in their fictional world (pseudo swear words). Both types occur in our dataset, but this distinction is not essential to our study of the form of fictional swear words. What is essential is whether, when inventing a new word or recycling an existing one, authors are biased to building or selecting words with unsonorous consonants. If so, this would confirm the unconscious association between swear words and certain consonants that was observed with real life swear words.

For this study, we used two existing lists of fictional words, one for English and one for French, and compared the swear words they contained to the regular lexicon, using the same databases as in the first case study.

The list of English fictional swear words comes from the collaborative Wiktionary page "Fictional English curse words" (Wiktionary 2009).[13] It contains 44 items, with a broad scope in terms of genres

13 https://en.wiktionary.org/wiki/Appendix:Fictional_English_curse_words Last accessed on

encompassing fantasy, science-fiction, superhero comics, space operas, and others. We listed *frak* or *frack*, appearing in *Battlestar Galactica* and *Warhammer 40K*, only once as *frak* because it is a case of homophony. This resulted in a list of 43 English fictional swear words, In order to determine whether a given item was an invented or a pseudo swear word, we checked whether or not it occurred in the online Oxford English Dictionary (OED) [14]. Our list includes 21 invented words, sixteen pseudo swear words – see above for the difference between invented and pseudo swear words – and six words of debatable status: *drek* is a homophone of the existing word *dreck*, *frack* and *slitch* are obsolete words of English, *mogadored* is based on the existing word *Mogador*, *sithspit* is based on the existing words *sith* and *spit*, and *stoom* is dialectal. With respect to the annotation of phonemes, we have assumed that fictional words follow the usual pronunciation rules of English because authors would want their audience to know how to pronounce them. As in the first case study, we removed fictional swear words that happened to be among the top 10% in the MRC Psycholinguistic Database, to avoid comparing swear words with swear words. The list of the 43 English fictional swear words, with their frequencies, is given below in Table 38.

Fictional English swear word	IPA	Source	Origin	Frequency (MRC)
algebraic	/ˌældʒɪˈbreɪɪk/	*Adventure Time*	pseudo	0
bastich	/ˈbɑːstɪtʃ/	*Judge Dredd*	invented	NA

31 May, 2019.

14 http://www.oed.com/ Last accessed on 28 January, 2020.

Fictional English swear word	IPA	Source	Origin	Frequency (MRC)
Belgium	/ˈbeldʒəm/	*Hitchhiker's Guide to the Galaxy*	pseudo	0
bleep	/ˈbliːp/	*Known Space* stories	pseudo	0
cabbage	/ˈkæbɪdʒ/	*Adventure Time*	pseudo	1
censor	/ˈsensə/	*Known Space* stories	pseudo	0
censored	/ˈsensəd/	*Podkayne of Mars* (book)	pseudo	0
cuss	/ˈkʌs/	*Fantastic Mr. Fox* (film)	pseudo	0
delete	/dɪˈliːt/	*Podkayne of Mars* (book)	pseudo	0
drek	/ˈdrek/	*Shadowrun*	(dreck)	NA
dren	/ˈdren/	*Farscape*	invented	NA
drokk	/ˈdrɒk/	*Judge Dredd*	invented	NA
felgercarb	/ˈfelgəkɑːb/	*Battlestar Galactica*	invented	NA
feth	/ˈfeθ/	*Warhammer 40K*	invented	NA
frag	/ˈfræg/	*Shadowrun*	pseudo	NA
frak	/ˈfræk/	*Battlestar Galactica*	(obsolete)	NA
frell	/ˈfrel/	*Farscape*	invented	NA
frinx	/ˈfrɪŋks/	*Star Trek*	invented	NA
Glob	/ˈglɒb/	*Adventure Time*	pseudo	NA
grife	/ˈgraɪf/	*Legion of Super-Heroes* comics	invented	NA
grud	/ˈgrʌd/	*Judge Dredd*	invented	NA
Hippikaloric	/ˌhɪpɪkaˈlɒrɪk/	*Ozma of Oz*	invented	NA
lump	/ˈlʌmp/	*Adventure Time*	pseudo	1
math	/ˈmæθ/	*Adventure Time*	pseudo	0
mee krob	/miːˈkrɒb/	*South Park*	pseudo	NA
mivonks	/mɪˈvɔːŋks/	*Farscape*	invented	NA

Fictional English swear word	IPA	Source	Origin	Frequency (MRC)
mogadored	/ˈmɒgədɔːd/	*Discworld series*	(Mogador)	NA
petaQ	/ˈpetæk/	*Star Trek*	invented	NA
shards	/ˈʃɑːdz/	*Dragonriders of Pern* series	pseudo	0
shazbot	/ˈʃɑːzbɒt/	*Mork and Mindy*	invented	NA
shock	/ˈʃɒk/	*Marvel 2099* comics	pseudo	3
sithspit	/ˈsɪθspɪt/	*Star Wars* Expanded Universe	(sith)	NA
slitch	/ˈslɪtʃ/	*Friday* (novel)	(obsolete)	NA
smeg	/ˈsmeg/	*Red Dwarf*	invented	NA
smurf	/ˈsmɜːf/	*The Smurfs* (2011 film)	invented	NA
sprock	/ˈsprɒk/	*Legion of Super-Heroes* comics	invented	NA
squaj	/ˈskwɑːdʒ/	*Legion of Super-Heroes* comics	invented	NA
squill	/ˈskwɪl/	*Firebird* (Tyers novel)	pseudo	0
stomm	/ˈstɒm/	*Judge Dredd*	(dialectal)	NA
tanj	/ˈtɑːndʒ/	*Known Space* stories	invented	NA
veruul	/ˈveruːl/	*Star Trek*	invented	NA
yarbles	/ˈjɑːbəlz/	*A Clockwork Orange*	invented	NA
zark	/ˈzɑːk/	*Hitchhiker's Guide to the Galaxy*	invented	NA

Table 38: 43 fictional English swear words by alphabetical order

One could argue that some fictional swear words are coined based on a similarity with an existing swear word (for example *frak* is similar to *fuck*). However, as seen in the list, few fictional swear words show such close similarity. We only count five arguable candidates out of 43: *bastich* which is close to *bastard* and *bitch*, *frak* close to *fuck*, *Glob* and *grud* close to *God*, and *slitch* close to *slut* and *bitch*. The same is true for French fictional swear words, where

we find no example of such close similarity at all.

The list of fictional swear words for French is drawn from a source that is extremely well-known among French speakers: the language used by the character Captain Haddock in the Belgian comic book series *The Adventures of Tintin*, by cartoonist Hergé, published in a period ranging from the 1930s to the 1970s. Captain Haddock uses words or phrases whose semantic meaning is obscure or irrelevant to the narrative context and which can then be interpreted as swearing. Captain Haddock's pseudo swear words are so well-known and frequent throughout the series that they are documented – along with other expressions – on a French-language Wikipedia page entitled *Vocabulaire du Capitaine Haddock* ("Captain Haddock's vocabulary") (Wikipedia 2004).[15] A limitation of our choice to use that list is that those fictional swear words were all created by one writer, whereas the English fictional swear words come from multiple writers. Despite this limitation, they are extremely relevant to our purposes. In a 1964 interview for French television, the author Hergé confirmed explicitly that he selected those words based on his gut feeling about how good they sounded as swear words:

> Already in [the volume] *The Crab with the golden claws*, yes, that's it, that's when he starts uttering those swear words, but they are not at all monstrous, as you say. Those are words that have a certain sonority to them, like *ectoplasme* ["ectoplasm" or "superficial person"] or, I don't know, *moule à gaufre* ["waffle iron"] but that are not insults in themselves. *Analphabète* ["illiterate"], true, I admit that it is an insult, but... [...] *Anacoluthe* ["anacoluthon"] is not an insult

15 https://fr.wikipedia.org/wiki/Vocabulaire_du_capitaine_Haddock. Last accessed on 31 May, 2019.

> but in Captain Haddock's mouth, I feel, maybe I am wrong, that it takes some depth, some allure that is really, really violent. There is something. [...] [I used] words, then, that were not insults but had a sonority which could pass, which made it possible that they would pass as insults. (Institut National de l'Audiovisuel n.d.) (our translation)[16]

Hergé discusses only swearing insults in this interview. Most of Captain Haddock's pseudo swear words are indeed insults, but they also include some interjections, for example *Mille millions de mille sabords!* literally meaning "One thousand million of a thousand gun ports!"

A potential concern here is that most existing swear words are insults (like *fucker*), interjections (*fuck!*), or intensifiers (*fucking* preceding a noun or adjective), but the reverse is not always true: for example, *crook* and *idiot* are insults but not swear words in English. As a consequence, one can have the impression that Captain Haddock's words are actually only (pseudo) insults and (pseudo) interjections, but not (pseudo) swear words. Maybe Haddock is insulting or interjecting, in his own fictional way, but he is not swearing. In that case, those fictional words would not be relevant to our research question. However, in the quote above Hergé confirms that they are indeed swear words, not just insults, so that concern can be safely dismissed.

While one should be careful with open source data, the Wikipedia page on Captain Haddock's vocabulary is reliable given how famous *The Adventures of Tintin* are across the French-speaking community, ensuring considerable peer control. In total, the page lists 421 expressions. In order to

16 https://fresques.ina.fr/europe-des-cultures-fr/fiche-media/Europe00025/herge-a-propos-du-capitaine-haddock.html. Last accessed on 6 June, 2019.

allow a comparison with our French lexical database, all multi-word expressions were removed, which reduced the list to 221. In addition, expressions that had no reference to a specific page of the comic books (whose actual presence in the series could thus not be verified) and two onomatopoeic expressions with no clear pronunciation (*KRRTCHMVRTZ* and *MRKRPXZKRMTFRZ*) were also left out, which reduced the list to 186 words.

Collaborators of the "Captain Haddock's Vocabulary" Wikipedia page have collected all items they considered as rather typical of Captain Haddock. They cast their net wide, so to speak. Consequently, our list ends up with many debatable cases. There is a continuum of prototypicality between words that are clearly pseudo swear words because their semantics are completely irrelevant (like *anacoluthe* "anacoluthon" or *moule à gaufre* "waffle iron"), words that are possibly used as non-swearing creative insults or interjections because their semantics allow it (like *doryphore* "Colorado potato beetle" or *ectoplasm* "ectoplasm" or "superficial person"), and words that are very probably used as mere non-swearing insults, interjections, or intensifiers (like *bandit* "bandit").

To address this issue, we did some further clean-up to reach the final selection. First of all, we removed all words that are documented in French dictionaries as existing interjections or intensifiers, like *tonnerre!* (literally "thunder! [interjection]"), or *satané* ("dreadful [intensifier before a noun]"). For this, we checked the largest online French dictionary, the *Trésor de la langue Française informatisé* (*TLFi* for short)[17] and the Larousse online French-English dictionary[18] – the English glosses for each word listed in the Appendix are also

17 http://atilf.atilf.fr/. Last accessed on 29 September, 2024.

18 https://www.larousse.fr/dictionnaires/anglais-francais/English/578460 Last accesssed on 29

retrieved from those dictionaries. Since they are existing interjections and intensifiers, the chances are too high that they were simply used as regular interjections or intensifiers and not for their sonority. This reduced the list to 179 words.

The 179 words include those referring to inanimate objects, like *anacoluthe* ("anacoluthon"). As it is much more common to insult people by comparing them to animate beings, words referring to inanimate objects are probably used solely for their sonority, so we kept all of them (25) in the final list.

We also removed unreliable items within the 154 animate words, based on two criteria. The first criterion is their frequency (in spoken French) in Lexique. Obscure infrequent words are probably used as pseudo swear words solely for their sonority, whereas well-known more frequent words such as *bandit* ("bandit") are probably used as real-life, non-swearing insults. The second criterion is their frequency of use throughout the comic series. Repeated use can be interpreted as an indication that the author consistently felt that it sounded convincing as a pseudo swear word. Hergé provided two examples of pseudo swear words in his interview quoted above: *analphabète* ("illiterate") and *ectoplasme* ("ectoplasm" or "superficial person"). *Analphabète* ("illiterate") occurs three times in the comics and has a frequency of 0.35 per million words (pmw) in spoken French according to Lexique. *Ectoplasme* ("ectoplasm") occurs sixteen times in the comics and has a frequency of 0.6 per million words in spoken French. Words that occur at least three times in the comics and have a frequency below 0.6 pmw are thus likely to be prototypical Captain Haddock pseudo swear words. We removed all others which leaves us with 23 animate

September, 2024.

words.

We also removed the only item that is certainly not a pseudo swear word given its context of use, that is *moussaillon* ("ship's boy"), which Captain Haddock uses to address the protagonist Tintin in a friendly and fatherly manner.

The final list contains 47 words, 25 with inanimate reference and 22 with animate reference. The list is given below complete with English glosses in Table 39.

Fictional French swear word	IPA	Category	Comics/ Lexique freq.
ectoplasme ("ectoplasm"/[superficial person])	/ɛktoplasm/	ANIMATE	16/0.6
bachi-bouzouk ("bashi-bazouk")	/baʃibuzuk/	ANIMATE	14/0
zouave ("clown/fool")	/zwav/	ANIMATE	13/0.46
iconoclaste ("iconoclast")	/ikonoklast/	ANIMATE	8/0.1
flibustier ("freebooter"/"buccaneer")	/flibystje/	ANIMATE	7/0.04
anthropopithèque ("Anthropopithecus")	/ɑ̃tʁopopitɛk/	ANIMATE	6/0
forban ("freebooter"/"crook")	/fɔʁbɑ̃/	ANIMATE	6/0.03
sapajou ("sapajou"/"capuchin monkey")	/sapaʒu/	ANIMATE	6/0.01
va-nu-pieds ("tramp"/"beggar")	/vanypje/	ANIMATE	6/NA
anacoluthe ("anacoluthon")	/anakolyt/	INANIMATE	5/0.1
boit-sans-soif ("drunk"/"lush")	/bwasɑ̃swaf/	ANIMATE	5/NA

Fictional French swear word	IPA	Category	Comics/ Lexique freq.
écraseur (“road hog“)	/ekʁazœʁ/	ANIMATE	5/0.19
emplâtre (“plaster“/[unefficient person])	/ɑ̃platʁ/	ANIMATE	5/0.12
olibrius (“oddball“)	/olibʁijys/	ANIMATE	5/0.03
cloporte (“woodlouse“)	/klopɔʁt/	ANIMATE	4/0.13
doryphore (“Colorado potato beetle“)	/doʁifɔʁ/	ANIMATE	4/0.01
Zoulou (“Zulu“)	/zulu/	ANIMATE	4/0.06
analphabète (“illiterate“)	/analfabɛt/	ANIMATE	3/0.35
anthropophage (“anthropophagite“)	/ɑ̃tʁopofaʒ/	ANIMATE	3/0.01
Canaque (“Kanak“)	/kanak/	ANIMATE	3/0
catachrèse (“catachresis“)	/katakʁɛz/	INANIMATE	3/0.01
coléoptère ([specific type of beetle])	/koleɔptɛʁ/	ANIMATE	3/0.38
coloquinte (“colocynth“)	/kolokɛ̃t/	INANIMATE	3/0
naufrageur (“[ship]wrecker“)	/nofʁaʒœʁ/	ANIMATE	3/0
troglodyte (“troglodyte“/“cave dweller“)	/tʁoglodit/	ANIMATE	3/0.1
ophicléide (“ophicleide“)	/ofikleid/	INANIMATE	2/NA
aérolithe (“meteorite“)	/aeʁolit/	INANIMATE	1/0
ascenseur (“lift“/“elevator“)	/asɑ̃sœʁ/	INANIMATE	1/22.87
bibelot (“curio“/“bibelot“/“trinket““)	/bibəlo/	INANIMATE	1/0.63
cake-walk ([name of a dance])	/kɛkwɔk/	INANIMATE	1/0.03
calembredaine (“nonsense“)	/kalɑ̃bʁədɛn/	INANIMATE	1/0

Fictional French swear word	IPA	Category	Comics/ Lexique freq.
cataplasme (“poultice“/“cataplasm“)	/kataplasm/	INANIMATE	1/0.08
cornemuse (“bagpipes“)	/kɔʁnəmyz/	INANIMATE	1/0.98
cyanure (“cyanide“)	/sjanyʁ/	INANIMATE	1/2.23
cyclone (“cyclone“/“hurricane“)	/siklon/	INANIMATE	1/1.33
cyclotron (“cyclotron“)	/siklotʁɔ̃/	INANIMATE	1/0.16
gargarisme (“gargling“/“mouthwash“)	/gaʁgaʁizm/	INANIMATE	1/0.02
gyroscope (“gyroscope“)	/ʒiʁoskɔp/	INANIMATE	1/0.46
hydrocarbure (“hydrocarbon“)	/idʁokaʁbyʁ/	INANIMATE	1/0.11
logarithme (“logarithm“)	/logaʁitm/	INANIMATE	1/0.05
mégacycle (“a thousand cycles“)	/megasikl/	INANIMATE	1/NA
pantoufle (“slipper“)	/pɑ̃tufl/	INANIMATE	1/0.57
poussière (“dust“)	/pusjɛʁ/	INANIMATE	1/22.77
rocambole (“rocambole“/“sand leak“)	/ʁokɑ̃bɔl/	INANIMATE	1/0
sinapisme ([specific type of food])	/sinapizm/	INANIMATE	1/0
topinambour (“Jerusalem artichoke“)	/topinɑ̃buʁ/	INANIMATE	1/0.02
vermicelles (“vermicelli“)	/vɛʁmisɛl/	INANIMATE	1/0.14

Table 39: List of 47 fictional French swear words by order of frequency of use in the comics.

We compare those French fictional swear words to the regular words of French drawn from Lexique. The comparisons between fictional swear words versus regular words were done following the same method as in the first case study. For the density comparison, 0.33 was chosen as the threshold based on the

prototypical Captain Haddock pseudo swear words provided by the author in the interview quoted above. The words *analphabète* ("illiterate") and *ectoplasme* ("ectoplasm" or "superficial person") have densities of 0.33 and 0.44 respectively. This implies that a 0.33 density is enough to sound like a convincing swear word. We used this threshold for fictional swear words, but also for all the other tests throughout this dissertation, so that the results can be compared easily. The results of the previous study on real-life swear words in English and French indicated that unsonorous consonants are more frequent in swear words. For that reason, we compared the unsonority density of English and French swear words vs. regular words.

3.2.2 Results for Fictional swear words

The results confirm that just like real-life swear words, fictional swear words tend to contain a higher proportion of unsonorous consonants: significantly more English fictional swear words have an unsonority density of 0.33 or more, which means that at least one phoneme out of three is one of the least sonorous consonants (see Table 40). French shows a similar tendency, which however does not reach statistical significance ($p<0.08$) (see Table 41). These results are presented visually in Figure 23.

	English RW	English SW	English RW	English SW
have a 0.33 density or higher	2,930	37	**60.5%**	**86.0%**
have a density lower than 0.33	1,916	6	39.5%	14.0%
total	4,846	43	100%	100%

*Table 40: significantly more fictional English swear words have a 0.33 unsonority density or more (**p = 0.0004288**)*

	French RW	French SW	French RW	French SW
have a 0.33 density or higher	6,936	28	**46.5%**	**59.6%**
have a density lower than 0.33	7,986	19	53.5%	40.4%
total	14,922	47	100%	100%

*Table 41: more fictional French swear words have a 0.33 unsonority density or more (**p = 0.07947**) (N.S.)*

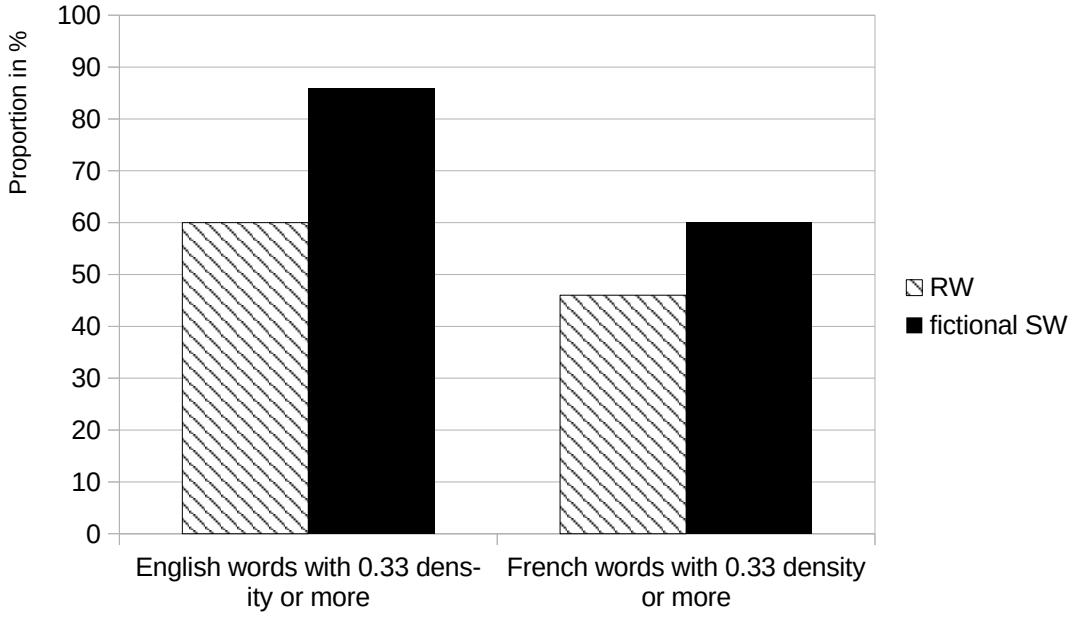

Figure 13: Tendencies for fictional swear words to contain at least one unsonorous consonant out of three phonemes (0.33 density) (p<0.01 for English while p<0.08 for French)

In sum, even if the results do not reach statistical significance for French, fictional swear words in both English and French follow the same tendency as their real-life counterparts. This strongly suggests that there is an unconscious sound-meaning tendency related to swear words that speakers exploit when creating new words or selecting existing words to serve as swear words in their fictional universes (RQ2a). Our study shows limitations in the fact that they gather open source data collected non-systematically by online contributors, and that they are the creative product of only authors – in the case of the French words, of one author – who might not be representative of their entire speech community. Our next study aims to avoid those limitations by devising an experimental protocol to elicit new words from English and French speakers.

3.3 Case Study 3: Experimental swear words

Our third and last case study aims to provide more definite proof on whether there is an unconscious sound-meaning relationship in English and French speakers. We ask native speakers to playfully invent words from an alien, i.e., extraterrestrial language of the kind that may be invented for science-fiction works. The results are significant for both English and French: invented swear words contain more unsonorous consonants compared to the invented non-swear words. This suggests, again, that there is an unconscious sound-meaning association that speakers unconsciously exploit, which is the object of our second research question (RQ2a).

3.3.1 Method for Experimental swear words

In order to overcome the limitations of the previous study on English and French fictional swear words, we devised an experimental protocol to check whether speakers of English and French indeed use more unsonorous consonants when asked to playfully and spontaneously invent swear words. The next subsections describe the protocol, participants, and data annotation process for this last case study.

3.3.1.1 Protocol

The experiment could happen either in in-person interaction with an instructor, or online via videoconference with the same instructor. In the in-person protocol, the respondent was sitting in front of a computer connected to an online questionnaire for the experiment. They were guided through the experiment by the instructor standing next to them, who also recorded their spoken answers with their informed consent. In the online protocol, the respondent met the instructor via videoconference. They shared their screen so the instructor could see what question they were at and guide them accordingly. Their spoken answers were also recorded with their informed consent.

The questionnaire was in English for English speakers, French for French speakers. Before the questionnaire proper, on the first page of the questionnaire, the respondent answered questions about their profile (the same as Case Study 1, but also in English). The questions for English speakers, and

their equivalents in French, were:

- *How old are you? / Quel est ton âge ?*
- *What is your gender? / Quel est ton genre ?*
- *What is your job or occupation? / Quel est ton métier ou occupation ?*
- *What is the highest diploma you got? / Quel est le diplôme le plus avancé que tu as obtenu ?*
- *Is English your native language, and in what country were you born? / Est-ce que le français est ta langue maternelle, et quel est ton pays de naissance ?*

The last question about native language aimed to confirm that the respondent was a native speaker of English or French and collect what national variety of English or French they were a native speaker of, e.g., British English vs. American English, or French of France vs. Belgian French. The other questions provide us a basic profile of our respondents.

When the respondent arrived at the second page of the questionnaire, the instructor explained to them in more detail what the experiment consisted in. Before that, they had been told that the experiment was about inventing new words, but not that it was about inventing *alien* words, to make sure that the respondents could not think in advance about what a truly alien, extra-terrestrial word would sound like and come up with sounds difficult or impossible to transcribe into IPA, which was not the point of the experiment. The point of alien words was to obtain data as unrelated to existing words as possible, rather than variations of existing swear words like *fuck*, *shit*, or others. Our aim is to observe tendencies independently of existing words.

The same general explanation below was also written down on the

page to avoid any misunderstandings.

In this questionnaire, I'm asking you to playfully invent words from ALIEN, EXTRA-TERRESTRIAL languages – the kind of alien words you can find in science-fiction, TV series, films, comic books, etc.

For every answer, please first SAY your answer ALOUD, then TYPE IT on this web page.

I will record your audio answers, with your permission.

Please be as SPONTANEOUS as possible, and look at only 1 question at a time. (Don't scroll down too fast!)

Pour ce questionnaire, je te demande d'inventer ludiquement des mots EXTRA-TERRESTRES, des mots ALIENS, comme il y a dans les B.D.s de science-fiction, séries télé, films de science-fiction, etc.

Pour chaque question, DIS ta réponse À VOIX HAUTE, puis TAPE-LA sur la page web.

Je vais enregistrer les réponses à voix haute, avec ta permission.

Sois le plus SPONTANÉ possible, et ne lis qu'une question à la fois (ne descends pas sur la page trop vite!).

The general instructions were also illustrated by a picture of the same style used for the questions (see below).

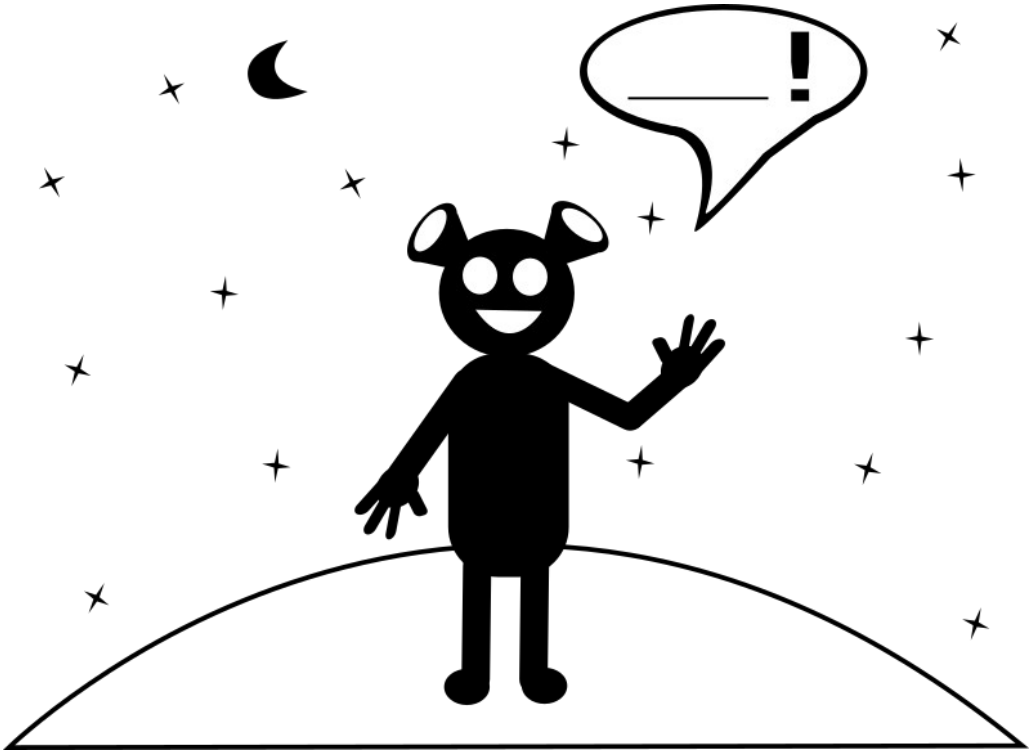

Figure 14: Illustration to help respondents conceptualise what kind of words they were going to invent in this questionnaire

Once the respondent gave their permission to be recorded, they went to the next page of the questionnaire, where they answered eight questions. Each question asked to invent a specific word with a pragmatic, contextual-emotional meaning, e.g., a greeting, or thanks, given that swear words have a contextual-emotional meaning, which for every question, we put in capital letters – see below. Each question was illustrated to help visualise the type of context where the fictional word could be used. We drew these illustrations

with the graphic editor software Inkscape[19] and used minimal variation so that the differences between pictures would influence the respondents' answers as little as possible. The order of questions was randomised for every respondent. We asked four questions about swear words and four questions about non-swear words as a baseline – to exclude that *all* fictional alien words are heavy on unsonorous consonants, not just alien swear words. The list of questions in English and French is given below along with the corresponding illustrations.

19 https://inkscape.org/ Last accessed on 30 September, 2024.

1. *Invent an alien word used to GREET EACH OTHER*

   Invente un mot alien qu'on utilise pour SE SALUER

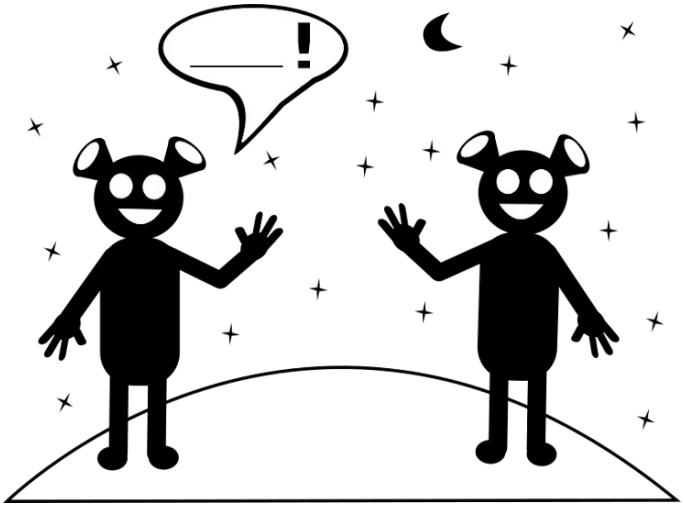

Figure 15: Illustration for question 1

2. *Invent an alien word used when LEAVING EACH OTHER*

   Invente un mot alien qu'on dit lorsqu'on SE QUITTE

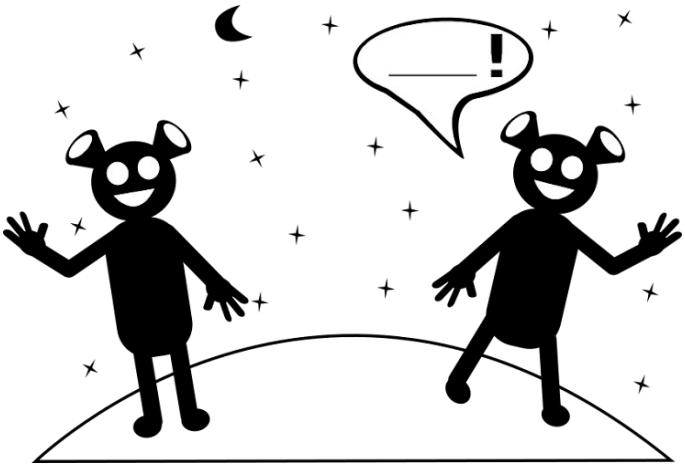

Figure 16: Illustration for question 2

3. *Invent an alien word used to REQUEST SOMETHING POLITELY*

   Invente un mot alien qu'on dit pour DEMANDER QUELQUE CHOSE POLIMENT

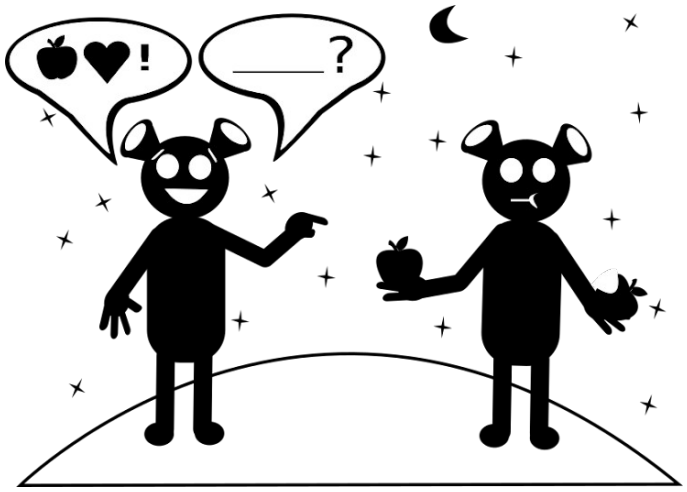

Figure 17: Illustration for question 3

4. *Invent an alien word used to EXPRESS GRATITUDE*

   Invente un mot alien qu'on dit pour EXPRIMER SA GRATITUDE

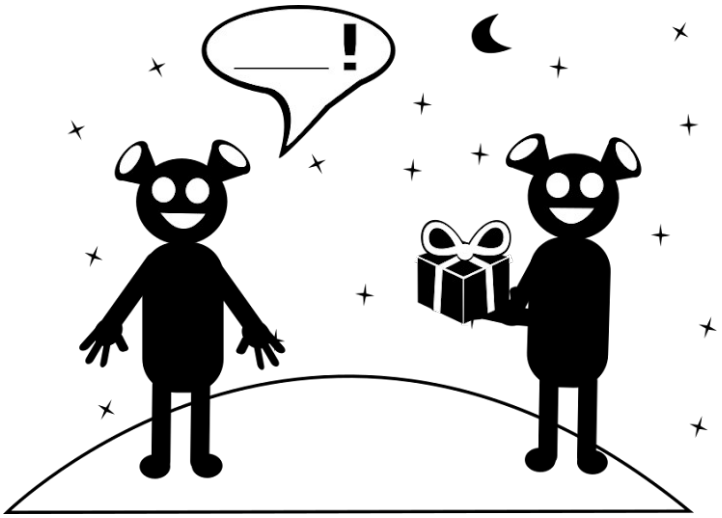

Figure 18: Illustration for question 4

5. *Invent an alien SWEAR WORD used when YOU ARE SURPRISED*

   Invente un GROS MOT alien qu'on dit quand ON EST SURPRIS

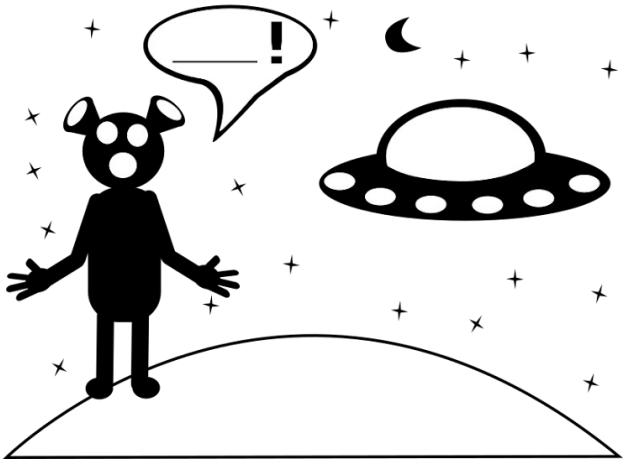

Figure 19: Illustration for question 5

6. *Invent an alien SWEAR WORD used when YOU FEEL ANGRY*

   Invente un GROS MOT alien qu'on dit quand ON EST EN COLÈRE

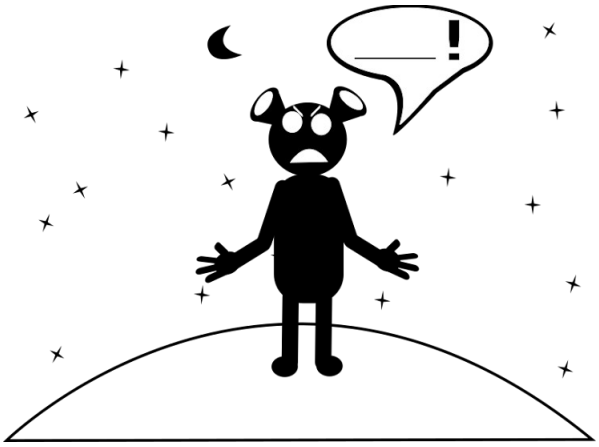

Figure 20: Illustration for question 6

7. *Invent an alien SWEAR WORD which is an INSULT*

   Invente un GROS MOT alien qui est UNE INSULTE

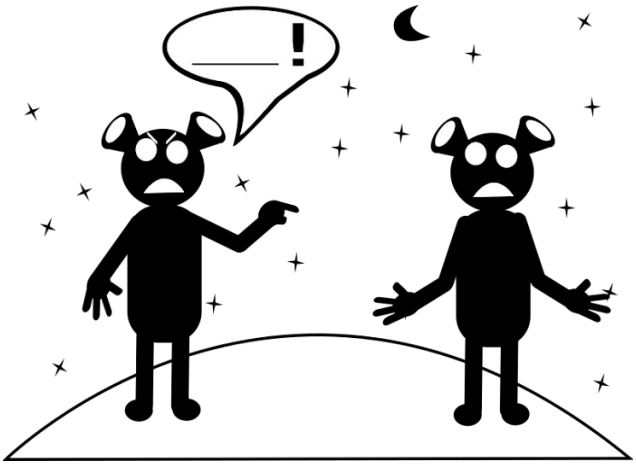

Figure 21: Illustration for question 7

8. *Invent alien INSULTS against OTHER ALIEN GROUPS*

   Invente des INSULTES aliens contre D'AUTRES GROUPES ALIENS

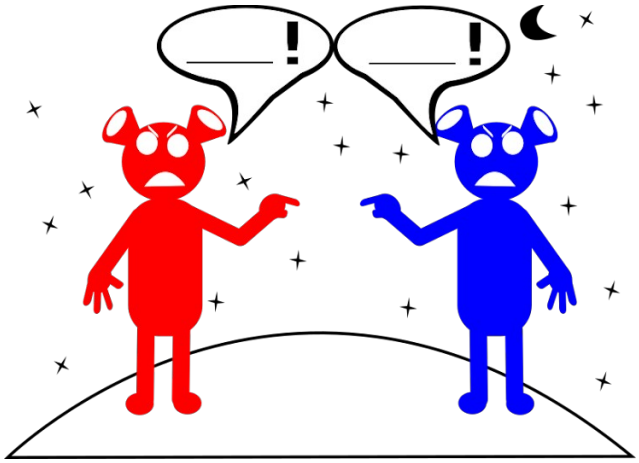

Figure 22: Illustration for question 8

Question 7 and question 8 were meant to elicit two different types of insults: question 7 asks for swearing insults, i.e., alien equivalents of *bastard* or *asshole*, while question 8 asks for slurs, i.e., aliens equivalents of *spic* or *chink*. Question 8 prompted respondents to invent two words: "invent alien words that are insults against other alien groups" with two speech bubbles for two aliens of differing colours insulting each other, suggesting that each belong to a separate group of aliens – see Figure 22. The question was formulated and illustrated in this symmetrical way in order to discourage respondents from taking inspiration from either of the two colours when inventing the insult.

After the respondent said aloud and typed down their answers, they were shown a screen with a thank-you note for their participation: *That's it! Thanks a lot! Please click on "Submit" below to send your written answers.* In French: *C'est tout ! Merci beaucoup ! Clique sur le bouton ci-dessous pour envoyer tes réponses écrites.* The instructor then stopped the audio recording.

3.3.1.2 Participants

Participants were recruited mostly among university students and students' own acquaintances, from September 2020 to January 2023. We explicitly asked for native speakers of English, or native speakers of French – a non-native speaker would not be allowed to participate in the experiment. Depending on the respondents' availabilities, they participated either in-person or online – 13 (40.6%) English speakers participated in person and 19 participated online. 15 (46.9%) French speakers participated in person and 17 (53.1%) participated online.

Our English-speaking respondents were aged from 18 to 40 years old – a majority 23 respondents (71.9%) were in their twenties or younger. 22 (69.8%) were women, the rest were men. Eighteen participants (56.3%) reported that they were born in the United States, ten (31.3%) in the United Kingdom, one in Canada, one in Ireland, and two born in China and Malaysia reported that they were native speakers of American English. Eleven (34.4%) were students, and the others reported various occupations like teaching, engineering, or bartending. Their diplomas ranged from high-school level, for example A-levels in the United Kingdom, to PhD.

Respondent	Country of birth	Age	Gender	Job/occupation	Diploma	Online or in person
1	UK	19	F	Student	A-Levels	online
2	UK	19	F	Student	A-Levels	online
3	UK	21	M	Student	A-Levels	online
4	UK	19	M	Student	A-Levels	online

Respondent	Country of birth	Age	Gender	Job/occupation	Diploma	Online or in person
5	UK	25	F	Teacher	Bachelor's degree	online
6	UK	34	F	Research Fellow	PhD	online
7	USA	22	M	Student	High school level	online
8	Ireland	23	M	College lecturer	Bachelor's degree	in person
9	UK	24	F	College lecturer	Master's degree	in person
10	UK	23	F	College lecturer	Bachelor's Degree	in person
11	USA	31	F	College lecturer	Master's degree	online
12	USA	23	M	College lecturer	Bachelor's degree	in person
13	USA	31	F	Researcher	PhD	online
14	USA	32	M	PhD student	Bachelor's degree	online
15	USA	24	F	College lecturer	Bachelor's degree	online
16	USA	29	F	Tech Sales representative	Bachelor's degree	online
17	Malaysia	28	M	Software engineer	Bachelor's degree	online
18	USA	31	F	Consultant	Master's degree	online
19	Canada	31	M	Personal Trainer	College degree	online
20	USA	40	M	Union Business representative	College degree	online
21	UK	32	F	Marketing Consultant	Bachelor's degree	online
22	USA	32	F	English teacher	Master's degree	in person
23	China	27	F	English teacher	Master's degree	in person
24	USA	22	F	College lecturer	Bachelor of Arts	online
25	USA	29	M	Graphic Designer	Master	online
26	USA	21	F	Student	Associate degree	in person
27	USA	18	F	Student	High school level	in person
28	USA	20	F	Bartender	High school level	in person

Respondent	Country of birth	Age	Gender	Job/occupation	Diploma	Online or in person
29	USA	19	F	Bartender	High school level	in person
30	USA	20	F	Student	High school level	in person
31	UK	23	F	Student	A-Levels	in person
32	USA	21	F	Student / Server at a bar	High school level	in person

Table 42: Profiles of English-speaking participants to our experiment

Our French-speaking respondents were aged from 18 to 41 years old – the overwhelming majority, i.e., 31 respondents (96.9%) were in their twenties or younger. Twenty-six (81.3%) were women, the rest were men. Thirty (93.8%) were born in France, one was born in Algeria and one was born in England, but all were native speakers of French from France. Twenty-six (81.3%) were students. Their diplomas ranged from French middle school level (*Brevet des collèges*) to Master's degree.

Number	Country of birth	Age	Gender	Job/occupation	Diploma	Online or in person
1	UK	25	M	IT engineer	Brevet des collèges	in person
2	France	25	F	Educational designer	Master's degree	in person
3	France	20	M	Student	Baccalauréat	online
4	France	25	F	Teacher	Master's degree	online
5	France	26	F	Unemployed	Master's degree	online
6	France	20	F	Student	Baccalauréat	online
7	France	22	M	Student	Baccalauréat	online
8	France	22	F	Student	Licence	online
9	France	22	M	Student	Licence	online
10	France	41	F	Teacher	Master's degree	online

Number	Country of birth	Age	Gender	Job/occupation	Diploma	Online or in person
11	France	19	F	Student	Baccalauréat	online
12	France	21	F	Student	Licence	online
13	France	19	F	Student	Baccalauréat	online
14	France	18	F	Student	Baccalauréat	in person
15	France	26	F	Student	BTS	in person
16	France	19	F	Student	Baccalauréat	in person
17	France	18	F	Student	Baccalauréat	online
18	France	18	F	Student	Baccalauréat	online
19	France	18	F	Student	Baccalauréat	online
20	France	18	F	Student	Baccalauréat	online
21	France	18	F	Student	Baccalauréat	online
22	France	18	M	Student	Baccalauréat	online
23	France	21	F	Teacher	Licence	in person
24	France	21	M	Student	Licence	in person
25	France	21	F	Student	Licence	in person
26	France	22	F	Student	Licence	in person
27	France	20	F	Student	Baccalauréat	in person
28	France	24	F	Student	Licence	in person
29	France	21	F	Student	Baccalauréat	in person
30	France	23	F	Student	Licence	in person
31	Algeria	19	F	Student	Baccalauréat	in person
32	France	22	F	Student	Baccalauréat	in person

Table 43: Profiles of French-speaking participants to our experiment

3.3.1.3 Transcription and inter-rater variability testing

The invented words were transcribed into IPA by ear based on the audio recording, but also based on the written answers.

When respondents wrote their invented words, we assumed that they would follow the usual spelling rules of English and of French, and by all appearances they did, although somewhat liberally in a few cases so it required a small amount of personal interpretation – examples of such cases are explained below. An inter-rater test confirmed that there was only a limited amount of potential bias – we explain the methodology for this inter-rater test further below.

We decided to listen to the recording and read the written answers at the same time. This limited the number of possible interpretations right away, allowing us to transcribe more practically and efficiently. For example, in a recording it can be difficult to distinguish by ear between the French voiceless postalveolar fricative /ʃ/ and its voiced equivalent /ʒ/. However, the first can only be spelled *ch*, *sch*, or *sh* according to the spelling rules of French, while the second can only be spelled *j* or *g*. When a French-speaking respondent invented a word that could be interpreted by ear as /fʁuʒib/ or /fʁuʃib/, and she spelled it as *froujib*, the most likely interpretation is that the word was [fʁuʒib], following the usual grapheme-phoneme correspondence rules of French.

In a few cases, the audio and the written answer contradicted themselves. For those, it was decided that the audio was more reliable. Respondents were explicitly instructed to first say the word, then write it, so the audio recording should be more representative of the first spontaneous answer compared to the written answer. It is also possible that respondents did not pay

too much attention to how they wrote, or made typos. For example, a French-speaking respondent invented a word that she pronounced [ʃpumf]. However, she transcribed it as *Spoumf*. Based on this written answer and the spelling rules of French, one could argue that the first consonant she pronounced was not /ʃ/ but /s/. Since the audio is more reliable, we decided that the correct IPA transcription is /ʃpumf/.

Our transcriptions are phonemic transcriptions rather than phonetic, i.e., it is a more theoretical transcription, with standard sounds of English or French, not taking into account idiosyncratic details of the respondents' pronunciation during the experiment, details which would be irrelevant to our study. This makes transcription easier but also allows comparisons with existing words of the English and French lexicon – for which we use phonemic transcriptions – and comparisons between invented words themselves.

We also assumed that when inventing alien words, respondents meant to use existing phonemes of their own native language. By all appearances, they did – so our choice for phonemic transcription is even more justified. One could imagine, e.g., a native English speaker who learned how to pronounce the sound [ɲ] (as in Spanish *niña* "girl" pronounced /ˈniɲa/) in a language class at school, and would use it in their answers to our questionnaire: this sort of cross-linguistic creativity almost never happened, at least based on our interpretation of the audio and the written answers. Here are the only three exceptions, which have been kept in the data as they correspond to the sounds produced by the respondents:

- two French speakers proposed /**h**ajli/ and /**h**ejo/ as alien words for greeting each other: /h/ is not a phoneme of French, and they probably

borrowed it from English. The indirect inspiration for those two words is probably the English word *hello*, or Swedish *hej*.

- an English speaker proposed *Gaston* → /ˈgæst**ɔ̃**/ with the typically French nasal vowel /ɔ̃/, as an alien word for leaving each other. Prompted by the instructor, she confirmed explicitly that she was borrowing the French word *Gaston* → /gastɔ̃/.

In total, our experiment yielded 511 alien words.[20] Those three examples of cross-linguistic use of phonemes amount to only 0.6%. Inventing completely new words is a task already quite difficult in itself. Even if it was easier, participating in an experiment in your native language is a strong incentive to focus on the phonemes of that language and not on others you might know. For these reasons, it is no surprise that respondents have used overwhelmingly the phonemes of their native language.

This also makes the transcription process more coherent and less subject to personal interpretation. For example, a French-speaking respondent invented a word that we clearly hear as a monosyllabic [ʁaːv] – with a long [a] – and she wrote it as *raav*. It is possible, likely even, that the respondent meant the *a* vowel to be pronounced long in the alien language, even if /aː/ is not a phoneme of French, since vowel length is not distinctive in French. It is even more likely that she used the typically non-French spelling *aa* to express that intention. However, because we want to avoid subjective interpretation as much as possible, we adopted a more conservative solution and transcribed it as /a/ (/ʁav/) like all other [a]-like sounds we heard in the answers from French

20 Thirty-two English speakers and the same number French speakers invented 8 words each, i.e., 512 words minus one because a respondent skipped a question, escaping the instructor's notice.

speakers.

Question 8 asked respondents to invent two words, but they were allowed to invent only one. Some invented two words, but many invented only one word and then went on to the next question. When respondents invented two words, we included in the data only the first one and removed the second one. That way the data are more coherent and comparable: they only include the first word that came to the respondent's mind when answering the question.

In order to determine whether our IPA annotation for invented words was not subjectively biased, we conducted an inter-rater variability test. For both languages, we solicited a fellow linguistics researcher who was a native speaker of either English or French, and familiar with IPA annotation. The rater was presented with a series of 26 randomly selected words, i.e., 10% of our data, comprising 255 words for English and 256 words for French. In accordance with the methodological choices described above, the rater was asked to give a phonemic rather than phonetic transcription of the invented word, using only the phonemes of English or French. In order to make the rater's transcription less susceptible to personal bias, they were not presented with the question that prompted each word – so they did not know if they were transcribing a taboo or non-taboo invented word. Instead, they were presented only with the audio recording of the spoken word and the same word in regular alphabet written by the respondent.

For English, there was no disagreement between our annotation and our rater's with regards to the number of unsonorous consonants and the number of phonemes, i.e., the two parameters relevant to our research question. For French, that variation was 2.9%: there were 2 disagreements about the

number of unsonorous consonants and 2 disagreements about the number of phonemes, which makes 4 disagreements out of 135 annotated phonemes. This confirms that our IPA annotation of the invented words is minimally subjective and does not bias the results.[21] The full list of 255 experimental words invented by native speakers of English is given below in Table 44. The full list of 256 experimental words invented by native speakers of French is given below in Table 45.

Respondent (English)	Meaning and swear word status (SW)		Written answer	IPA
1	GREETING EACH OTHER	-	Haj	/ˈhædʒ/
1	LEAVING EACH OTHER	-	Teepo	/ˈtiːpəʊ/
1	REQUESTING POLITELY	-	Tarmeng	/ˈtɑːmeŋ/
1	EXPRESSING GRATITUDE	-	Tarm	/ˈtɑːm/
1	SWEAR WORD when ANGRY	SW	Marp	/ˈmɑːp/
1	SWEAR WORD when SURPRISED	SW	Op!	/ˈɒp/
1	SWEAR WORD and INSULT	SW	Tark	/ˈtɑːk/
1	INSULT for OTHER ALIEN GROUPS	SW	Porf	/ˈpɔːf/
2	GREETING EACH OTHER	-	pi	/ˈpaɪ/
2	LEAVING EACH OTHER	-	tieo	/ˈtaɪɒ/
2	REQUESTING POLITELY	-	kai	/ˈkaɪ/
2	EXPRESSING GRATITUDE	-	canyo	/ˈkænjɒ/
2	SWEAR WORD when ANGRY	SW	flitz	/ˈflɪts/

21 The number of inter-rater disagreements rises to 22% (English) and 5.2% (French) if we look at disagreements about any phoneme, including more sonorous consonants and vowels. The English-speaking rater disagreed with our annotation about 23 vowels and 3 consonants, i.e., 26 out of 118 phonemes (22%). The French-speaking rater disagreed with our annotation about 3 vowels and 4 consonants, i.e., 7 out of 135 phonemes (5.2%). As this is not relevant to our research question, this variation can be ignored.

Respondent (English)	Meaning and swear word status (SW)		Written answer	IPA
2	SWEAR WORD when SURPRISED	SW	pourl	/ˈpɔːl/
2	SWEAR WORD and INSULT	SW	anyon	/ˈænjɒn/
2	INSULT for OTHER ALIEN GROUPS	SW	fallop	/ˈfɔːlɒp/
3	GREETING EACH OTHER	-	Blarp	/ˈblɑːp/
3	LEAVING EACH OTHER	-	Dosta	/ˈdɒstə/
3	REQUESTING POLITELY	-	Zdra	/ˈzdrɑː/
3	EXPRESSING GRATITUDE	-	Baank	/ˈbɑːŋk/
3	SWEAR WORD when ANGRY	SW	Draank	/ˈdrɑːŋk/
3	SWEAR WORD when SURPRISED	SW	Jart	/ˈdʒɑːt/
3	SWEAR WORD and INSULT	SW	Tornk	/ˈtɔːŋk/
3	INSULT for OTHER ALIEN GROUPS	SW	Sprank	/ˈspræŋk/
4	GREETING EACH OTHER	-	Tara	/təˈrɑː/
4	LEAVING EACH OTHER	-	Zoobaiy	/ˈzuːbaɪ/
4	REQUESTING POLITELY	-	Zizz	/ˈzɪz/
4	EXPRESSING GRATITUDE	-	Cheego	/ˈtʃiːgəʊ/
4	SWEAR WORD when ANGRY	SW	Ninn	/ˈnɪn/
4	SWEAR WORD when SURPRISED	SW	BAP	/ˈbæp/
4	SWEAR WORD and INSULT	SW	Cov	/ˈkɒv/
4	INSULT for OTHER ALIEN GROUPS	SW	Bash	/ˈbæʃ/
5	GREETING EACH OTHER	-	Hetz	/ˈhets/
5	LEAVING EACH OTHER	-	Ooglie	/ˈəʊgliː/
5	REQUESTING POLITELY	-	Squeegle	/ˈskwiːgəl/
5	EXPRESSING GRATITUDE	-	Poigle	/ˈpɔɪgəl/
5	SWEAR WORD when ANGRY	SW	Wapzer	/ˈwɒpsə/
5	SWEAR WORD when SURPRISED	SW	Boggle	/ˈbɒgəl/

Respondent (English)	Meaning and swear word status (SW)		Written answer	IPA
5	SWEAR WORD and INSULT	SW	Futze	/ˈfʌtsi/
5	INSULT for OTHER ALIEN GROUPS	SW	Ruffian	/ˈrʌfiən/
6	GREETING EACH OTHER	-	loao	/lɒˈaʊ/
6	LEAVING EACH OTHER	-	da	/ˈdɑː/
6	REQUESTING POLITELY	-	nacar	/næˈkɑː/
6	EXPRESSING GRATITUDE	-	Assi	/æˈsiː/
6	SWEAR WORD when ANGRY	SW	zakar	/zæˈkɑː/
6	SWEAR WORD when SURPRISED	SW	ack	/ˈæk/
6	SWEAR WORD and INSULT	SW	Blut	/ˈblʌt/
6	INSULT for OTHER ALIEN GROUPS	SW	zoboots	/zɒˈbuːts/
7	GREETING EACH OTHER	-	Jasmorg	/dʒɑːˈʃmɔːg/
7	LEAVING EACH OTHER	-	Jabye	/dʒɑːˈbaɪ/
7	REQUESTING POLITELY	-	Jagreetful	/dʒɑːˈgriːtfəl/
7	EXPRESSING GRATITUDE	-	Jafull	/dʒɑːˈfuːl/
7	SWEAR WORD when ANGRY	SW	Jacooty	/dʒɑːˈkuːti/
7	SWEAR WORD when SURPRISED	SW	Jaboul	/dʒɑːˈbuːl/
7	SWEAR WORD and INSULT	SW	Jasuh	/dʒɑːˈsʌh/
7	INSULT for OTHER ALIEN GROUPS	SW	Jaspeal	/dʒɑːˈspiːl/
8	GREETING EACH OTHER	-	Alia	/ˈæliə/
8	LEAVING EACH OTHER	-	Salutane	/ˌsæluˈteɪn/
8	REQUESTING POLITELY	-	Sweetie	/ˈswiːti/
8	EXPRESSING GRATITUDE	-	Giftus	/ˈgɪftəs/
8	SWEAR WORD when ANGRY	SW	Boggler	/ˈbɒglə/
8	SWEAR WORD when SURPRISED	SW	Fudging	/ˈfʊdʒɪŋ/
8	SWEAR WORD and INSULT	SW	Crataneous	/krəˈteɪniəs/

Respondent (English)	Meaning and swear word status (SW)		Written answer	IPA
8	INSULT for OTHER ALIEN GROUPS	SW	Bog	/ˈbɒg/
9	GREETING EACH OTHER	-	scubudu	/ˈʃkuːbəduː/
9	LEAVING EACH OTHER	-	schoochdidouch	/ˈʃkuːʃkɪduːʃ/
9	REQUESTING POLITELY	-	Chnubudu	/ˈʃnuːbəduː/
9	EXPRESSING GRATITUDE	-	flaffy	/ˈflæfi/
9	SWEAR WORD when ANGRY	SW	skamoosh	/ʃkæˈmuːʃ/
9	SWEAR WORD when SURPRISED	SW	zbaltty	/ˈzblæti/
9	SWEAR WORD and INSULT	SW	zblingadorous	/ˌzblɪŋgəˈdɔːrəs/
9	INSULT for OTHER ALIEN GROUPS	SW	pfofqdodoado	/ˌpfɒfəˈduːdəduː/
10	GREETING EACH OTHER	-	Jodol	/dʒəˈdɒl/
10	LEAVING EACH OTHER	-	Modol	/ˈmɒdəl/
10	REQUESTING POLITELY	-	Jimble	/ˈdʒɪmbəl/
10	EXPRESSING GRATITUDE	-	Mombil	/ˈmɒmbɪl/
10	SWEAR WORD when ANGRY	SW	Schkah	/ˈʃkɑː/
10	SWEAR WORD when SURPRISED	SW	Hukadi	/hʊkəˈdiː/
10	SWEAR WORD and INSULT	SW	Pimrie	/ˈpɪmriː/
10	INSULT for OTHER ALIEN GROUPS	SW	Djakal	/dʒæˈkæl/
11	GREETING EACH OTHER	-	vana	/ˈvɑːnə/
11	LEAVING EACH OTHER	-	leeloo	/ˈliːluː/
11	REQUESTING POLITELY	-	tudava	/ˌtʊdəˈvɑː/
11	EXPRESSING GRATITUDE	-	flalon	/flæˈlɒn/
11	SWEAR WORD when ANGRY	SW	faku	/ˈfæku/
11	SWEAR WORD when SURPRISED	SW	hupo	/ˈhʌpəʊ/
11	SWEAR WORD and INSULT	SW	zazoo	/ˈzɑːzuː/
11	INSULT for OTHER ALIEN GROUPS	SW	pudoo	/puˈduː/

Respondent (English)	Meaning and swear word status (SW)		Written answer	IPA
12	GREETING EACH OTHER	-	Flarg	/ˈflɑːg/
12	LEAVING EACH OTHER	-	Ploond	/ˈpluːnd/
12	REQUESTING POLITELY	-	Plord	/ˈplɔːd/
12	EXPRESSING GRATITUDE	-	Schmeen	/ˈʃmiːn/
12	SWEAR WORD when ANGRY	SW	Bwaah	/ˈbwɑː/
12	SWEAR WORD when SURPRISED	SW	Floop	/ˈfluːp/
12	SWEAR WORD and INSULT	SW	Stub	/ˈstʌb/
12	INSULT for OTHER ALIEN GROUPS	SW	Schoombo	/ˈskuːmbəʊ/
13	GREETING EACH OTHER	-	Velka	/ˈvelkə/
13	LEAVING EACH OTHER	-	Sayo	/ˈseɪɒ/
13	REQUESTING POLITELY	-	Hegee	/ˈhegiː/
13	EXPRESSING GRATITUDE	-	Hav hav	/ˈhævhæv/
13	SWEAR WORD when ANGRY	SW	Aflaat	/ˈæflɑːt/
13	SWEAR WORD when SURPRISED	SW	Ahchoo	/æˈtʃuː/
13	SWEAR WORD and INSULT	SW	Hellu	/heˈlʊ/
13	INSULT for OTHER ALIEN GROUPS	SW	Lamouf	/ˈlæmʌf/
14	GREETING EACH OTHER	-	Pleatz	/ˈpliːts/
14	LEAVING EACH OTHER	-	Dwippp	/ˈdwɪp/
14	REQUESTING POLITELY	-	Blooo	/ˈbluː/
14	EXPRESSING GRATITUDE	-	Hruuuuuu	/ˈhruː/
14	SWEAR WORD when ANGRY	SW	Kah	/ˈkɑː/
14	SWEAR WORD when SURPRISED	SW	Blah	/ˈblɑː/
14	SWEAR WORD and INSULT	SW	Djou	/ˈdʒuː/
14	INSULT for OTHER ALIEN GROUPS	SW	Jah	/ˈʒɑː/
15	GREETING EACH OTHER	-	Olaki	/ˈɒləkaɪ/

Respondent (English)	Meaning and swear word status (SW)		Written answer	IPA
15	LEAVING EACH OTHER	-	Yahowa	/jæˈhaʊə/
15	REQUESTING POLITELY	-	Pogawa	/pɒˈgæwə/
15	EXPRESSING GRATITUDE	-	Angua	/ˈæŋgwə/
15	SWEAR WORD when ANGRY	SW	Frip	/ˈfrɪp/
15	SWEAR WORD when SURPRISED	SW	Joop	/ˈdʒuːp/
15	SWEAR WORD and INSULT	SW	Burfip	/ˈbɜːfɪp/
15	INSULT for OTHER ALIEN GROUPS	SW	Bratine	/ˈbreɪtən/
16	GREETING EACH OTHER	-	Nalia	/ˈnɑːliə/
16	LEAVING EACH OTHER	-	Chulia	/tʃuˈliːə/
16	REQUESTING POLITELY	-	xinguma	/zɪŋˈguːmɑː/
16	EXPRESSING GRATITUDE	-	Hikumie	/ˈhaɪkʊmiː/
16	SWEAR WORD when ANGRY	SW	Kaminga	/kəˈmɪŋgə/
16	SWEAR WORD when SURPRISED	SW	Mcgorebles	/mækˈgɔːbəlz/
16	SWEAR WORD and INSULT	SW	Werchen	/ˈwɜːtʃən/
16	INSULT for OTHER ALIEN GROUPS	SW	Zingles	/ˈzɪŋgəlz/
17	GREETING EACH OTHER	-	harouphm	/hɑːˈruːmf/
17	LEAVING EACH OTHER	-	Sa	/ˈsɑː/
17	REQUESTING POLITELY	-	systoum	/ˈsɪstuːm/
17	EXPRESSING GRATITUDE	-	kaala	/ˈkɑːlɑː/
17	SWEAR WORD when ANGRY	SW	rashet	/ˈrɑːʃet/
17	SWEAR WORD when SURPRISED	SW	kiish	/ˈkiːʃ/
17	SWEAR WORD and INSULT	SW	krakout	/krɑːˈkuːt/
17	INSULT for OTHER ALIEN GROUPS	SW	barbar	/ˈbɑːbɑː/
18	GREETING EACH OTHER	-	how	/ˈhaʊ/
18	LEAVING EACH OTHER	-	bahku	/ˈbɑːkuː/

Respondent (English)	Meaning and swear word status (SW)		Written answer	IPA
18	REQUESTING POLITELY	-	gow	/ˈgaʊ/
18	EXPRESSING GRATITUDE	-	jink	/ˈdʒɪŋk/
18	SWEAR WORD when ANGRY	SW	zoo	/ˈzuː/
18	SWEAR WORD when SURPRISED	SW	zouch	/ˈzaʊtʃ/
18	SWEAR WORD and INSULT	SW	zugu	/ˈzuːguː/
18	INSULT for OTHER ALIEN GROUPS	SW	swo	/ˈswəʊ/
19	GREETING EACH OTHER	-	Bish	/ˈbɪʃ/
19	LEAVING EACH OTHER	-	Bika	/ˈbɪkɑː/
19	REQUESTING POLITELY	-	Mooka	/ˈmuːkɑː/
19	EXPRESSING GRATITUDE	-	Hallo	/ˈhɑːləʊ/
19	SWEAR WORD when ANGRY	SW	HARKAR	/ˈhɑːkɑː/
19	SWEAR WORD when SURPRISED	SW	Globo	/ˈgləʊbəʊ/
19	SWEAR WORD and INSULT	SW	Jeema	/ˈdʒiːmɑː/
19	INSULT for OTHER ALIEN GROUPS	SW	Karot	/ˈkeərɒt/
20	GREETING EACH OTHER	-	Oy	/ˈɔj/
20	LEAVING EACH OTHER	-	Garff	/ˈgɑːf/
20	REQUESTING POLITELY	-	Targoul	/ˈtɑːguːl/
20	EXPRESSING GRATITUDE	-	Tits	/ˈtɪts/
20	SWEAR WORD when ANGRY	SW	Vermit	/ˈvɜːmɪt/
20	SWEAR WORD when SURPRISED	SW	Shepegal	/ˈʃpiːgəl/
20	SWEAR WORD and INSULT	SW	Shperlack	/ˈʃpɜːlæk/
20	INSULT for OTHER ALIEN GROUPS	SW	Sploof	/ˈspluːf/
21	GREETING EACH OTHER	-	Blob	/ˈblɒb/
21	LEAVING EACH OTHER	-	Partak	/ˈpɑːtæk/
21	REQUESTING POLITELY	-	Brockle	/ˈbrɒkəl/

Respondent (English)	Meaning and swear word status (SW)		Written answer	IPA
21	EXPRESSING GRATITUDE	-	Merkani	/ˈmɜːkɑːni/
21	SWEAR WORD when ANGRY	SW	Shatok	/ˈʃætɒk/
21	SWEAR WORD when SURPRISED	SW	Wotok	/ˈwɒtɒk/
21	SWEAR WORD and INSULT	SW	Tropple	/ˈtrɒpəl/
21	INSULT for OTHER ALIEN GROUPS	SW	Florben	/ˈflɔːbən/
22	GREETING EACH OTHER	-	haho	/ˈhæhəʊ/
22	LEAVING EACH OTHER	-	blew	/ˈbluː/
22	REQUESTING POLITELY	-	spool	/ˈspuːl/
22	EXPRESSING GRATITUDE	-	quermit	/ˈkɜːmɪt/
22	SWEAR WORD when ANGRY	SW	phunk	/ˈfʌŋk/
22	SWEAR WORD when SURPRISED	SW	flonk	/ˈflɒŋk/
22	SWEAR WORD and INSULT	SW	MERFIT	/ˈmɜːfɪt/
22	INSULT for OTHER ALIEN GROUPS	SW	flerk	/ˈflɜːk/
23	GREETING EACH OTHER	-	He Heyo	/heˈhejəʊ/
23	LEAVING EACH OTHER	-	Oubli	/ˈuːbli/
23	REQUESTING POLITELY	-	I ken ha dee	/ˌɪ ken hæ ˈdiː/
23	EXPRESSING GRATITUDE	-	Merks	/ˈmɜːks/
23	SWEAR WORD when ANGRY	SW	Bah	/ˈbɑː/
23	SWEAR WORD when SURPRISED	SW	Dah	/ˈdɑː/
23	SWEAR WORD and INSULT	SW	Flaggibit	/ˈflægɪbɪt/
23	INSULT for OTHER ALIEN GROUPS	SW	Yurken glab splee	/ˌjɜːkən ˈglæb spliː/
24	GREETING EACH OTHER	-	alora	/əˈlɔːrə/
24	LEAVING EACH OTHER	-	praxis	/ˈpræksɪs/
24	REQUESTING POLITELY	-	meech	/ˈmiːtʃ/
24	EXPRESSING GRATITUDE	-	martu	/ˈmɑːtu/

Respondent (English)	Meaning and swear word status (SW)		Written answer	IPA
24	SWEAR WORD when ANGRY	SW	grackle	/ˈgrækəl/
24	SWEAR WORD when SURPRISED	SW	womp	/ˈwɒmp/
24	SWEAR WORD and INSULT	SW	gonk	/ˈgɒŋk/
24	INSULT for OTHER ALIEN GROUPS	SW	flarbo	/ˈflɑːbəʊ/
25	GREETING EACH OTHER	-	Bleezarp	/ˈbliːzɑːp/
25	LEAVING EACH OTHER	-	See-Sar	/ˈsiːsɑː/
25	REQUESTING POLITELY	-	Pachoom	/ˈpætʃʊm/
25	EXPRESSING GRATITUDE	-	Bleeker	/ˈbliːkə/
25	SWEAR WORD when ANGRY	SW	Nahr	/ˈnɑː/
25	SWEAR WORD when SURPRISED	SW	Blachoom	/blæˈtʃuːm/
25	SWEAR WORD and INSULT	SW	Klank	/ˈklæŋk/
25	INSULT for OTHER ALIEN GROUPS	SW	Feebark	/ˈfiːbɑːk/
26	GREETING EACH OTHER	-	biskof	/ˈbɪskɒf/
26	LEAVING EACH OTHER	-	gaston	/ˈgæstɔ̃/
26	REQUESTING POLITELY	-	tak	/ˈtɑːk/
26	EXPRESSING GRATITUDE	-	AHSSI	/ˈɑːsi/
26	SWEAR WORD when ANGRY	SW	DACK	/ˈdæk/
26	SWEAR WORD when SURPRISED	SW	scwhat	/ˈʃwæt/
26	SWEAR WORD and INSULT	SW	Swacker	/ˈswækə/
26	INSULT for OTHER ALIEN GROUPS	SW	Radahms	/ˈrɑːdɑːms/
27	GREETING EACH OTHER	-	augrek	/ˈɔːgrek/
27	LEAVING EACH OTHER	-	saumen	/ˈsɑːmən/
27	REQUESTING POLITELY	-	grieveu	/ˈgriːvuː/
27	EXPRESSING GRATITUDE	-	brogen	/ˈbrɒgən/
27	SWEAR WORD when ANGRY	SW	bruck	/ˈbrʌk/

Respondent (English)	Meaning and swear word status (SW)		Written answer	IPA
27	SWEAR WORD when SURPRISED	SW	[question skipped]	[question skipped]
27	SWEAR WORD and INSULT	SW	klug	/ˈklʌg/
27	INSULT for OTHER ALIEN GROUPS	SW	stchomper	/ˈʃtɒmpə/
28	GREETING EACH OTHER	-	zaja	/ˈzædʒə/
28	LEAVING EACH OTHER	-	miffda	/ˈmɪfdə/
28	REQUESTING POLITELY	-	veddy	/ˈvedi/
28	EXPRESSING GRATITUDE	-	miff	/ˈmɪf/
28	SWEAR WORD when ANGRY	SW	zara	/ˈzærə/
28	SWEAR WORD when SURPRISED	SW	waka	/ˈwækə/
28	SWEAR WORD and INSULT	SW	bragard	/ˈbrægɑːd/
28	INSULT for OTHER ALIEN GROUPS	SW	boya	/ˈbɒjə/
29	GREETING EACH OTHER	-	Oop	/ˈuːp/
29	LEAVING EACH OTHER	-	Blek Black	/ˈblekblæk/
29	REQUESTING POLITELY	-	Astro	/ˈæstrəʊ/
29	EXPRESSING GRATITUDE	-	Blip	/ˈblɪp/
29	SWEAR WORD when ANGRY	SW	Merk	/ˈmɜːk/
29	SWEAR WORD when SURPRISED	SW	ooga booga	/ˌuːgəˈbuːgə/
29	SWEAR WORD and INSULT	SW	Snoob	/ˈsnuːb/
29	INSULT for OTHER ALIEN GROUPS	SW	Snoogas	/ˈsnuːgəz/
30	GREETING EACH OTHER	-	haba	/ˈhɑːbə/
30	LEAVING EACH OTHER	-	zabook	/ˈzæbʊk/
30	REQUESTING POLITELY	-	na	/ˈnɑː/
30	EXPRESSING GRATITUDE	-	baba	/ˈbɑːbə/
30	SWEAR WORD when ANGRY	SW	kra	/ˈkrɑː/
30	SWEAR WORD when SURPRISED	SW	oka	/ˈəʊkə/

Respondent (English)	Meaning and swear word status (SW)		Written answer	IPA
30	SWEAR WORD and INSULT	SW	kazka	/ˈkæzkə/
30	INSULT for OTHER ALIEN GROUPS	SW	puzhka	/ˈpʊzkə/
31	GREETING EACH OTHER	-	Imfloo	/ˈɪmfluː/
31	LEAVING EACH OTHER	-	bomkee	/ˈbɒmke/
31	REQUESTING POLITELY	-	mlaise	/ˈmlɛɪʒ/
31	EXPRESSING GRATITUDE	-	forsee	/ˈfɔːsiː/
31	SWEAR WORD when ANGRY	SW	frunack	/ˈfrʊnæk/
31	SWEAR WORD when SURPRISED	SW	amraf	/ˈæmræf/
31	SWEAR WORD and INSULT	SW	plunck	/ˈplʌŋk/
31	INSULT for OTHER ALIEN GROUPS	SW	vu nicte dine	/vuː ˈnɪktə dɑɪn/
32	GREETING EACH OTHER	-	chiao	/ˈtʃɑʊ/
32	LEAVING EACH OTHER	-	abit	/ˈæbɪt/
32	REQUESTING POLITELY	-	dingin	/ˈdɪŋgɪn/
32	EXPRESSING GRATITUDE	-	melow	/ˈmiːləʊ/
32	SWEAR WORD when ANGRY	SW	hap	/ˈhæp/
32	SWEAR WORD when SURPRISED	SW	how	/ˈhɑʊ/
32	SWEAR WORD and INSULT	SW	zamin	/ˈzæmɪn/
32	INSULT for OTHER ALIEN GROUPS	SW	lamins	/ˈlæmɪns/

Table 44: Experimental words invented by English-speaking participants

Respondent (French)	Meaning and swear word status (SW)		Written answer	IPA
1	LEAVING EACH OTHER	-	gaoubal	/gaubal/
1	REQUESTING POLITELY	-	malgh	/malg/
1	GREETING EACH OTHER	-	zouly	/zuli/

1	EXPRESSING GRATITUDE	-	mouly	/muli/
1	SWEAR WORD when SURPRISED	SW	sllop	/slop/
1	INSULT for OTHER ALIEN GROUPS	SW	goulbe	/gulb/
1	SWEAR WORD and INSULT	SW	zergh	/zœʁg/
1	SWEAR WORD when ANGRY	SW	zarrr	/zaʁ/
2	GREETING EACH OTHER	-	pilii	/pili/
2	EXPRESSING GRATITUDE	-	bolavi	/bolavi/
2	REQUESTING POLITELY	-	vlavibu	/vlaviby/
2	LEAVING EACH OTHER	-	raav	/ʁav/
2	SWEAR WORD when ANGRY	SW	krak	/kʁak/
2	SWEAR WORD when SURPRISED	SW	rabroc	/zabrok/
2	SWEAR WORD and INSULT	SW	gruvk	/gʁuvək/
2	INSULT for OTHER ALIEN GROUPS	SW	mvr	/mʁəvə/
3	REQUESTING POLITELY	-	Bushtuk !	/byʃtyk/
3	GREETING EACH OTHER	-	Zaklop !	/zaklɔp/
3	LEAVING EACH OTHER	-	Zazaklop !	/zazaklɔp/
3	EXPRESSING GRATITUDE	-	Zlaba	/zlaba/
3	INSULT for OTHER ALIEN GROUPS	SW	Schgronk !	/ʃgʁɔ̃k/
3	SWEAR WORD when SURPRISED	SW	Schongronk ?	/ʃɔ̃gʁɔ̃k/
3	SWEAR WORD and INSULT	SW	Prachta !	/pʁaʃta/
3	SWEAR WORD when ANGRY	SW	Greuncheunn	/gʁœnʃən/
4	EXPRESSING GRATITUDE	-	Tiki	/tiki/
4	REQUESTING POLITELY	-	Spizy	/ʃpizi/
4	GREETING EACH OTHER	-	Wafa	/wafa/
4	LEAVING EACH OTHER	-	Boumi	/bumi/
4	SWEAR WORD when SURPRISED	SW	Spoumf	/ʃpumf/

4	INSULT for OTHER ALIEN GROUPS	SW	Craquibak	/kʁakibak/
4	SWEAR WORD when ANGRY	SW	Grapoum	/gʁapum/
4	SWEAR WORD and INSULT	SW	Froujib	/fʁuʒib/
5	LEAVING EACH OTHER	-	fiote	/fjot/
5	EXPRESSING GRATITUDE	-	draki	/dʁaki/
5	REQUESTING POLITELY	-	mikou	/miku/
5	GREETING EACH OTHER	-	houlou	/ulu/
5	SWEAR WORD when ANGRY	SW	gloethe	/gløt/
5	SWEAR WORD and INSULT	SW	chtram	/ʃtʁam/
5	INSULT for OTHER ALIEN GROUPS	SW	zitak	/zitak/
5	SWEAR WORD when SURPRISED	SW	klam	/klam/
6	GREETING EACH OTHER	-	tchoo	/tʃo/
6	LEAVING EACH OTHER	-	plak	/plak/
6	REQUESTING POLITELY	-	kii	/ki/
6	EXPRESSING GRATITUDE	-	kaduk	/kadyk/
6	SWEAR WORD when SURPRISED	SW	kluk	/kluk/
6	SWEAR WORD and INSULT	SW	fiak	/fjak/
6	SWEAR WORD when ANGRY	SW	chniouk	/ʃnjuk/
6	INSULT for OTHER ALIEN GROUPS	SW	palzak	/palzak/
7	REQUESTING POLITELY	-	chuyba	/ʃɥiba/
7	EXPRESSING GRATITUDE	-	tablou	/tablu/
7	LEAVING EACH OTHER	-	souli	/suli/
7	GREETING EACH OTHER	-	marcia	/maʁsja/
7	INSULT for OTHER ALIEN GROUPS	SW	pastiblo	/pastiblo/
7	SWEAR WORD when SURPRISED	SW	chabal	/ʃabal/
7	SWEAR WORD when ANGRY	SW	taboulasba	/tabulazba/

7	SWEAR WORD and INSULT	SW	salbui	/salbɥi/
8	EXPRESSING GRATITUDE	-	issi	/isi/
8	REQUESTING POLITELY	-	pliz	/pliz/
8	GREETING EACH OTHER	-	yi	/ji/
8	LEAVING EACH OTHER	-	va	/va/
8	SWEAR WORD and INSULT	SW	bi	/bi/
8	INSULT for OTHER ALIEN GROUPS	SW	bor	/bɔʁ/
8	SWEAR WORD when ANGRY	SW	argh	/aʁ/
8	SWEAR WORD when SURPRISED	SW	wo	/wo/
9	LEAVING EACH OTHER	-	Gra	/gʁa/
9	EXPRESSING GRATITUDE	-	Glaï	/glaj/
9	GREETING EACH OTHER	-	Oï	/ɔj/
9	REQUESTING POLITELY	-	Olo	/olo/
9	SWEAR WORD when SURPRISED	SW	Soka	/soka/
9	SWEAR WORD when ANGRY	SW	GRO	/gʁo/
9	SWEAR WORD and INSULT	SW	GLO	/glo/
9	INSULT for OTHER ALIEN GROUPS	SW	KLO	/klo/
10	LEAVING EACH OTHER	-	Tschussi	/tʃusi/
10	REQUESTING POLITELY	-	Soli	/soli/
10	EXPRESSING GRATITUDE	-	Mierdi	/mjɛʁdi/
10	GREETING EACH OTHER	-	Ula	/yla/
10	INSULT for OTHER ALIEN GROUPS	SW	Diked	/diked/
10	SWEAR WORD when ANGRY	SW	Atata	/atata/
10	SWEAR WORD when SURPRISED	SW	Oushi	/uʃi/
10	SWEAR WORD and INSULT	SW	Pouriton	/puʁitɔ̃/
11	EXPRESSING GRATITUDE	-	Zuckblerg	/zykblɛʁg/

11	LEAVING EACH OTHER	-	Ripgu	/ʁipgy/
11	REQUESTING POLITELY	-	Bloubla	/blubla/
11	GREETING EACH OTHER	-	Joujue	/ʒuʒy/
11	INSULT for OTHER ALIEN GROUPS	SW	Tabpie	/tabpi/
11	SWEAR WORD and INSULT	SW	Gruss	/gʁys/
11	SWEAR WORD when ANGRY	SW	Greugah	/gʁøga/
11	SWEAR WORD when SURPRISED	SW	Plar	/plaʁ/
12	GREETING EACH OTHER	-	Borpmip	/bɔʁpmip/
12	LEAVING EACH OTHER	-	Miborp	/mibɔʁp/
12	EXPRESSING GRATITUDE	-	Korvit	/kɔʁvit/
12	REQUESTING POLITELY	-	Azur	/asuʁ/
12	INSULT for OTHER ALIEN GROUPS	SW	Madoud	/madud/
12	SWEAR WORD and INSULT	SW	Marklik	/maʁklik/
12	SWEAR WORD when SURPRISED	SW	Polar	/polaʁ/
12	SWEAR WORD when ANGRY	SW	Azmar	/azmaʁ/
13	LEAVING EACH OTHER	-	gout'gout'	/gutgut/
13	GREETING EACH OTHER	-	Flipouf	/flipuf/
13	REQUESTING POLITELY	-	Pitapou	/pitapu/
13	EXPRESSING GRATITUDE	-	Cilia	/silja/
13	SWEAR WORD when ANGRY	SW	Guabouk!	/gabuk/
13	SWEAR WORD and INSULT	SW	Sabidouk	/sabiduk/
13	INSULT for OTHER ALIEN GROUPS	SW	Zacon, Karik	/zakɔ̃/
13	SWEAR WORD when SURPRISED	SW	Olalou !	/olalu/
14	LEAVING EACH OTHER	-	Saloïd	/saloid/
14	REQUESTING POLITELY	-	pli	/pli/
14	GREETING EACH OTHER	-	bzaru	/bzaʁy/

14	EXPRESSING GRATITUDE	-	Pizra	/pizʁa/
14	SWEAR WORD when SURPRISED	SW	bizgou	/bizgu/
14	SWEAR WORD and INSULT	SW	plaroïd	/plaʁoid/
14	SWEAR WORD when ANGRY	SW	Claru	/klaʁy/
14	INSULT for OTHER ALIEN GROUPS	SW	martrion	/maʁtʁijɔ̃/
15	LEAVING EACH OTHER	-	quitarus	/kwitaʁys/
15	GREETING EACH OTHER	-	saturi	/satyʁi/
15	REQUESTING POLITELY	-	lunatimatus	/lunatimatys/
15	EXPRESSING GRATITUDE	-	lunati	/lunati/
15	SWEAR WORD and INSULT	SW	stropatus	/stʁopatys/
15	INSULT for OTHER ALIEN GROUPS	SW	grosatus	/gʁosatys/
15	SWEAR WORD when SURPRISED	SW	horsu	/ɔʁsy/
15	SWEAR WORD when ANGRY	SW	martalo	/maʁtalo/
16	GREETING EACH OTHER	-	Braxou	/bʁaksu/
16	EXPRESSING GRATITUDE	-	kirkoui	/kirkwi/
16	LEAVING EACH OTHER	-	Adrious	/adʁijus/
16	REQUESTING POLITELY	-	jaoui	/ʒawi/
16	SWEAR WORD and INSULT	SW	Kichtra	/kiʃtʁa/
16	INSULT for OTHER ALIEN GROUPS	SW	Fastrouya	/fastʁuja/
16	SWEAR WORD when SURPRISED	SW	Wouadziou	/wadzju/
16	SWEAR WORD when ANGRY	SW	Ralala	/ʁalala/
17	REQUESTING POLITELY	-	pléteu	/pletø/
17	GREETING EACH OTHER	-	bolo	/bolo/
17	LEAVING EACH OTHER	-	orvou	/ɔʀvu/
17	EXPRESSING GRATITUDE	-	verci	/vɛʁsi/
17	SWEAR WORD when ANGRY	SW	chtir	/ʃtiʁ/

17	SWEAR WORD when SURPRISED	SW	osti	/osti/
17	SWEAR WORD and INSULT	SW	slar	/slaʁ/
17	INSULT for OTHER ALIEN GROUPS	SW	sanar	/sanaʁ/
18	GREETING EACH OTHER	-	Bola	/bola/
18	EXPRESSING GRATITUDE	-	Calie	/kali/
18	REQUESTING POLITELY	-	Polavie	/polavi/
18	LEAVING EACH OTHER	-	Luro	/lyʁo/
18	SWEAR WORD when SURPRISED	SW	Sulafe	/sylaf/
18	SWEAR WORD when ANGRY	SW	Carnode	/kaʁnɔd/
18	SWEAR WORD and INSULT	SW	Doru	/doʁy/
18	INSULT for OTHER ALIEN GROUPS	SW	Romula	/ʁomula/
19	EXPRESSING GRATITUDE	-	damcacha	/damkaʃa/
19	REQUESTING POLITELY	-	pilatou	/pilatu/
19	GREETING EACH OTHER	-	haïli	/hajli/
19	LEAVING EACH OTHER	-	bayleau	/bajlo/
19	SWEAR WORD when ANGRY	SW	bichty	/biʃti/
19	SWEAR WORD and INSULT	SW	mektacha	/mektaʃa/
19	SWEAR WORD when SURPRISED	SW	parati	/paʁati/
19	INSULT for OTHER ALIEN GROUPS	SW	rikatau	/ʁikato/
20	REQUESTING POLITELY	-	Antipastoa	/ɑ̃tipastoa/
20	GREETING EACH OTHER	-	Karztorno	/kaʁztɔʁno/
20	EXPRESSING GRATITUDE	-	Darta	/daʁta/
20	LEAVING EACH OTHER	-	Orztina	/ɔʁztina/
20	SWEAR WORD when ANGRY	SW	Koriba	/koʁiba/
20	SWEAR WORD and INSULT	SW	Karzta	/kaʁzta/
20	INSULT for OTHER ALIEN GROUPS	SW	Gordoni	/gɔʁdoni/

20	SWEAR WORD when SURPRISED	SW	Ohhlaaaa	/ola/
21	EXPRESSING GRATITUDE	-	crate	/kʁat/
21	GREETING EACH OTHER	-	caroche	/kaʁɔʃ/
21	LEAVING EACH OTHER	-	rèche	/ʁɛʃ/
21	REQUESTING POLITELY	-	prage	/pʁaʒ/
21	INSULT for OTHER ALIEN GROUPS	SW	praquer saltige	/pʁake/
21	SWEAR WORD when ANGRY	SW	rique	/ʁik/
21	SWEAR WORD when SURPRISED	SW	rotte	/ʁɔt/
21	SWEAR WORD and INSULT	SW	nache	/naʃ/
22	EXPRESSING GRATITUDE	-	Shrimb	/ʃʁimb/
22	REQUESTING POLITELY	-	Croum ?	/kʁum/
22	LEAVING EACH OTHER	-	Zgrouni !	/zgʁuni/
22	GREETING EACH OTHER	-	houla	/ula/
22	SWEAR WORD when ANGRY	SW	Blangk !	/blɑ̃gk/
22	SWEAR WORD and INSULT	SW	Zglorb !	/zglɔʁb/
22	INSULT for OTHER ALIEN GROUPS	SW	Zomb et Zimb	/zɔ̃b/
22	SWEAR WORD when SURPRISED	SW	Blor !	/blɔʁ/
23	EXPRESSING GRATITUDE	-	plopli	/plopli/
23	GREETING EACH OTHER	-	chavi	/ʃavi/
23	REQUESTING POLITELY	-	rafi	/ʁafi/
23	LEAVING EACH OTHER	-	blouli	/bluli/
23	SWEAR WORD when ANGRY	SW	cracbou	/kʁakbu/
23	SWEAR WORD when SURPRISED	SW	flak	/flak/
23	SWEAR WORD and INSULT	SW	glotak	/glotak/
23	INSULT for OTHER ALIEN GROUPS	SW	rafij	/ʁafiʒ/
24	EXPRESSING GRATITUDE	-	Rog	/ʁɔg/

24	REQUESTING POLITELY	-	Agar	/agaʁ/
24	GREETING EACH OTHER	-	Riise	/ʁiz/
24	LEAVING EACH OTHER	-	Vine	/vin/
24	SWEAR WORD and INSULT	SW	Irsh	/iʁʃ/
24	INSULT for OTHER ALIEN GROUPS	SW	Chaine, tarf	/ʃɛn/
24	SWEAR WORD when ANGRY	SW	Aarz	/aʁz/
24	SWEAR WORD when SURPRISED	SW	Ran	/ʁan/
25	GREETING EACH OTHER	-	Saloupita	/salupita/
25	LEAVING EACH OTHER	-	Bilobu	/biloby/
25	EXPRESSING GRATITUDE	-	Marsamax	/maʁsamaks/
25	REQUESTING POLITELY	-	Paloupalou	/palupalu/
25	SWEAR WORD when ANGRY	SW	Psaposib	/psaposib/
25	INSULT for OTHER ALIEN GROUPS	SW	Badibule	/badibyl/
25	SWEAR WORD and INSULT	SW	Farolax	/faʁolaks/
25	SWEAR WORD when SURPRISED	SW	Patalou	/patalu/
26	GREETING EACH OTHER	-	heyo	/hejo/
26	EXPRESSING GRATITUDE	-	toro	/toʁo/
26	REQUESTING POLITELY	-	prali	/pʁali/
26	LEAVING EACH OTHER	-	auro	/oʁo/
26	INSULT for OTHER ALIEN GROUPS	SW	bapu	/bapy/
26	SWEAR WORD when ANGRY	SW	pracha	/pʁaʃa/
26	SWEAR WORD and INSULT	SW	parto	/paʁto/
26	SWEAR WORD when SURPRISED	SW	fram	/fʁam/
27	EXPRESSING GRATITUDE	-	cimer	/simɛʁ/
27	REQUESTING POLITELY	-	mikala	/mikala/
27	GREETING EACH OTHER	-	vavar	/vavaʁ/

27	LEAVING EACH OTHER	-	vivir	/viviʁ/
27	SWEAR WORD when SURPRISED	SW	bachi	/baʃi/
27	SWEAR WORD and INSULT	SW	derto	/dɛʁto/
27	INSULT for OTHER ALIEN GROUPS	SW	vapoutine	/vaputin/
27	SWEAR WORD when ANGRY	SW	ribouli	/ʁibuli/
28	EXPRESSING GRATITUDE	-	gratitudant	/gʁatitydɑ̃/
28	GREETING EACH OTHER	-	Broua	/bʁua/
28	REQUESTING POLITELY	-	Tiru	/tiʁy/
28	LEAVING EACH OTHER	-	Hayo	/ajo/
28	SWEAR WORD when SURPRISED	SW	PAF	/paf/
28	SWEAR WORD when ANGRY	SW	Barpe	/baʁp/
28	INSULT for OTHER ALIEN GROUPS	SW	Farpo	/faʁpo/
28	SWEAR WORD and INSULT	SW	Florpo	/flɔʁpo/
29	GREETING EACH OTHER	-	blab	/blab/
29	LEAVING EACH OTHER	-	neka	/neka/
29	EXPRESSING GRATITUDE	-	zanka	/zɑ̃ka/
29	REQUESTING POLITELY	-	ziork	/ziɔʁk/
29	SWEAR WORD when ANGRY	SW	breb	/bʁɛb/
29	SWEAR WORD when SURPRISED	SW	gronk	/gʁɔ̃k/
29	SWEAR WORD and INSULT	SW	tika	/tika/
29	INSULT for OTHER ALIEN GROUPS	SW	kabe	/kabe/
30	EXPRESSING GRATITUDE	-	dos	/dos/
30	GREETING EACH OTHER	-	gloubi	/glubi/
30	LEAVING EACH OTHER	-	lai	/laj/
30	REQUESTING POLITELY	-	mi	/mi/
30	SWEAR WORD when SURPRISED	SW	breuz	/bʁøz/

30	INSULT for OTHER ALIEN GROUPS	SW	bozor	/bozɔʁ/
30	SWEAR WORD when ANGRY	SW	zourbi	/zuaʁbi/
30	SWEAR WORD and INSULT	SW	zor	/zɔʁ/
31	EXPRESSING GRATITUDE	-	Flagealable	/flaʒəlabœl/
31	GREETING EACH OTHER	-	Gloski	/gloski/
31	LEAVING EACH OTHER	-	Ikslog	/iksl͡ɔg/
31	REQUESTING POLITELY	-	Coopi	/kupi/
31	SWEAR WORD when SURPRISED	SW	Junda	/dʒunda/
31	INSULT for OTHER ALIEN GROUPS	SW	Bitah	/bita/
31	SWEAR WORD when ANGRY	SW	Pertix	/pɛʁtiks/
31	SWEAR WORD and INSULT	SW	Vermor	/vɛʁmɔʁ/
32	EXPRESSING GRATITUDE	-	zglomb	/zgl͡ɔ̃mb/
32	GREETING EACH OTHER	-	oumdae	/umde/
32	LEAVING EACH OTHER	-	tshienah	/tʃena/
32	REQUESTING POLITELY	-	skol	/skɔl/
32	SWEAR WORD when SURPRISED	SW	oiyshi	/ɔjiʃi/
32	INSULT for OTHER ALIEN GROUPS	SW	stiumpeu	/stjɔ̃mpø/
32	SWEAR WORD when ANGRY	SW	pitsha	/pitʃa/
32	SWEAR WORD and INSULT	SW	ploumtenia	/plumtenja/

Table 45: Experimental words invented by French-speaking participants

3.3.2 Results for Experimental swear words

When comparing experimental words, the results show that just like real-life swear words, the experimental swear words invented by native speakers of

English contain a higher proportion of unsonorous consonants, compared to the invented non-swear words. Significantly more of those swear words have an unsonority density of 0.33 or more, which means that at least one phoneme out of three is one of the least sonorous consonants (see Table 46). The same is true for French (see Table 47). These results are presented visually in Figure 23.

	Experimental RW (English)	**Experimental SW (English)**	**Experimental RW (English)**	**Experimental SW (English)**
have a 0.33 density or higher	88	102	**68.8%**	**80.3%**
have a density lower than 0.33	40	25	31.2%	19.7%
total	128	127	100%	100%

*Table 46: significantly more experimental swear words by English-speaking participants have a 0.33 unsonority density or more (**p = 0.04391**)*

	Experimental RW (French)	**Experimental SW (French)**	**Experimental RW (French)**	**Experimental SW (French)**
have a 0.33 density or higher	54	88	**42.2%**	**68.8%**
have a density lower than 0.33	74	40	57.8%	31.2%
total	128	128	100%	100%

*Table 47: significantly more experimental swear words by French-speaking participants have a 0.33 unsonority density or more (**$p=3^{-05}$**)*

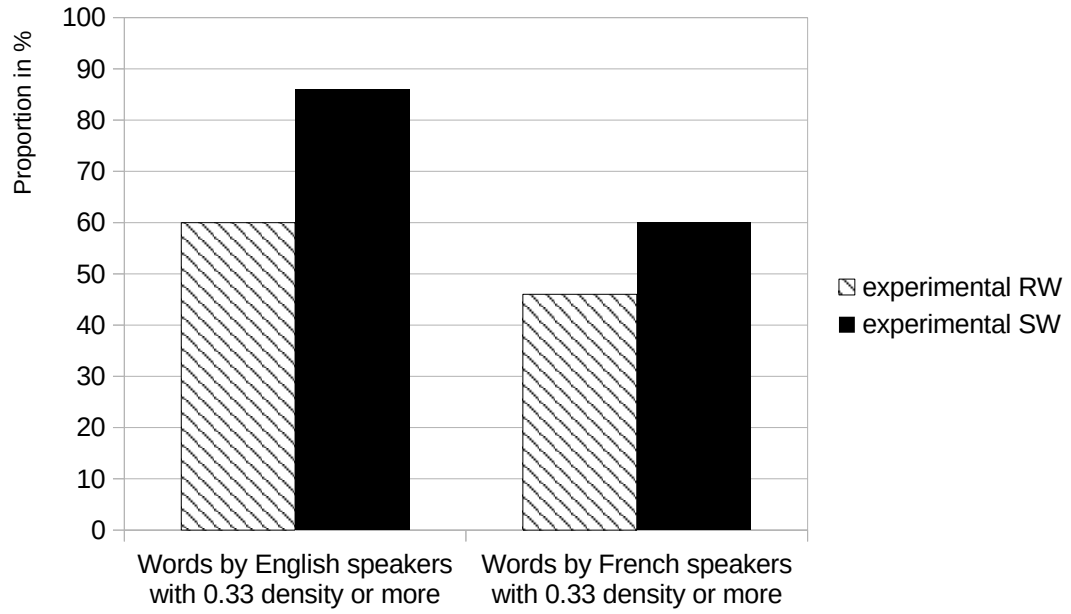

Figure 23: Tendencies for experimental swear words to contain at least one unsonorous consonant out of three phonemes (0.33 density) ($p<0.05$ for both)

In sum, invented swear words in both English and French follow the same tendency as their real-life counterparts. The results of our experimental study eliciting invented swear words from English and French speakers confirm the hypothesis that there is a particular sound-symbolic association, which speakers exploit unconsciously when inventing swear words and regular words in an experimental setting (RQ2a). We discuss those results for our first research questions (RQ1 and RQ2a), as well as the meaning of swear words involved in that sound-meaning association (RQ2b), and the implications that such sound-meaning associations have for the notions of double articulation and arbitrariness (RQ3) in the discussion in the next chapter.

Chapter 4. Discussion

In this chapter, we discuss the implications of our results for the study of swear words and how it relates to the historical processes that lead to the interdiction of specific words. We also describe data from the first case study on real-life French swear words which confirm elements of our definition for swear words and how this helps us to reach a proposition for the meaning of swear words, involved in our sound-meaning association. We then discuss how our own study and other studies on sound symbolism invite to reanalyse double articulation and arbitrariness.

4.1 The best phonemes for swearing: data analysis

The results of our three studies can be summarized as follows. Our first study shows a tendency of real life swear words of English and French to contain a higher number of the most unsonorous consonants, i.e., more plosives, voiceless fricatives, and affricates. Namely, the number of swear words containing at least one unsonorous consonants out of three phonemes is significantly higher in swear words compared to the rest of the lexicon. We observe the same tendency in our second study about swear words in English-fiction, compared to the same non-swearing lexicon – the tendency does not reach the conventional threshold for significance in French, but is observable nonetheless. Finally in our third study, experimentally elicited alien swear words show the

exact same significant tendency: the alien swear words tend to contain more unsonorous consonants compared to the alien non-swear words. This suggests that not only do swear words show a statistical tendency to contain those sounds in the authentic lexicons of English and French, that pattern is also picked up unconsciously by speakers, so that a sort of "unsonorous-swearing" form-meaning association is part of their linguistic knowledge: otherwise, why would speakers follow the same pattern when selecting or inventing words to serve as swear words in their fictional works, and why would they follow the same pattern when asked in an experimental setting to spontaneously invent alien words? This suggests, first of all, that for native speakers of English unsonorous phonemes feel intuitively like the best sounds for swearing; secondly, it seems unlikely that for the two languages in question, a word with no or a low proportion of unsonorous phonemes, would end up as a swear word. This may explain anecdotal observations we can make about words or phrases that never – or never as clearly – became the target of the swear word taboo, despite their taboo referential meanings, contrary to (near-)synonymous items with higher unsonority densities.

For English, the contrast in terms of unsonority densities can be represented as in Table 48 below, opposing each swear word with a referential equivalent that is not a swear word.

English non-swear word (with density)	**(Near-)synonymous swear word (with density)**
vagina /vəˈ**dʒ**aɪnə/ (0.17)	*cunt* /ˈ**k**ʌn**t**/ (0.5) *pussy* /ˈ**p**ʊ**s**i/ (0.5)
penis /ˈ**p**iːnɪ**s**/ (0.4)	*dick* /ˈ**dɪk**/ (0.66) *cock* /ˈ**k**ɒ**k**/ (0.66)

English non-swear word (with density)	(Near-)synonymous swear word (with density)
testicles /ˈ**testɪkə**lz/ (0.44)	*bollocks* /ˈ**b**ɒlə**ks**/ (0.5)
bottom /ˈ**bɒtə**m/ (0.4)	*ass* /ˈɑː**s**/ (0.5)
excrement /ˈe**ksk**rəmən**t**/ (0.4) *stool* /ˈ**st**uːlz/ (0.4)	*shit* /ˈ**ʃɪt**/ (0.66)
urine /ˈjʊərɪn/ (0)	*piss* /ˈ**pɪs**/ (0.66)
make love /ˈmeɪ**k** ˈlʌv/ (0.17) *have sex* /ˈ**h**æv ˈ**se**k**s**/ (0.57)	*fuck* /ˈ**f**ʌ**k**/ (0.66)
Lord /ˈlɔː**d**/ (0.33) *Heavens* /ˈ**h**evənz/ (0.17)	*God* /ˈ**g**ɒ**d**/ (0.66)

Table 48: Unsonority densities of non-swearing alternatives vs. swear words, in English

For French, the same contrast in the unsonority densities of non-swear words versus swear words can be illustrated with the examples in Table 49 below.

French non-swear word (with density)	(Near-)synonymous swear word (with density)
vagin (“vagina”) /vaʒɛ̃/ (0)	*chatte* (“pussy”) /**ʃ**a**t**/ (0.66)
verge (“penis”/“rod”) /vɛʁʒ/ (0) *pénis* (“penis”) /**p**eni**s**/ (0.4)	*bite* (“dick”/“cock”) /**b**i**t**/ (0.66)
urine (“urine”) /yʁin/ (0)	*pisse* (“piss”) /**p**i**s**/ (0.66)
faire l’amour (“to make love”) /**f**ɛʁ l amur/ (0.13)	*baiser* (“to fuck”) /**b**eze/ (0.25)
lupanar (“brothel”) /ly**p**anaʁ/ (0.14) *maison close* (literally “closed house”, i.e., “brothel”) /mɛzɔ̃ **k**loz/ (0.13)	*bordel* (“brothel” / “mess” / “fuck!”) /**b**ɔʁ**d**ɛl/ (0.33)
Jésus Christ (“Jesus Christ”) /ʒezy **k**ri/ (0,14) *Seigneur* (“Lord”) /**s**ɛɲœʁ/ (0,2) *Vierge Marie* (“Virgin Mary”) /vjɛʁʒ maʁi/ (0)	*Dieu* (“God”) /**d**jø/ (0.33)

Table 49: Unsonority densities of non-swearing alternatives vs. swear words, in French

The central focus of our thesis has been to evaluate whether swear words contain more unsonorous sounds than regular words. While our studies indeed suggest that there is a statistical and cognitively real distinction between swear words and non-swear words in terms of their sonority, we need to remain aware of some of the limitations of our study, because our analysis is only a first tentative at identifying a sound-meaning pattern among swear words. These concern the threshold in proportion of unsonorous consonants, how the categories of phonemes are involved in the tendency, what categories are involved, and what words the sound-meaning tendency exists in.

First of all, in order to analyse a tendency in our data for our three studies, we used a 0.33 threshold for the proportion of unsonorous consonants,

based on words spontaneously listed by Hergé in an interview about his fictional swear words. Hergé's spontaneous list suggested that this proportion was enough to sound like a swear word, but this does not necessarily mean that said proportion corresponds to a cognitively real threshold for the word to be felt like a good swearword in the pattern we are investigating. More fine-grained analysis of the data, and more experimental data with native speakers, is needed to confirm that the suggested somewhat artificial threshold of 0.33 is cognitively real. We can imagine, for example, that the actual threshold is lower, like 0.25. Alternatively, we can imagine that any word with comparatively more unsonorous phonemes is more likely to feel like a swear word, so that the pattern is more gradual.

Moreover, we have grouped the three most unsonorous categories of phonemes together (plosives, affricates, voiceless fricatives), but their unsonority levels are different: voiceless plosives are clearly more unsonorous than voiced plosives. It will be worth investigating in future more refined studies whether that relative unsonority has any impact, for example whether the most unsonorous phonemes, i.e., voiceless plosives are significantly more present in swear words than voiceless fricatives or voiced plosives.

We also grouped the three categories and looked for a tendency to contain more of them, but a comparable analysis could be conducted if the reverse tendency is also observable: do swear words contain fewer sonorous consonants, i.e., fewer voiced fricatives, or fewer sonorants? Our thesis has been centred around the question whether swear words contain more unsonorous sounds than regular words. However, when it comes to sound-meaning associations, other articulatory dimensions should also be looked at to complete or nuance our present results, such as the proportion of vowels or consonants in

swear words, or the type of vowels (high, short, etc.). If such differences turn out to be observed, it would mean that the unsonority pattern is actually even more systematic than what our statistical tests detected here, indicating that the unsonority pattern involves other tendencies than the ones that we have studied.

The central parameter investigated in our studies is the unsonority to test Yardy's (2010) claim that swear words contain more unsonorous consonants. Another perspective to analyse acoustic differences between swear words and non-swear words is looking at the frequency spectrum. This is what underlies Ohala's (1994, 1984) frequency code, positing that phonemes feel more or less aggressive depending on the concentration of intensity at different frequencies in the spectrum. Ohala (1984: 9) says for example that voiceless obstruents should feel less aggressive because they have a higher frequency, and that the reverse for voiced obstruents. However, our data show that voiced plosives are more present, voiced fricatives less present, and voiceless obstruents more present in swear words. Another issue is that the frequency code is most often mentioned in the context of vowels – see Section 2.1.3. It could nevertheless be interesting to look into this more exhaustively.

Finally, we focused our attention on swear words, but it would be worth investigating if the tendency does not apply to words of the informal register in general – we develop on that issue in Section 5.2 of our conclusion.

In other words, the patterns that we have observed in our three studies deserve to be exploited further looking at other dimensions, such as the number of vowels versus consonants and the frequency code of the sounds making up swear words versus non swear words. Despite the more narrow focus of our studies looking at the dimension of sonority only, the

"unsonorous-swearing" tendency they suggest leads to the question of what motivates such an unsonorous pattern in terms of meaning.

4.2 On the loss of referential meaning, and the *gros mots* categorisation

Our definition of swear words discussed in Section 2.2 allows us to discuss what meaning is involved in the sound-meaning association in swear words. Before addressing this issue in the next section, we need to show how our results confirm two claims, based on our intuitions as a native speaker of French, and that allowed us to reach this conceptualisation. The first claim, discussed in Section 2.2.2, was that numerous swear words of French lost their referential meanings to taboo subjects like religion, sexuality, or body waste. The second claim, discussed in Section 2.2.5, was that the French *gros mots* category includes slurs related to sexuality, but excludes slurs unrelated to sexuality. We confirm the two claims here by comparing the answers of our respondents to the different questions of the questionnaire on real-life French swear words, described in Section 3.1.

The first claim was that some prototypical French swear words lost their referential meanings to a taboo subject. The six swear words we took as examples were *putain* ("Fuck!"/"fucking (X)"), *con* ("jerk"/"dumb"), *bâtard* ("bastard"), *bordel* ("mess"), *enfoiré* ("scumbag"), and *foutre* ("to do"/"to make"/"to put"). The lost taboo domain that they all referred to, as discussed in Section 2.2.2, is sexuality. One question aimed at collecting swear words related

to sexuality (Question 3 "List *gros mots* that are related to sexuality."), while six other thematic questions aimed at collecting other swear words. Varying numbers of respondents gave those six words throughout the seven questions – from 5 (9%) who gave *foutre* ("to do"/"to make"/"to put") as a *gros mot* to 54 (96%) who gave *putain* ("Fuck!"/"fucking (X)") as a *gros mot*. What is noticeable is that for each one of the six words, a large majority of these respondents (from 60% to 100%) did not give the same word when asked for sexual swear words at Question 3. This strongly suggests that speakers do not think of these words as conveying a sexual meaning anymore. The data also suggest that a few other words might have followed that same historical tendency: *niquer* ("to fuck"/"to screw") from 3 to 1 respondent (-66.7%), *pétasse* ("tart") from 5 to 1 respondent (-80.0%), and *salaud* ("dirty man"/"scumbag") from 15 to 3 respondents (-80.0%).

French swear word	**Frequency as a swear word**	**Frequency as a sexual swear word**	**Difference**
foutre ("to do" / "to make"/ "to put")	5	2	-60.0%
niquer ("to fuck" / "to screw")	3	1	-66.7%
pétasse ("tart")	5	1	-80.0%
salaud ("dirty man")	15	3	-80.0%
putain ("Fuck!"/"fucking (X)")	54	10	-81.5%
con ("jerk"/"dumb")	25	3	-88.0%
bordel ("mess")	19	2	-89.5%
bâtard ("bastard")	21	1	-95.2%
enfoiré ("scumbag")	18	0	-100.0%

Table 50: French swear words who lost their referential sexual meaning

One could imagine that our respondents shied away from giving those words precisely because of the taboo on sexuality. However, our data reveal by contrast how truly sexual swear words did not make respondents uncomfortable and were easily given for Question 3. These words were given as swear words by respondents, and a majority of those respondents also gave them as sexual swear words. This shows how clearly these words retain their sexual meaning.

French swear word	Frequency as a swear word	Frequency as a sexual swear word	Difference
chienne (“bitch”)	4	4	0.0%
suceuse (“(female) (cock)sucker”)	4	4	0.0%
baiser (“to fuck”)	3	3	0.0%
chatte (“pussy”)	3	3	0.0%
tarlouze (“fag” / “sissy”)	3	3	0.0%
couilles (“balls [testicles]”)	2	2	0.0%
gouine (“dyke”)	2	2	0.0%
puceau (“(male) virgin”)	2	2	0.0%
queutard (“sex-obsessed man”)	2	2	0.0%
saligaud (“dirty man” [scumbag])	2	2	0.0%
suce-boules (“ball-sucker”)	2	2	0.0%
bite (“cock”/“dick”)	7	6	-14.3%
pédé (“fag”)	20	17	-15.0%
tapette (“fag” / “sissy”)	5	4	-20.0%
enculer (“to sodomise”)	4	3	-25.0%
pédale (“fag”)	4	3	-25.0%
pute (“whore”)	41	28	-31.7%

salope (“dirty woman” / “slut”)	45	30	-33.3%
cul (“ass”)	3	2	-33.3%
enculé (“sodomised man”)	43	26	-39.5%

Table 51: French swear words from our questionnaire data who retain their referential sexual meaning

The data on *putain* (“Fuck!”/“fucking (X)”), *con* (“jerk”/“dumb”), *bâtard* (“bastard”), *bordel* (“mess”), *enfoiré* (“scumbag”), and *foutre* (“to do”/“to make”/“to put”) confirm our intuitions that French swear words often lost their referential taboo meaning but not their swear word status. It contributes again to the idea that the reference to taboo subjects like religion, sexuality, or body waste, is irrelevant to describe the synchronic value of swear words in a language, and thus irrelevant when trying to determine what their common meaning might be.

Our second claim based on intuitions as a native speaker was that the French *gros mots* (“big words”) category includes sexual slurs, but excludes non-sexual slurs. We analyse below the data relevant to this issue from the same questionnaire on real-life French swear words described in Section 3.1.

As already mentioned, we asked native speakers of French to list the *gros mots* (“big words”) and insults they knew via nine thematic questions. Question 6 asked for the overlap between the two, i.e., for insults that are also *gros mots* (“List *gros mots* that are insults.”). Question 7 asked for insults that are not gros mots ((“List insults that are NOT *gros mots*.”). Question 8 asked for slurs, regardless of whether they considered them as *gros mots,* using a paraphrase because there is no equivalent of English *slur* in French (“List insults (*gros mots* or not) that must not be said because they target a group of people”).

If speakers considered some slurs to be *gros mots* (“big words”), they could have listed slurs on several occasions before, in particular when asked for

insults that are also *gros mots* on Question 6. The first questions prompted many occurrences of sexual slurs, e.g., *pédé* ("fag") or *pute* ("whore"): 26 respondents (46%) gave that sort of word for first six questions, and repeated the same word for the question on slurs. To put it differently, the answers of 46% of respondents our respondents show unambiguously that they consider these words to belong to both the *gros mots* ("big words") category and the slurs category. So it is safe to say that the French *gros mots* category includes sexual slurs.

For non-sexual slurs, a few respondents disagree but the data suggest that they are not indeed included in the *gros mots* category. The answers of 25 respondents (45%) hint that for them, non-sexual slurs are not *gros mots*. They went through six questions on *gros mots* before, including one that asked for *gros mots* that happen to be insults, and they gave non-sexual slurs only when asked for slurs on Question 8. Their 40 answers include 11 slurs for physically or mentally disabled people (*attardé*, *autiste*, *cotorep*, *débile*, *éclopé*, *handicapé*, *mongol*, *schizo*, *taré*, *triso*, *trisomique*), 8 anti-Arab slurs (*Arabe*, *bougnoule*, *boukak*, *bicot*, *crouille*, *petit gris*, *rebeu*), 5 anti-black slurs (*bamboula*, *nègre*, *négro*, *noir*, *sale nègre*), 5 slurs against East Asian people (*bouffeur de chiens*, *bridé*, *chinetoque*, *jaune*, *niakoué*), 3 antisemitic slurs (*Juif*, *gros feuj*, *youpin*), 5 miscellaneous slurs (*gonzesse* ("chick"), *nain* ("dwarf"), *portos* (an anti-Portuguese slur), *polack* ("Polack"), *sale trans* ("dirty transgender")) and finally, 3 far-fetched items that could be considered as slurs in particular circumstances (*has been* ("has been") which could be seen as a slur if used against old people, *nazi* ("nazi") which could be seen as a slur if used against German people, and *terroriste* ("terrorist") which could be seen as a slur if used against Muslim people). The sudden appearance of these 40 non-sexual slurs

only when asking about slurs, and not earlier in the questionnaire when asking for *gros mots*, strongly suggests that these are not categorised as *gros mots*.

Even more, the answers of six respondents (11%) show unambiguously that for them, non-sexual slurs are not *gros mots*. They gave nine non-sexual slurs in the questions on slurs, and also gave them when asked them insults that were not *gros mots* ("List insults that are not *gros mots*"): *attardé* ("retarded"), *bougnoule* (an anti-Arab slur), *débile* ("moron"), *éclopé* ("cripple"), *Juif* ("Jew"), *jaune* ("yellowman"), *petit gris* (an anti-Arab slur), *taré* ("defective", "loony") *trisomique* ("person with Down syndrome"). By contrast, the answers of only three respondents (5%) show unambiguously that for them, some non-sexual slurs are also *gros mots*. They gave four non-sexual slurs in the questions on *gros mots*, and repeated them in the question on slurs: *bougnoule* (an anti-Arab slur), *feuj* (a slang form of *Juif* "Jew"), *sale Juif* ("dirty Jew"), and *youpin* (an anti-Jew slur, equivalent kike). If we were to include non-sexual slurs in the French *gros mots* ("big words") category, we would account for the answers of 5% of our respondents who think non-sexual slurs are *gros mots*, but not the larger 11% who they think the opposite. It would also fail to explain why 25% of respondents listed non-sexual slurs only when asked for slurs and never before when asked for *gros mots*. One might imagine that slurs trigger more unease for participants and they did not feel comfortable telling them before being asked explicitly for slurs. However, this explanation would fail to account for why 46% of respondents felt comfortable listing sexual slurs when asked for *gros mots* in the first questions, before the specific question on slurs.

As a side remark, if we look at the four answers of the 5% who seem to think that some non-sexual slurs count as gros mots, we can see that three out of four were given for a question on religious swear words ("List *gros mots*

that are related to religion") which apparently confused respondents. They overwhelmingly did not respond (55.4%), or gave far-fetched answers. For example, five respondents gave *cul-bénit* (literally "blessed ass"), a rather rare insult against overly devout people, probably because it contains the word *bénit* ("blessed") and the common swear word *cul* ("ass"). This is certainly due to the fact that French religious swear words – like the interjection *nom de Dieu*, literally "name of God" – have lost their swear word status and become benign, even quaint interjections in today's French. If we ignore the answers to that visibly confusing question on religion, there remains only one last respondent out of 56 who considers that a non-sexual slur – namely, *bougnoule* (an anti-Arab slur) – also counts as a *gros mot*.

In short, our intuitive categorisation explains the French data much more efficiently – it accounts for the answers of 95% of the respondents – than alternative categorisations. For this reason, and because *gros* ("big") in *gros mots* ("big words") comes historically from *grossier* ("rude") (Rouayrenc 1996: 3–6), we propose that *gros mots* ("big words") is best translated to English as *rude swear words*. For most French speakers, the French *gros mots* ("big words") category includes sexual slurs but excludes non-sexual slurs. For some, but not all English speakers, the *swear word* category includes all slurs. We saw in Section 2.2.6 that non-sexual slurs are apparently not often included in the swear word category, not least because the cultural concept of slurs does not exist in every speech community that otherwise has swear words. These differences in categorisation around slurs suggest that they are somewhat peripheral to the collective conceptualisation of the category. Consequently, we can safely ignore the specificities of slurs when looking for the common meaning of swear words involved in our sound-meaning association. This is

what we aim to do in the next section.

4.3 A hypothesis on the meaning of swear words

As will be recalled from our discussion in Section 2.2, swear words are a small (sub)register of the most informal words in a given language that share a number of characteristics. They are strongly forbidden words that express and trigger powerful emotions, to the point of unease, laughter, or measurable physical reactions like pain relief or sweating. They are referred to with an everyday term, showing that they correspond to a category in the speakers' minds. The taboo nature applies to the words themselves, not to their referential meaning: there is a non-taboo alternative for each word. The literature on swear words suggests that, consistently with those characteristics, they express a kind of contextual, emotional meaning. That specific meaning may be, as suggested by Haiman (2018: 209–212) a "violation of hearer's space", i.e., "familiarity or aggression". It can be "hostility" (Yardy 2010: 12–20; 71–78). Both Haiman and Yardy suggest that this is done via iconic pairings. From a diachronic point of view, both pairings may have played a role in the development of the tendency we observe in today's English and French swear words. From a synchronic point of view, however, it is difficult to determine whether any of those two proposed meanings is still relevant in the speakers' minds, and if so, which one. None of the two hypotheses is completely satisfactory by itself. Haiman proposes an association between plosives and "familiarity or aggression", yet our results show that not just plosives, but also other consonants are more present in swear words. Yardy proposes a meaning

that is only negatively valenced, whereas we know that swear words can be used with positive effects, as in in-group swearing among friends, and many swearing utterances are positively valenced, as interjections or (positive) intensifiers.

In an attempt to reconcile those two propositions with observations from the literature as well as the findings from our three case studies data, we could say that swear words associate unsonorous consonants (not just plosives, contrary to Haiman's suggestion) on the one hand with a meaning of "violation of hearer's space" / "familiarity or aggression" on the other hand. That meaning is a powerful contextual, emotional meaning that can be either positive or negative. This association exists at least in English and French, but more empirical cross-linguistic studies on swear words in unrelated languages would have to confirm whether that proposed pairing is indeed motivated or even iconic. At the very least, given that almost no English and French swear words are cognates – the only possible candidates are *cunt* and *con* ("jerk"/"idiot", historically "pussy"/"cunt") – the association can indeed be considered iconic.

This is the hypothesis we can derive directly from the literature. Based on our personal reflection, we can also make an alternative and simpler, more elegant iconic hypothesis. The contextual emotional meaning of swear words, which are taboo words, is perhaps merely the breaking of the taboo itself. The unconscious association of "taboo-breaking" meaning with unsonority might be an historical accident, but that doesn't exclude that it also makes sense to speakers that sounds that are are felt intuitively harsher or noisier, are considered quite suitable for a meaning like "I am breaking a taboo". Unconsciously and emotionally, speakers could interpret harsh, noisy, "hard-object-breaking" sounds as breaking a taboo: collectively, they would

tend to forbid words with such phonemes, while ignoring words where they are absent or less present.

4.4 The meanings of sounds: on double articulation

Sound-meaning associations of the kind we discuss in this thesis, seem to contradict two classical tenets of linguistics, namely Saussure's arbitrariness and Martinet's double articulation. The discussion around these issues is fraught with polemic claims and potential misunderstandings. Mauro (2005 [1967]), for example, argues at length that arguments for arbitrariness are self-evident and implicitly compares researchers questioning arbitrariness to "an imbecile" (2005 [1967]: 446). On the other side, and in equally strong terms, Winter boldly states that "iconicity, not arbitrariness, is a design feature of language" (2021), and that "there is nothing arbitrary about the linguistic sign" (2023). Both supporters and critics of Saussure and Martinet consider that sound-meaning associations fundamentally contradict Saussure's and Martinet's tenets. Either they exist and Saussure and Martinet were wrong, or they do not exist and Saussure and Martinet were right. We argue for an alternative, more nuanced position, that sound symbolism is not so incompatible with arbitrariness and double articulation. In order to see how it could be so, it is imperative to clarify issues involved. As the following discussion will show in this section and the next, Saussure's arbitrariness and Martinet's double articulation refer to different kinds of meaning-making than sound symbolism. Namely, Saussure's arbitrariness and Martinet's double articulation are concerned with how language conveys referential meaning in a

way that is compositional, deterministic, and conscious. In sound-meaning associations on the contrary, the meaning may or may not be referential, the form is usually non-compositional, the association is probabilistic and contextual, and it is typically unconscious.

Moreover, what emerges from findings in the literature and our own findings necessitates to reanalyse arbitrariness and redefine it more specificly. Following that more specific definition, arbitrariness does not entail double articulation, contrary to Saussure's implicit understanding, and it is fundamentally distinct from it. It also allows us to analyse notions around sound-meaning associations more precisely.

We start our discussion in this section with a reassessment of double articulation because it is comparably simpler than Saussure's arbitrariness, and because it will allow us to contrast it with arbitrariness in the next section.

We propose that swear words in English and French display an unconscious sound-meaning association. As indicated in the introductory discussion (Section 2.1.3), such unconscious sound-meaning associations seem to run counter to a classical tenet of linguistics that phonemes are meaningless, known as duality of patterning or double articulation. However, we will show that it is possible to reconcile double articulation with the existence of sound-meaning associations.

There is a functional argument for saying that we need phonemes to be meaningless. We illustrate this with a fictional, theoretical case for the sake of the argument. Let us imagine that in a given language, speakers associate phonemes with meanings. By definition, there is a limited number of existing phonemes in a language. For the sake of making the argument simpler to follow, let us give this language only three vowels: /a/ /i/ and /u/. Let us give

this language six consonants: /p/ /t/ /k/ /f/ /ʃ/ /s/. That makes nine phonemes.

Let us give a list of possible meanings to each of the nine phonemes. For the sake of showing how strong the argument for double articulation is, let us give abstract meanings that seem easily adaptable and relevant to a large number of communicative contexts:

/a/ ↔ "big"	/p/ ↔ "light"	/f/ ↔ "good"
/i/ ↔ "small"	/t/ ↔ "grey"	/ʃ/ ↔ "neutral"
/u/ ↔ "deep"	/k/ ↔ "dark"	/s/ ↔ "bad"

Thus in that language, /pa/ could mean "a great light", /asu/ could mean "shark", as it is a big, bad animal that dwells in the depths of the sea, /ʃik/ could mean "ants", as they are neutral, small animals that live in the dark, etc.

We can easily see the functional problems that such a system poses. First of all, speakers can parse the meaning of a sound sequence from the meanings of each individual sound, for example in our system, the meaning of /pa/ "big light" can be parsed from the meaning of /p/ "light" and /a/ "big". That parsability means that one sound sequence can be interpreted as referring to other concepts combining the same meanings, and create too much ambiguity and confusion. For example, if /asu/ means "shark" because it refers to something combining the "big", "bad", and "deep" properties, how does one refer to a different big, bad animal that dwells in the depths of the sea, like "whale" ? How does one refer to an animal that is big, bad, but dwells in the depths of the forest instead, like "bear" ? We could imagine that order plays a role, but that would only take us so far. If /asu/ means "shark", /sau/ means "whale", /usa/ means "bear", etc. that leaves us with $3^3 = 27$ possible combinations. One can certainly imagine more than 27 concepts that would combine the "big", "bad", and "deep" properties.

To avoid that ambiguity, we can also try to give more specific meanings to our phonemes, like “has teeth”, “lives in the water”, “lives in the forest”, etc. but then the reverse problem arises, i.e., how to express generic, abstract concepts like “animals” if the meanings of our phonemes refer to such specific properties ? How does one express other specific meanings like “water”, or “apple” ? How does one express other abstract meanings unrelated to those properties, like “people”, or “life” ?

This is the fundamental logical problem that double articulation addresses: if phonemes have meanings, then other meanings, and combinations of such meanings, cannot be expressed. A sound-meaning system like the one we described here allows to express some meanings, but it fails to capture many more. The infinity of concepts required by human communication would imply that, in order to express them in language, we would need not just nine phonemes, or a larger number of phonemes like twenty or thirty, but actually an infinite number of meaningful phonemes.

Instead, a limited number of meaningless sounds, ordered in meaningful sequences, can refer to an infinity of meanings (Martinet 2008 [1960]: 37–44; Ohala 1994). The sounds they contain are multi-purpose tools, adaptable to any concept. To take an example from English, the three sounds /l/ /aɪ/ and /f/ combined together, can refer to completely unrelated concepts, with vastly different degrees of abstractness, depending on the order in which they are used: /laɪf/ ↔ “life” and /faɪl/ ↔ “file”.

Double articulation is useful to describe the fact that we cannot associate phonemes with meanings in the way we have described here. However, that rationale describes only a specific type of meaning-making. It looks for phoneme-meaning associations exclusively after the model of

word-meaning and morpheme-meaning pairings, and only one specific kind of meaning. This is a relevant, but narrow perspective.

The first issue with that rationale is that it holds a purely compositional view of meaning-making. In that view, the meaning of a form is supposed to be logically predictable, at least in part, from the meaning of its parts, for example we can derive the meaning of a word like *readable* if we know the meanings of the morphemes *read* and *-able*. We can logically derive the meaning of a sentence like *I am hungry* because we know the meanings of the words *I*, *am*, and *hungry*. We know, however, that this is not always the case in language, typically with idiomatic phrases, for example the meaning of *The cat is out of the bag* ("the secret is revealed") cannot be logically derived from the meaning of its individual words. We also observe the same discrepancy with more frequent items like phrasal verbs, for example the meaning of *to look after* ("to take care of") cannot be derived from the meanings of *to look* and *after*. The meaning of a form can be, but is not necessarily based on the meaning of its parts. Therefore, if words have meanings, and if phonemes have meanings, the meanings retrievable at those two levels are, however, not necessarily related. For example, it is true that the meaning of *life* /laɪf/ and the meaning of *file* /faɪl/ are completely unrelated, so that they cannot be derived combinations of the meanings of /f/, /aɪ/, and/or /l/, if any. Yet it does not necessarily follow from this, that /f/, /aɪ/, and /l/ are meaningless and do not have other meanings. Language can convey meaning in unpredictable, non-compositional ways.

A second issue is that the double articulation rationale is concerned only with referential meanings, which are the most obvious, immediately retrievable, context-independent meanings that one can define for words and compile in dictionaries. That sort of meaning is more easily described, and

predictable. Yet, language also conveys other (non-referential) meanings that are more emotional and/or attitudinal, for example the meanings of interjections like *Wow!*, or the meaning of swear words. It makes sense that the same could happen at the level of phonemes, and even more for specific categories of words whose unique effects are unrelated to reference, like interjections, ideophones or swear words.

The third issue with Martinet's rationale is that it is concerned with meaning conveyed consciously. It leaves no room for more unconscious parts of meaning-making, which are also the norm for sound-meaning associations. Again, English speakers do not know consciously about the English *gl-* phonaestheme, but it is part of their unconscious linguistic knowledge and they exploit it when interpreting words. Otherwise, this phonaestheme and others would not have priming effects on word recognition comparable to morphemes (Bergen 2004).

Finally, the fourth issue is that Martinet's rationale describes a deterministic, one-to-one unequivocal form-meaning system. One phoneme corresponds to one meaning, one meaning corresponds to one phoneme. It is easy to see that this is not how existing languages attribute meanings to form, not even to words. A one-to-one unequivocal form-meaning system leaves no place for ambiguity or homonymy, which are widely pervasive phenomena when accounting for the meaning of words alone. It also leaves no space for more probabilistic, less easily predictable form-meaning associations that are actually the norm for the sound-meaning tendencies we observe: to take the example of the *gl-* phonaestheme (Bergen 2004: 293) described in Section 2.1.3, it means "light"/"vision" in 39% but not all words where the form appears. Meaning is not necessarily attributed following a binary, either/or logic, but can

be attributed by subtle tendencies or in specific contexts instead. Even at the level of words, insights from pragmatics tell us that meaning has to be interpreted from context, alongside the more predictable context-independent meaning. This allows for flexibility in meaning attribution, and for a simple one-way mechanism that makes phonemes both meaningful and ambiguous: a given meaning corresponds to a sound, but a sound does not always correspond to that meaning. We can make a useful comparison here with other ambiguous forms: colours. Fire engines, fire extinguishers, fire hydrants, and fire alarms are all red, for the obvious and simple reason that red is a fitting form to convey the meaning "fire", because fire is red, i.e., it is a case of iconicity, where the form resembles the meaning. Following a similar iconic logic, red is also chosen to express the meaning "blood" in other contexts, for example red flags on pirate ships meant that no opponent would be taken alive, or to take an example from cinema, the artistic direction of the film *Bram Stoker's Dracula* (Coppola 1992) makes constant use of red hues to evoke blood which the titular character famously drinks in order to live forever. In yet other contexts, and this time non-iconically, red in European politics is traditionally associated with left-wing parties, in particular with socialists and communists. This does not imply that whenever humans see red, they unconsciously think of fire, or blood, neither does it imply than Europeans, whenever they see red, think unconsciously of left-wing politics. This only implies that when conveying the meaning "left-wing politics" through colour, Europeans are more likely to use red than other colours, and that in order to convey the meanings "fire" or "blood", humans across cultures are more likely to use red than other colours. But red is also present in many other lived experiences, for example on the leaves of trees in autumn, so that in terms of potential

unconscious associations, it can evoke many other meanings to the human mind. Red can be associated with meanings, sometimes iconically, sometimes culturally, but a question like "What is the meaning of red?" is impossible to answer out of context, precisely because it is associated with so many different meanings. The difficulty is not that a colour is meaningless, but on the contrary, it is overwhelmingly meaningful. Even "natural", i.e., iconic associations like red ↔ "fire" or red ↔ "blood" cannot be accessed out of context, i.e., iconicity is contextual (Winter 2021).

If humans can, consciously or unconsciously, attribute meanings in specific contexts to extremely ambiguous forms like colours, they should be able to do the same for the similarly ambiguous phonemes of their language, in such probabilistic and contextual manner. Attributing meaning to a given phoneme does not entail that it always corresponds to the same meaning. For example, if unsonorous phonemes are associated in swear words with a meaning of "familiarity", "aggression", or "taboo-breaking", as we hypothesised above, we do not claim that somehow, whenever speakers hear unsonorous phonemes in any context, they interpret them as familiar, or aggressive, or taboo. A specific meaning can be retrieved from an ambiguous form only given a context that allows that specific interpretation. In our hypothesis, swear words are used in contexts that allow for the specific, unconscious interpretation that the sounds mean something like "familiarity", "aggression", or "taboo-breaking".

The existence of sound-meaning associations also connects with insights from Construction Grammar (Goldberg 2006). Construction Grammar holds that the whole of linguistic knowledge is captured by a network of Constructions, defined as form-meaning pairings. This applies to words of

course, but also to what are traditionally analysed as grammatical rules and categories, analysed instead as more schematic, i.e., more abstract form-meaning pairings containing variables. For example, instead of describing some verbs as syntactically ditransitive, Construction Grammarians will consider that the DITRANSITIVE Construction consists of the form VERB + NOUN PHRASE + NOUN PHRASE where VERB and the two NOUN PHRASE are three variables that can be filled with other linguistic items that happen to be compatible in terms of meaning, and that form is associated with an abstract meaning, which is approximately "to give something to someone". Grammatical categories like VERB and NOUN are also considered as associated with abstract meanings like respectively, "action" or "entity". In that framework, a sentence like *Pass me the salt* contains the DITRANSITIVE construction VERB + NOUN PHRASE + NOUN PHRASE ↔ "to give something to someone". The variables in that construction are filled each with other constructions, i.e., with the forms VERB, NOUN PHRASE, and NOUN PHRASE associated with their respective meanings, something like "action", "entity" and "entity". They are themselves filled with the compatible constructions *pass*, *me*, and *the salt,* each one a form associated with its meaning; *the salt* itself contains the determiner *the* and the noun *salt,* each one a form with its own meaning. Thus language is "constructions all the way down" (Goldberg 2006: 18), i.e., nothing but a set of form-meaning pairings that can be combined and contained in each other, in order to express and convey meaning. This blurs the classical distinction between lexicon and grammar: that distinction is not understood as a fundamental difference in nature, but instead, as a difference in degree of schematicity of forms, specificity of meanings, of how exactly the meaning is associated with the form, of how clearly the constructions are compatible with

each other, etc. To our knowledge, Construction Grammar has not yet been applied to sound-meaning associations. But if it is correct in seeing linguistic knowledge as a set of form-meaning associations, then this should also be applicable to the lower levels of language, at the level below words and morphemes. Sound-meaning associations of the kind studied in this thesis can be seen as constructions, contained within the words and morphemes. They confirm and expand the Construction Grammar insight that, indeed, language can be made of constructions all the way down, even at the lowest level, for the most basic linguistic forms that speakers exploit in their activity of meaning-making. It is also a view compatible with systemic-functional perspectives on grammar. As Halliday (1985) has pointed out, "a language is a system for making meanings" (Halliday 1985 :xvii). Meaning is not conveyed solely through the lexicon, but "by grammar as well as by vocabulary" (ibid.) – and we would add, sometimes by phonemes as well. As also observed by Wierzbicka, in language, "everything 'conspires' to convey meaning" (1988: 2). In the end, Martinet, and Saussure as well, were talking about different kinds of meaning-making than the ones analysed in sound-meaning associations. We now discuss the issues specific to Saussure's arbitrariness, and how related but difference it is from Martinet's double articulation, in the next section.

4.5 Sound symbolism, double articulation, and arbitrariness

In this section, we aim to reanalyse Saussure's notion of arbitrariness and show

how it requires a more specific definition that allows to distinguish between different kinds of sound-meaning associations.

At the heart of the notion of arbitrariness is a straightforward, practical insight that is part of everyone's lived experience: the meanings of words just have to be learnt, they cannot be guessed. If the link between the word and meaning is random, or to say it differently, unpredictable, unmotivated, then the question is how words originally got their (unpredictable) meaning. Is it just a matter of random social conventions or is there some hidden logic to it?

This unpredictability has of course been noticed long before Saussure. It is debated in Plato's *Cratylus* dialogue written around 360 BCE. It also underlies Shakespeare's well-known phrase *What's in a name? That which we call a rose by any other name would smell as sweet* in Romeo and Juliet.

Enter Saussure. His *Cours de Linguistique Générale* ("Course on General Linguistics") is posthumously written, compiled, and published by his former students based on university courses he gave (Bally & Sechehaye 2005 [1916]: 7–11; Mauro 2005 [1967]: 405–409). Along with contemporary linguists like Whitney (Mauro 2005 [1967]: 442), Saussure comments on the unpredictability of the link between words and their meanings (2005 [1916]: 100–102) – and other unmotivated aspects of language, such as the delimitation of phonemes and concepts (2005 [1916]: 155 – 157). He calls that notion *arbitraire* ("arbitrary") in French, translated in English as *arbitrary* – when used as an adjective – or *arbitrariness* – when used as a noun. Saussure also uses the term *immotivé* ("unmotivated") to mean the same thing.

French *arbitraire* and English *arbitrary* can both refer to the unpredictable, self-imposing, unmotivated decisions of an autocratic *arbiter*, i.e.,

a power figure, for example a king. *L'arbitraire du pouvoir* ("the arbitrariness of power") or *pouvoir arbitraire* ("arbitrary power") and their English equivalents describe situations where people are at the mercy of such autocratic decisions. There is also another potential cognitive association for French speakers: *libre-arbitre* (literally "free arbiter") is the French term for English *free will*, i.e., the unpredictability of individual decisions. In short, that choice of word by Saussure suggests that speech communities attribute words to meanings in the unpredictable, uncontrollable way that a king makes and imposes his decisions on everybody else, with no need for justification, or the way an individual person makes freewill choices. These are not bad metaphors. We can see how other choices of words like *conventional* used by Whitney (Mauro 2005 [1967]: 442) have different connotations, so they evoke different metaphors, and hint at slightly different conceptualisations of the same phenomenon: *convention(al)* can evoke people gathering and deciding together what rules they should give each other – it sounds somewhat more democratic, with possible justifications. In particular, *arbitrary* is more easily understood as *unmotivated*, *without reason*, *random*, as opposed to *determined*, so it implies the idea already mentioned that a meaning does not determine its corresponding word, or the other way around. The meaning is no motivation, no cause for its corresponding form, or the other way around. To illustrate the notion of arbitrariness, Saussure takes an example from French, and an example of cross-linguistic difference for a similar concept (2005 [1916]):

> The link that unites the signified and the signifier is arbitrary, or […] to put it more simply: *the linguistic sign is arbitrary*. Thus the idea of "sister" is linked, due to no inner relationship, with the sequence of

> sounds [/s/ /œ/ and /ʁ/ in the French word *sœur* ("sister")] that serves as its signifier; it could just as well be represented by any other: this is proven by differences between languages and the very existence of different languages: the signifier "ox" has as a signifier [/bœf/ *bœuf* ("ox") in French] on one side of the border, and [/ɔks/ *Ochs* ("ox") in German] on the other side. […] We mean that [the linguistic sign] is *unmotivated*, i.e., arbitrary with regards to the signified, with which it has no natural link in the reality. (2005 [1916]: 100–101; our translation, emphasis in original)

This is not to say that concepts are identical in every language though, or that it only refers to linguistic diversity: Saussure merely uses these differences to easily illustrate arbitrariness to his students (2005 [1916]: 155–157; Mauro 2005 [1967]: 442–443). The notion would also apply equally, should there be only one human language on the planet, referring to only one set of cultural concepts: the link between words and concepts in that language would be equally unpredictable, random, arbitrary.

Saussure however nuances that notion himself: actually sometimes, to a certain extent, we can predict the meanings of words, if we know words that share morphemes with them, so that some words are relatively motivated. This phenomenon that he calls *relative arbitrariness* accounts for the effects of morphology (2005 [1916]):

> *The sign can be relatively motivated*. Thus *vingt* ("twenty") is unmotivated, but *dix-neuf* (literally "ten-nine", i.e., "nineteen") is not to the same degree, because it evokes the terms it is composed of [*dix* ("ten") and *neuf* ("nine")] […]; taken separately, *dix* ("ten") and *neuf*

> ("nine") are on the same level as *vingt* ("twenty"), but *dix-neuf* (literally "ten-nine", i.e., "nineteen") is a case of relative motivation. It is the same for *poirier* ("pear tree") which brings to mind the word *poire* ("pear") and where the suffix *-ier* evokes *cerisier* ("cherry tree"), *pommier* ("apple tree"), etc. (2005 [1916]: 181; our translation, emphasis in original)

Presenting this as *relative arbitrariness* has the unfortunate effect of making the notion seemingly more complicated and limited, while it easily applies to morphemes, i.e., there is, to paraphrase Saussure, "no natural link" between the form *-teen* – as in *nineteen* – and its meaning, between *-ier* – as in *poirier* ("pear tree") – and its meaning. To account for such effects, Martinet (2008 [1960]: 39) proposes the term *moneme*, which includes words and morphemes: using that terminology, arbitrariness is the notion that the link between the form of a moneme, i.e., a monomorphemic word or a morpheme, and its meaning, is unpredictable. For the sake of simplicity and brevity, the reader should keep in mind that whatever we say about *words* in the next paragraphs also applies to *morphemes*, and despite Saussure's choice of words, the existence of morphemes changes nothing fundamental to his discussion of arbitrariness.

The intuitive unpredictability of the link between words (or morphemes) and their meanings is the issue that linguists allude to when they claim that sound-meaning associations contradict arbitrariness. If sounds can have meanings, then we can predict the meaning of words based on the sounds they contain, or the other way around, predict the sounds that a word contains based on its meaning.

Just like Martinet's double articulation discussed in the previous

section, Saussure's arbitrariness is focused on how language conveys referential meaning in a way that is compositional, deterministic, and conscious. The same remarks we made apply to his discussion. However the attentive reader, given the literature we reviewed in Section 2.1.3 and our discussion in the previous section, may feel that something else is off about this debate. This is due to the fact that Saussure's arbitrariness corresponds to the idea that the phonemes of *words* are not determined by their meanings, or the other way around. The fact that Saussure defines it at the level of words has important logical implications. As a result, Saussure's arbitrariness is the combination of three notions that are closely related but are to be considered separately when we investigate sound-meaning associations. These three notions are:

1. Phonemes are meaningless.
2. Since the parts of a word (its phonemes) have no meaning, the meaning of that whole (a word) cannot be predicted from the meaning of its parts (its phonemes).
3. A linguistic form is not determined by its meaning, or the other way around.

The first notion corresponds to double articulation but is already part of Saussure's understanding of arbitrariness – we already discussed this notion in the previous section. The second notion is entailed by the first one: if the parts of a whole are meaningless, then it is impossible to predict the meaning of the whole. Yet this second notion becomes problematic because it applies to words, but not exclusively to them: the impossibility of predicting the meaning of a whole from the meanings of its parts can be described at many other levels of linguistic forms, like idioms. If we understand *arbitrary* in this sense, the

meaning of fixed phrases like *The cat is out of the bag* ("the secret is revealed") is arbitrary because it cannot be logically derived from the meaning of its individual words. This applies in the same way to the meaning of many multi-word phrases, like phrasal verbs: the meaning of *to look after* ("to take care of") is arbitrary because it cannot be derived from the meanings of *to look* and *after*. If we take a Constructionist approach to grammar as discussed in the previous section, we can argue that the DITRANSITIVE construction is arbitrary in that sense, i.e., that it is impossible to derive its meaning of the whole ("to give something to someone") from the meaning of its parts (the meanings of VERB, of NOUN PHRASE, and of NOUN PHRASE). It is possible to understand arbitrariness in that whole-from-parts sense. This is however quite far from Saussure's initial preoccupations that were so focused on words, and Martinet's understanding of arbitrariness that was still focused on words and morphemes. When we investigate sound-meaning associations and claim that we can then predict to a certain extent the meaning of words based on the meaning(s) of (some of) its parts, we might say that this contradicts arbitrariness – see for example Gutiérrez et al. (2016) who investigate English phonaesthemes and present them as "non-arbitrary" form-meaning correspondences. But then we discuss a phenomenon that is relatively common in language and not restricted to words or morphemes, even though it is particularly salient and systematic there, more radically unpredictable, compared to other levels of linguistic form. Moreover, it also implies that we can discuss arbitrariness without any consideration on cross-linguistic variation or convergence: *to look after* ("to take care of") is arbitrary in English because its meaning cannot be derived from the meanings of *to look* and *after*, and that specific observation can be made by looking at English alone. However Saussure's immediate mention of cross-linguistic

variation, and the concern of many sound symbolism studies for cross-linguistic convergence, shows how arbitrariness in that whole-from-parts sense is often not actually at the core of the discussion.

We would argue instead that the actual notion that Saussure was touching upon in his course, the core notion of arbitrariness, is the third notion above. It is the notion that a linguistic *form* – whether it is a word, morpheme, multi-word phrase, abstract grammatical Construction, or phoneme – is not determined by its meaning, or the other way around. As a logical consequence, one cannot predict a linguistic form from its meaning, or the other way around. This one notion is related to cross-linguistic variation or convergence. If we find significant form-meaning tendencies across unrelated languages, then there has to be some cross-linguistic determinism at play between form and their meanings, and arbitrariness is not so pervasive in languages as initially thought by Saussure. This applies to any linguistic form, including phonemes, which is the specific form that cross-linguistic studies like the ones reviewed in Section 2.1.3 investigate for such motivated form-meaning associations.

We propose to redefine arbitrariness specifically as this third notion, i.e., the notion that a form is not determined by its meaning or the other way around. This is a separate issue from double articulation, i.e., from the meaninglessness of phonemes. As such, we can say that some sound-meaning associations are arbitrary. For example, we can say that there is no apparent motivation, no determinism predicting that the sound sequence /gl/ should be associated with the meaning "light" or "vision" as in the English *gl-* phonaestheme (Bergen 2016a: 60–63; 2004: 293). According to this understanding of arbitrariness, the *gl-* phonaestheme is an arbitrary sound-meaning pattern, contrary to Gutiérrez et al. (2016)'s understanding of

the term (*non-*)*arbitrary*. This redefinition of arbitrariness allows to see more clearly that there are language-specific sound-meaning associations, which are a part of speakers' unconscious linguistic knowledge and such a field of inquiry for linguistic research. At the same time, it allows to see how these language-specific pairings are fundamentally different from motivated, cross-linguistic sound-meaning associations also discussed in the literature, typically in iconicity studies. Our own studies on English and French swear words point to the existence of a sound-meaning association in them, but our data is insufficient to conclude whether this association is motivated or arbitrary – in the sense that we redefine arbitrariness. This understanding of arbitrariness also illustrates how it is related to but fundamentally distinct from double articulation. Double articulation entails arbitrariness – if sounds have no meanings, then at level of words and morphemes, their meanings are unpredictable from the form (their phonemes) or the other way around. Yet arbitrariness, redefined in that specific sense, does not entail double articulation – again, the English *gl-* ↔ "light"/"vision" phonaestheme (Bergen 2004: 293) is arbitrary, but it contradicts double articulation.

We would argue that this redefinition of arbitrariness also allows to distinguish their opposites. Arbitrariness, as we define it, is opposed to motivated sound-meaning associations: we propose to call this notion *sound motivation*. Double articulation is opposed to sound-meaning associations, whether they are motivated or unmotivated: we propose to call this notion *sound symbolism*, because a *symbol* is – in the literature and also in a more everyday sense – a form-meaning pairing that can be arbitrary. For example, doves and olive branches are symbols for peace, and there is no obvious reason why one should consider these pairings as non-arbitrary.

To conclude this section of our discussion:

- Sounds can be associated with meanings in other ways than discussed by Saussure and Martinet, i.e., they can be part of unconscious, and/or non-compositional, and/or probabilistic form-meaning pairings, that convey either referential or non-referential meaning. This completes Martinet's tenet of double articulation, implicit in Saussure's rationale, which holds that sounds have no meanings, but only applies to conscious, compositional, deterministic meaning-making conveying referential meaning. We propose to call this notion *sound symbolism*.
- Sounds can be associated to meanings in ways that are undetermined, unmotivated, random, unpredictable, and for that reason are language-specific, i.e., sound-meaning associations can be arbitrary, following our redefinition of arbitrariness. It shows that double articulation and arbitrariness are fundamentally distinct, although related issues, and that there are two types of sound-meaning associations.
- Sounds can also be associated to meanings in ways that are determined, motivated, predictable, and for that reason are found cross-linguistically. This completes and nuances Saussure's notion of arbitrariness. We propose to call this notion *sound motivation*.

To conclude our discussion, the tendency of English and French swear words to contain unsonorous consonants (RQ1), and the tendency of speakers to use unsonorous consonants for fictional or experimental swear words is the sign of an unconscious sound-meaning association (RQ2a). The meaning involved in that association is a meaning that is common to swear words, and we argue that

it can be either "violation of hearer's personal space", i.e., (positive) familiarity or (negative) aggression depending on the context, or "taboo-breaking" (RQ2b). Such sound-meaning associations do not actually contradict Saussure's arbitrariness and Martinet's double articulation, as long as one accepts that they do not describe the only possible kind of meaning-making in language (RQ3). Our own study and some of the other sound-meaning associations describe sound-meaning associations that may or may not be motivated, e.g., may or may not be iconic. Because of that, they illustrate the need for a more specific definition of arbitrariness compared to Saussure's original understanding (RQ3).

Chapter 5. Conclusion

5.1 Summary

Why are swear words swear words and why are they taboo while their near-synonyms are not? This is the general, overarching research question that this PhD thesis has sought to address. Drawing on previous research on swear words and sound symbolism allowed us to define more specific research questions concerning one factor, among many, that may have contributed to the interdiction of English and French swear words:

1. Is there a tendency for swear words to be closed monosyllabics, or to contain more unsonorous consonants?
2. (a) Does the tendency correspond to a cognitively real, unconscious form-meaning pattern for swear words? (b) If so, what is this meaning?
3. If such form-meaning pairings exist, what are the theoretical implications for the (generally accepted) Saussurian idea that the linguistic form (of words) is arbitrary, and for the (equally accepted) idea that phonemes are meaningless?

In order to identify what meaning is involved in this potential unconscious sound-meaning association (the second part of RQ2), we first need to define what swear words are, for which we proposed four criteria to help identify

swear words in the discussion of our theoretical framework. First, swear words are taboo words, typically forbidden to or in front of children; second, they express and trigger powerful emotions, either positive or negative; third, they belong to a prototype category, usually named with an everyday term like English *swear words* or French *gros mots* ("big words"/"swear words"); and fourth, the taboo on swear words is an autonomous taboo that applies to the words themselves, and the emotional meaning these specific words carry, irrespective of their referential meanings. In other words, the meaning of swear words has to be non-referential, emotional instead. It can also be either positively or negatively valenced depending on the context, as a speaker might create a positive emotional effect by using swear words as positive intensifiers or exclamations, typically among friends.

Our Real-life swear words study allowed us to answer the first research question (RQ1): in English and French, swear words show a statistically significant tendency to contain more unsonorous consonants, compared to the regular lexicon. We counted for the proportion of words containing at least one of the least sonorous consonants – i.e., plosives, voiceless fricatives, affricates – out of three phonemes (0.33 unsonority density). The proportion of words with that minimal density is significantly higher among swear words (84.5% for English, 62.8% for French) compared to the regular lexicon (60.4% for English, 46.5% for French). The data collected on French also contribute to our definition of what swear words are, and consequently what their common specific meaning might be (RQ2b). More precisely, they show that it is possible and even frequent in French for a swear word to have lost any reference to a taboo subject like religion, sexuality, or body waste, and remain a swear word.

Our Fictional swear words study gave a first confirmation that the unsonorous pattern among swear words seems to be exploited by authors inventing swear words (RQ2a). Instead of using real-life swear words in their works, authors can recycle existing words or craft new words to serve as swear words in their fictional universes. We used English fictional swear words from a variety of works by different authors as our data. Compared to the regular lexicon of English and just like real-life swear words, they tend to contain more often one unsonorous consonant out of three phonemes. The same is true for fictional swear words from a French-speaking comic book series who makes frequent uses of obscure words to serve as swear words uttered, although the tendency was not significant in that case. This second study had limitations: the collection of English swear words was not systematic but an informal collection by online contributors, the words in both languages are creations or selections by authors of fiction only, and the French words were selected by only one author. The tendencies in their fictional productions might not be representative of sound-meaning patterns in the minds of other speakers.

To find further confirmation that the sound association is cognitively real, we set up an experimental study where we asked speakers to invent alien swear words. This study confirmed that not just authors but also other speakers are biased towards unsonorous consonants when they are asked to invent swear words. The Experimental swear words study illustrates that when they playfully create words from scratch for a hypothetical alien language, speakers will indeed use more unsonorous consonants for swear words but fewer unsonorous consonants for non-swearing meanings like “hello”, “goodbye”, “please”, or “thanks”. Importantly, for both English and French, neither these more neutral words nor the swear words resemble their real-life equivalents.

This suggests that English and French participants in our study did not build their lexical creations by reusing existing words but seem to be exploiting an unconscious association between swear words and unsonorous consonants, which they do not exploit for creating non-swear words.

In sum, our three case studies provide further evidence for the existence of probabilistic sound-meaning pairings in language already described in the literature. Based on these considerations, we suggest two plausible meanings of swear words that would be carried by the sound symbolism. The first candidate, inspired by a suggestion in Haiman (2018), where we paraphrase the meaning of swear words as a "violation of hearer's space" (ibid.), i.e., familiarity or aggression depending on the context. The second plausible meaning we suggested is that of taboo-breaking. The two meanings are not mutually exclusive, and both can actually be at play simultaneously in a given usage event.

Our findings are sufficiently reliable to propose that there is an unconscious sound-meaning association for swear words in the minds of English and French speakers. However, that does not allow us to conclude that this sound-meaning pattern is motivated, i.e., non-arbitrary, in which case it should be found across unrelated languages. That crucial distinction between unmotivated language-specific associations, and motivated cross-linguistic associations has to be made and taken into account when discussing sound-meaning associations. For that reason, we propose to redefine Saussure's arbitrariness of the linguistic sign more as the notion that there is no deterministic mechanism between a form (e.g., a phoneme or group of phonemes) and its meaning or the other way around (i.e., not necessarily between a word and its meaning, contrary to Saussure's thinking). So one

cannot predict the meaning of a given form, or what form corresponds to a given meaning. This redefined notion may apply not just to words or morphemes, but to any unmotivated form-meaning pairing, including sound-meaning associations. Defined as such, arbitrariness is fundamentally distinct from, and does not logically entail Martinet's double articulation, also known as duality of patterning, which posits that phonemes are meaningless units. However, as we have argued, these two classical tenets of linguistics do not actually invalidate the existence of unconscious, non-compositional, probabilistic sound-meaning patterns, as in the case of swear words.

In particular for double articulation, it is true that phonemes have to be multi-purpose units and that a given phoneme does not have a meaning out of context. However, in the reverse direction, it is not excluded that one given meaning can correspond to one (group of) phoneme(s). That way, when they express meaning, speakers can attribute (phonemic) form to a given meaning following specific sound-meaning correspondences. This is what our experimental study revealed: speakers tended to use more unsonorous consonants when they expressed the meaning of swear words. Sound-meaning associations are not a one-to-one form-to-meaning correspondence system, but a one-to-many system. From the other side of a communicative event, when addressees interpret unconsciously the meanings expressed by the speaker, only the communicative context, and the unconscious knowledge of existing patterns, can help interpret the otherwise extremely ambiguous forms that phonemes are. In other words, sound-meaning associations are – and can only be – probabilistic, contextual, and dependent on the meaning expressed.

The same asymmetry between forms and meanings, creating ambiguity that can only be partly resolved in context, applies for cases of

non-arbitrary, i.e., motivated phoneme-meaning associations, like iconic pairings for example: one given phoneme is not associated iconically with one meaning irrespective of context, otherwise there would exist only one human language on the planet. Yet it is possible to associate one given meaning with one (kind of) phoneme(s) along iconic correspondences – as in the case of the *bouba-kiki* phenomenon, or the cross-linguistic correspondence of close vowels with "small" and open vowels with "big", following Ohala's (1994, 1984) frequency code. One given meaning is iconically associated with one (group of) phoneme(s), not the other way around. Just like unmotivated sound-meaning correspondences, motivated sound-meaning correspondences are meaning-dependent and contextual – see Winter (2021) on the fact that iconicity is contextual. Saussure's arbitrariness of the sign and Martinet's double articulation do not invalidate the existence of such probabilistic, contextual associations dependent on the meaning.

In an attempt to clarify the different distinctions, we propose the term *sound motivation* to describe motivated form-meaning associations, and *sound symbolism* to describe sound-meaning associations which may or may not be motivated, depending on the specific association under study. The distinction brings more awareness to the existence of unmotivated language-specific sound-meaning pairings. It also allows researchers to discuss and accumulate evidence for the existence of a sound-meaning pattern in a given (group of related) language(s) without making unnecessary claims about whether the pattern is or is not motivated and thus found in unrelated languages. A sound-meaning association is worth studying and a useful piece of linguistic knowledge, even if in the end it turns out to be language-specific. Even if the sound-meaning pattern we investigated in this thesis is specific to

English and French, or some group of languages they belong to, it is part of the unconscious knowledge of English and French speakers and therefore can be studied.

5.2 Future research

While our case studies have confirmed sound-meaning patterns for swear words for English and French, there are still some limitations to our study which could be refined in many ways.

First of all, our findings indicate that the group of sounds involved in the sound-meaning association are plosives, voiceless fricatives, and affricates. To keep things feasible within the framework of this thesis, we have only looked at these three consonant groups. However, it is not excluded that other relatively unsonorous consonants, like voiced fricatives, are slightly more frequent than the more sonorous consonants, like nasals or laterals, which is something that our statistical analysis has not yet looked at. Similarly, we have not looked at vowel sounds, where it might be that short vowels are significantly more frequent than long vowels. Bringing these dimensions into the analysis would be a relevant refinement, since it might indicate that the larger feature of unsonority itself is associated with swear words, not just the three most unsonorous categories of plosives, voiceless fricatives, and affricates that we have looked at. Such a more elaborate analysis including voiced fricatives and vowels would be particularly interesting as since our findings seem to contradict Ohala's (1984; 1994) frequency code, which predicts that all voiced obstruents would be more frequent in aggressive words, which arguably

include swear words – although this is not our position, as already discussed. Moreover, the tendency itself might play out differently depending on the degree of sonority, creating a hierarchy within the tendency: for example, the least sonorous consonants, i.e., voiceless plosives may be more frequent in swear words compared to voiceless fricatives, which in turn are more frequent than voiced plosives, which in turn are more frequent than voiced fricatives, etc.

Our experimental study where participants had to invent alien swear words study also opens up perspectives for research on creativity. In our study, we asked speakers to follow their creative intuition to craft entirely new words. One can wonder for example to what extent participants in our study followed the phonotactics of their own native language, even though they were trying to create alien words. Moreover, it is unusual to ask speakers to invent alien words, since most linguistic studies of this kind, like those on the *bouba–kiki* phenomenon, ask speakers to evaluate nonsense words created by the experimenter. Yet this experimental protocol lead to relevant findings, which suggests that this could be extended to the study of sound symbolism in other languages and for other domains.

Another interesting case for further research is whether euphemisms for swear words, such as *fudge* or *shoot* follow the same unsonority tendency. Such euphemisms, sometimes also called "minced oaths" have an ambivalent status: they are not swear words, but they replace an existing swear word which is most likely activated in the speakers' and hearers' minds. Otherwise, the euphemism would not be recognised as such. Given their ambivalent status, it would be interesting to see if minced oaths follow the same unsonority-swearing tendency or not – one study in Lev-Ari and McKay (2022)

suggests that they do not. Their form is also paradoxical: they must be different from, but very close to the swear word they replace, so that speakers understand what swear word is replaced. *fudge* /ˈfʌdʒ/ is only one phoneme away from *fuck* /ˈfʌk/, same for *shoot* /ˈshuːt/ compared to *shit* /ˈshɪt/. With those two examples, we can see that /dʒ/ is more sonorous than /k/ (/ˈfʌdʒ/ ← /ˈfʌk/) and /ɪ/ is a short vowel and thus arguably more sonorous than a long vowel like /uː/ (/ˈshuːt/ ← /ˈshɪt/). We can hypothesise that minced oaths tend to be near-identical to the swear words they replace, but the small change they bring consists in adding sonority, i.e., in making the word sound less "swearish", given the unsonority-swearing tendency. A more exhaustive, systematic study of minced oaths would allow to confirm or disconfirm this provisional hypothesis.

Moreover, it would be interesting to see if the tendency also exists in the larger categories that swear words belong to, like informal words or words used to express strong emotions, like non-swearing interjections. If the same tendency is observed here, we could refine or even reconsider the definition of the meaning that we have suggested for this sound-meaning association.

Our data also showed how the French *gros mots* ("big words" / "(rude) swear words") category does not include slurs unrelated to sexuality. This illustrates how the taboo on *gros mots* or (rude) swear words is not the same as the taboo on slurs, even though sexual slurs like *fag* or *whore* form an overlap between these two taboos and the two categories they create. This overlap creates ambivalence: *gros mots* are relatively harmless compared to slurs and are used cathartically, while slurs express bigotry. One can wonder if in the case of sexual slurs, the fact that they are also used cathartically can make it harder for speakers to understand the harm they do, and to stop using them.

Considerations on the unconscious interpretation and emotional impact of sounds has been common in the study of art and literature. It is taken as a given that poets and artists would use sounds to convey meaning consciously or unconsciously with e.g., alliterations or iconic correspondences. However, our three case studies, as well as some others found in the literature, show that speakers follow such tendencies not just when they are making or enjoying art. Those unconscious sound-meaning pairings can even have a lasting impact on conscious linguistic conventions, like the interdiction of swear words.

Primary sources

ATILF - CNRS & Université de Lorraine. (2002). TLFi : Trésor de la Langue Française informatisé. http://www.atilf.fr/tlfi.

Benoit, C. (2013). Le Soldat et la Putain. Editions Pierre de Taillac.

Benioff, D., & Weiss, D. (2011–2019). Game of Thrones. HBO.

Chautard, E. (2013 [1931]). La Vie étrange de l'argot. Bartillat.

Chollet, M. (2012, June). La maman et la putain sont de retour. Le Monde diplomatique. https://www.monde-diplomatique.fr/2012/06/CHOLLET/47883

Coppola, F. F. (Director). (1992). Bram Stoker's Dracula [Film]. Columbia Pictures; American Zoetrope.

Glowager, M. (Director). (2011). Whores' Glory [Film]. Kino Lorber.

Eustache, J. (Director). (1973). La Maman et la Putain. Elite Films; Ciné Qua Non; Les Films du Losange; Simar Films; V.M. Productions.

Institut National de l'Audiovisuel. (N.d.) Hergé à propos du capitaine Haddock. https://fresques.ina.fr/europe-des-cultures-fr/fiche-media/Europe00025/herge-a-propos-du-capitaine-haddock.html.

Larousse (2020). Dictionnaire français-anglais. https://www.larousse.fr/dictionnaires/francais-anglais/.

Late Night with Seth Meyers (2015, April 7). Seth Brings Jon Snow to a Dinner Party - Late Night with Seth Meyers [video]. Youtube. https://www.youtube.com/watch?v=BabsgCQhpu4

Minier, B. (2015). Une putain d'histoire. XO Editions.

New, B. & Pallier, C. (2019.) Lexique. http://www.lexique.org/.

Oxford University Press. 2020. OED Online. http://www.oed.com/.

Philip, B. (2018, February, 20). Bangkok croule sous les lupanars. Le Monde. https://www.lemonde.fr/asie-pacifique/article/2018/02/20/bangkok-croule-sous-les-lupanars_5259504_3216.html

Première (2015, April 8). Game of Thrones : quand John Snow s'invite à un dîner et pourrit l'ambiance. https://www.premiere.fr/Series/Game-of-Thrones-quand-John-Snow-s-invite-a-un-diner-et-pourrit-l-ambiance

Radio France (2023, September 19). Pourquoi la prostitution était-elle largement acceptée au Moyen Âge ? https://www.radiofrance.fr/franceculture/podcasts/le-pourquoi-du-comment-histoire/pourquoi-la-prostitution-etait-elle-largement-acceptee-au-moyen-age-2819752

Tarantino, Q. (Director) (1994). Pulp Fiction [Film]. Miramax.

University of Western Australia, School of Psychological Science. (1997). MRC Psycholinguistic Database. http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm.

Valeix, V. (2023). La prostitution sous le Premier Empire. Histoire Magazine. https://histoiremagazine.fr/la-prostitution-sous-le-premier-empire/

Wikipedia, (2004). Vocabulaire du Capitaine Haddock. https://fr.wikipedia.org/wiki/Vocabulaire_du_capitaine_Haddock.

Wiktionary, (2009). Appendix: Fictional English swear words. https://en.wiktionary.org/wiki/Appendix:Fictional_English_curse_words

Zidi, C. (Director) (1999). Asterix and Obelix take on Caesar [Film]. Pathé.

Bibliography

Aryani A., Conrad, M., Schmidtke, D., Jacobs, A. (2018). Why 'piss' is ruder than 'pee'? The role of sound in affective meaning making. PLoS ONE 13 (6): e0198430. https://doi.org/10.1371/journal. pone.0198430

Bergen, B. K., (2016a). What the F: What Swearing Reveals about Our Language, Our Brains, and Ourselves. Basic Books.

Bergen, B. K., in Talks at Google (2016b, October 31). Benjamin Bergen: 'What the F' | Talks at Google. https://www.youtube.com/watch?v=rQws2yRfJ7c

Bergen, B. K. (2004). The psychological reality of phonaesthemes. Language 80(2). 290–311.

Broadcasting Standards Authority (2010). What not to swear: The acceptability of words in broadcasting. https://www.bsa.govt.nz/oldsite/assets/Research/What-Not-to-Swear-Full-BSA2010.pdf.

Childs, G. Tucker. (2001). Research on ideophones, whither hence? The need for a social theory of ideophones. In Voeltz, F.K. Erhard & Christa Kilian-Hatz (eds.), Ideophones, 63–70. John Benjamins Publishing.

Blasi, D. E., Wichman, S., Hammarström, H., Peter F. Stadler, P. F., Christiansen, M. H., (2016). Sound–meaning association biases evidenced across thousands of languages. PNAS, 113 (39) 10818–10823 https://doi.org/10.1073/pnas.1605782113

Bowers, J. S., & Pleydell-Pearce, C. W. (2011). Swearing, Euphemisms, and Linguistic Relativity. PLOS ONE

https://doi.org/10.1371/journal.pone.0022341

Chiang, W., & Schlatter, T., (ms.) "Sound Symbolism in Swearing: The Phylogenetic Origin of Profanity" https://rpubs.com/ThomasSchlatter/922544

Culpeper, J. (2011). Impoliteness: Using Language to Cause Offence. Cambridge University Press

Culpeper, J., & Kytö, M. (2010). Pragmatic noise: meaning and development, in Culpeper, J. & Kytö, M. (2010). Early Modern English Dialogues: Spoken Interaction in Writing, 284–305. Cambridge University Press.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., Monaghan, P., (2015). Arbitrariness, Iconicity, and Systematicity in Language. Trends in Cognitive Sciences 19(10). 603–615

Drummond, Rob. (2020). Teenage swearing in the UK. English World-Wide 41(1). 59–88. [pre-print].

Egbokhare, F. O. (2001). Phonosemantic correspondences in Emai attributive ideophones. In F. K. Erhard Voeltz & Christa Kilian-Hatz (eds.), Ideophones, 87–96. John Benjamins Publishing. https://doi.org/10.1075/tsl.44.08egb

Finkelstein, S. R., (2018). Swearing as emotion acts. In Pizarro Pedraza, A., (ed.) Linguistic Taboo Revisited: Novel Insights from Cognitive Perspectives. De Gruyter Mouton.

Goldberg, A., (2006). Constructions at Work. Oxford University Press.

Greenberg, J. H., (1987). Languages in the Americas. Stanford University Press.

Gutiérrez, E. D., Levy, R. & Bergen, B. K., (2016). Finding non-arbitrary form-meaning systematicity using string-metric learning for kernel regression. In Erk, K. & Smith, N. A. Proceedings of the 54th annual

meeting of the Association for Computational Linguistics 1, 2379–2388.

Haiman, J., (2018). Ideophones and the Evolution of Language. Cambridge University Press.

Halliday, M.A.K. 1985. An Introduction to Functional Grammar. Arnold.

Hughes, G., (2006). An Encyclopedia of Swearing. Armonk, Sharpe.

Ibarretxe-Antuñano, I. (2024). Onomatopoeia in Basque. In Körtvélyessy, L. & Štekauer, P. (eds), Onomatopoeia in the World's Languages: A Comparative Handbook, 489–499. De Gruyter Mouton. https://doi.org/10.1515/9783111053226

Ibarretxe-Antuñano, I. (2023). Vindicating the role of ideophones as a typological feature of Basque. In J. P. Williams (ed.), Expressivity in European Languages 313–334. Cambridge University Press.

Janschewitz, K. (2008). Taboo, emotionally valenced, and emotionally neutral word norms. Behavior Research Methods, 40 (4), 1065–1074.

Jay, T. (2018). Swearing, moral order, and online communication. Journal of Language Aggression and Conflict 6 (1), 107–126 https://doi.org/10.1075/jlac.00005.jay

Jay, T. & Janschewitz, K. (2008). The pragmatics of swearing. Journal of Politeness Research, 4(2), 267–288. https://doi.org/10.1515/JPLR.2008.013

Joo, I. (2023). The sound symbolism of food: the frequency of initial /PA-/ in words for (staple) food. Linguistics, 61 (1), 33–46. https://doi.org/10.1515/ling-2021-0127

Jumanto, Sulistyorini H., (2019). The pragmatics of swearing: how it contextually counts. Proceedings of the English Language and Literature International Conference, 3, 324–333 https://jurnal.unimus.ac.id/index.php/ellic/article/view/4722

Kakehi, H., Tamori, I., Schourup, S. (2011). Dictionary of Iconic Expressions in Japanese. De Gruyter Mouton.

Kamp, V., Pélissier, M., Ferragne, M., (2024, May 15). Exploring the Foreign Language Effect: A Study on Emotional Responses to Taboo Words [Conference presentation]. AFLiCo 9, Lyon, France.

Kosugi, H. (2010). Performative power of language: Japanese and swearing. J Teknosastik, 8, 30–37.

Lev-Ari, S., & McKay, R., (2022). "The sound of swearing: Are there universal patterns in profanity?" Psychonomic Bulletin & Review 30. 1103–1114. https://doi.org/10.3758/s13423-022-02202-0

Ljung, M. (2011). Swearing: A Cross-Cultural Linguistic Study. Palgrave Macmillan.

Love, R. (2021). Swearing in informal spoken English: 1990s–2010s. Text & Talk, 41(5–6), 739–762. https://doi.org/10.1515/text-2020-0051

Martinet, A., (2008 [1960]). Éléments de linguistique générale, Armand Colin.

Martinet, A., (1957). "Arbitraire linguistique et double articulation" Cahiers Ferdinand de Saussure 15, 105–116.

Mauro (de), T., (2005 [1967]) in Saussure (de), F., (2005 [1916]). Cours de linguistique générale, Mauro (de), T., (éd.), Payot

McEnery, T. (2006). Swearing in English Bad Language, Purity and Power from 1586 to the Present. Routledge.

Millwood-Hargrave, A. (2000). *Delete expletives?* Ofcom. http://jastoy.plus.com/pdf/delete_expletives.pdf.

Owren, M.J. & Rendall , D. (1997). An affect-conditioning model of nonhuman primate vocal signaling. Perspectives in Ethology, 12. Plenum Press.

Owren, M.J., Rendall, D. (2001). Sound on the rebound: Bringing form and

function back to the forefront in understanding nonhuman primate vocal signaling. Evolutionary Anthropology, 10, 58–71.

Ohala, J. J. (1994). The frequency codes underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (eds.), Sound symbolism. Cambridge University Press. 325–347.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. Phonetica 41, 1–16.

Parker, S., (2008). Sound level protrusions as physical correlates of sonority Journal of Phonetics 36 (1). 55–90. https://doi.org/10.1016/j.wocn.2007.09.003

Pinker, S. (2007). The Stuff of Thought: Language as a Window Into Human Nature, Viking.

Reilly, J., Kelly, A., Zuckerman, B.M. et al. (2020). Building the perfect curse word: A psycholinguistic investigation of the form and meaning of taboo words. Psychonomic Bulletin Review 27, 139–148. https://doi.org/10.3758/s13423-019-01685-8

Reilly, J., Biun, D., Cowles, H. W., Peelle J., (2008). Where did words come from? A linking theory of sound symbolism and natural language evolution. Nature Precedings. http://precedings.nature.com/documents/2369/version/1

Robertson, O., Robinson, S. J., Stephens, R. (2017). Swearing as a response to pain: A cross-cultural comparison of British and Japanese participants. Scandinavian journal of pain, 17, 267–272. https://doi.org/10.1016/j.sjpain.2017.07.014

Rouayrenc, C. (1996). *Les gros mots.* Paris: Presses Universitaires de France.

Ruiz Martínez, D. (2019). Dictionaries of Japanese mimetic words: defining the mimetic category by the selection of lexical items. Lexicography ASIALEX

6, 1–19. https://doi.org/10.1007/s40607-019-00055-9

Saussure (de), F., (2005 [1916]). Cours de linguistique générale, Mauro (de), T., (éd.), Payot

Sidhu, D. M., (2019). Explorations of Sound Symbolism and Iconicity. Unpublished doctoral thesis, University of Calgary, Calgary, AB, Canada. Available at: http://hdl.handle.net/1880/111016

Sidhu, D. M., Deschamps, K., Bourdage, J. S., Pexman, P. M. (2019). Does the name say it all? Investigating phoneme-personality sound symbolism in first names. Journal of Experimental Psychology: General, 148(9), 1595–1614. https://doi.org/10.1037/xge0000662

Stapleton, K., Beers Fägersten, K., Stephens, R., Loveday, C. (2022). The power of swearing: What we know and what we don't", Lingua 277 (103406), https://doi.org/10.1016/j.lingua.2022.103406.

Stephens, R., Spierer, D. K., Katheis E., (2017). Effect of swearing on strength and power performance. Psychology of Sport and Exercise, 35, 111–117. https://doi.org/10.1016/j.psychsport.2017.11.014.

Stephens, R. and Umland, C. (2011). Swearing as a Response to Pain: Effect of Daily Swearing Frequency. The Journal of Pain 12 (12), 1274–1281. https://doi.org/10.1016/j.jpain.2011.09.004

Stephens, R., Atkins, J., and Kingston A. (2009). Swearing as a response to pain. NeuroReport, 20 (12), 1056–1060. doi: 10.1097/WNR.0b013e32832e64b1

Sulpizio, S. et al (2024). Taboo language across the globe: A multi-lab study. https://doi.org/10.31234/osf.io/r3wzq

Vallery, R., (2019). The sound of taboo: Exploring phonopragmatic associations of swear words in English and in French. Unpublished MA thesis, University de Lille, France

Van Lancker, D. & Cummings, J. (1999). Expletives: Neurolinguistic and neurobehavioral perspectives on swearing. Brain research reviews, 31 (1), 83–104.

Voeltz, F. K. Erhard & Kilian-Hatz C. (2001). Introduction. In F. K. Erhard Voeltz & Christa Kilian-Hatz (eds.), Ideophones, 1–8. John Benjamins Publishing. https://doi.org/10.1075/tsl.44.02voe

Wajnryb, R. (2005). Expletive deleted: A good look at bad language. Free Press

Ward, N. (2006). Non-lexical conversational sounds in American English. Pragmatics & Cognition, 14(1), 129–182.

Wharton, T. (2003). Interjections, language and the showing-saying continuum. Pragmatics and Cognition, 11(1), 39–91.

Wierzbicka, Anna. 1988. The Semantics of Grammar. Studies in Language Companion Series, 8. John Benjamins.

Winter, B. in ERC project ABSTRACTION (2023, February 23). There is nothing arbitrary about the linguistic sign [video]. Youtube. https://www.youtube.com/watch?v=Ol0YPGHQhvs

Winter, B. in ERC project Abralin (2021, May 5). Bodo Winter [video]. Youtube. https://www.youtube.com/watch?v=R1ETw21oCGE

Yardy, B. J., (2010). "Sound symbolism, sonority, and swearing: an affect induction perspective", PhD thesis, University of Lethbridge, AB, Canada. https://opus.uleth.ca/handle/10133/2556

Resumé substantiel en français

1. Introduction

Pourquoi les gros mots sont-ils des gros mots ? Peu de pratiques langagières semblent aussi évidentes aux locuteurs et locutrices que l'interdiction des gros mots. Pour beaucoup d'entre nous, les gros mots sont interdits parce qu'ils sont interdits, et la règle va de soi. Même les personnes les plus tolérantes n'acceptent les gros mots que partiellement : on peut bien les utiliser, mais sans en abuser, au risque de paraître, dit-on, vulgaire. Pourtant, cette interdiction ne vient pas de nulle part. Elle est notamment le produit de l'interdiction d'aborder des sujets tabous, c'est-à-dire socialement interdits : à travers les langues et cultures où les gros mots existent, on constate que ces gros mots ont, ou ont eu dans le passé, des sens appartenant à des domaines sémantiques tabous comme la religion, la sexualité, ou les déchets du corps humain (Stapleton et al. 2022 : 2 ; Bergen 2016a : 12–39). Cette remarque en appelle immédiatement une autre : si c'est le domaine sémantique tabou qui a créé cette interdiction, alors pourquoi certains mots sont-ils des gros mots, et pas leurs quasi-synonymes ? Pour prendre des exemples de gros mots en anglais, pourquoi *prick* ("bite") est-il un gros mot et pas *penis* ("pénis"), alors qu'il réfère au même concept appartenant au domaine tabou de la sexualité (Ljung 2011: viii) ? Pourquoi *shit* ("merde") est-il un gros mot et pas *excrement* ("excrément") ou *stool* ("selles"), alors qu'il réfère au même concept du domaine tabou des déchets du corps humain (Bergen 2016a: 14–15) ? La seule référence à un domain sémantique tabou ne peut pas expliquer pourquoi un mot donné a été

désigné collectivement comme objet d'interdiction par une communauté linguistique. D'autres paramètres entrent nécessairement en compte. Certains chercheurs et chercheuses suggèrent que la présence ou l'absence de certains phonèmes peut avoir joué un rôle dans ce processus d'interdiction. Il a notamment été observé empiriquement que les gros mots de l'anglais ont une tendance à être des monosyllabiques fermés, finissant par une consonne, et que cette tendance pourrait refléter un association forme-sens inconsciente dans l'esprit des locuteurs et locutrices anglophones (Bergen 2016a: 49–64). Autrement dit, les gros mots "sonnent" d'une certaine manière, et un mot qui "sonne comme un gros mot" est un meilleur candidat, toutes choses égales par ailleurs, pour devenir à terme un gros mot, car il correspond à cette consonance particulière. Une proposition alternative faite par Yardy (2010) d'après des données de l'anglais est que les gros mots de nombreuses langues non-apparentées, à cause d'une motivation iconique où la non-sonorité est interprétée inconsciemment comme expression d'une attitude hostile, devraient avoir tendance à contenir plus de consonnes relativement non-sonores sur l'échelle de sonority de Parker (2008), c'est-à-dire, plus d'occlusives (/p/, /t/, /k/, /b/, /d/, et /g/), plus de fricatives sourdes (/f/, /θ/, /s/, /ʃ/, et /h/), et plus d'affriquées sourdes (/tʃ/). D'autres observations intuitives (Wajnryb 2005 : 205–210 ; Hughes 2006 : 343 ; Pinker 2007) ou basées sur des données quantitatives (Lev-Ari & McKay 2022 ; Chiang & Schlatter, ms.) proposent différentes formes, c'est-à-dire différents phonèmes ou groupes de phonèmes, mais ces observations et propositions éparses correspondent assez largement à la proposition de Yardy (2010) sur la non-sonorité. Ces propositions divergent sur le sens à attribuer aux gros mots, et qui serait présent dans cette association, mais s'accordent sur le point que ce sens commun et spécifique aux gros mots

n'est pas de type référentiel (ou dénotatif, ou véric onditionnel, pour utiliser d'autres termes partiellement équivalents), mais plutôt de type émotionnel et contextuel.

Le Chapitre 2 présente un état de l'art et une présentation du cadre théorique. L'examen de la littérature scientifique existante se concentre sur trois axes principaux: d'abord les études sur les gros mots, qui permettent d'élaborer une première définition provisoire de ce que sont les gros mots, et de contribuer à l'une de nos questions de recherche; ensuite, les études sur les associations son-sens dans les gros mots, qui permettent d'élaborer nos deux premières questions de recherche; et les études sur les associations son-sens dans la langue, un sujet couramment appelé phonosymbolisme et qui mène à notre troisième question de recherche. Ces trois questions de recherche sur les gros mots de l'anglais et du français sont les suivantes :

1. Y a-t-il une tendance dans les gros mots à être plus souvent des monosyllabiques fermés, ou à contenir des consonnes non-sonores, comparés aux autres mots du lexique ?
2. (a) Si cette tendance est statistiquement établie dans les gros mots, correspond-elle alors à un schéma, une association forme-sens cognitivement réelle, inconsciente ? (b) Si oui, quel est le sens impliqué dans cette association ?
3. Si de telles associations forme-sens existent, quelles sont les implications théoriques pour l'idée saussurienne (généralement acceptée) que la forme linguistique d'un mot est arbitraire, et pour l'idée (tout aussi acceptée) que les phonèmes n'ont pas de sens ?

Afin de poser des bases théoriques nécessaires pour des études empiriques

visant à répondre à nos premières questions de recherche (QR1 et QR2a), et poser un cadre à notre réponse sur le sens des gros mots (QR2b), la seconde partie du Chapitre 2 offre le développement d'un cadre théorique sur la définition des gros mots, à partir de la littérature existante mais également de notre propre réflexion.

Le Chapitre 3 donne la méthode, les données et la description des résultats de trois études empiriques visant à répondre à nos premières questions de recherche (QR1 et QR2a). Une première étude a consisté à vérifier les propositions de Bergen (2016a) et Yardy (2010) sur les tendances statistiques des gros mots à correspondre à une certaine forme (QR1) en utilisant des données disponibles sur les gros mots de l'anglais d'une part, et d'autre part en collectant empiriquement les gros mots du français par un questionnaire donné à des locuteurs et locutrices francophones, ce qui à notre connaissance n'avait été jamais fait avant la période de collecte (2018). Les données récupérées sur le français permettent également de confirmer la définition théorique proposée pour les gros mots, et de répondre en partie à la question de recherche sur le sens des gros mots (QR2b). Les deuxième et troisième études tentent d'établir si la tendance statistique détectée lors de la première étude correspond à un schéma inconscient présent dans l'esprit des locuteurs et locutrices (QR2a). La deuxième étude consiste à analyser les gros mots fictionnels issues d'œuvres anglophones et francophones, et vérifier s'ils suivent la même tendance, ce qui suggèrerait que les auteurs et autrices exploitent inconsciemment cette association lorsqu'ils et elles suivent leur intuition créative et inventent des mots, ou réutilisent des mots existants, pour servir de gros mots dans leurs univers de fiction. La troisième étude est une étude expérimentale, consistant à demander par un questionnaire à des participants anglophones et francophones

d'inventer spontanément, ludiquement, des mots d'une langue alien, c'est-à-dire extra-terrestre, comme on en trouverait dans des œuvres de science-fiction. Ce questionnaire demande d'inventer des gros mots, et d'inventer des non-gros mots, pour comparer si les gros mots inventés suivent la même tendance, comparés aux non-gros mots inventés, que les gros mots authentiques, ce qui indiqueraient que les locuteurs et locutrices exploitent inconsciemment ce schéma, cette association forme-sens inconsciente.

Le Chapitre 4 discute les résultats de nos trois études empiriques. La première partie présente les implications logiques de l'existence de cette association forme-sens inconsciente dans les gros mots anglais et français, qui a été détectée dans nos données. Les deuxième et troisième parties montrent comment les données du français de la première étude valident notre cadre théorique présenté au Chapitre 2. La dernière partie du Chapitre 4 est consacrée à répondre, à partir de la littérature existante et de nos données, à notre dernière question de recherche sur les implications théoriques concernant l'arbitraire du signe de Saussure et la double articulation de Martinet (QR3).

2. État de l'art et cadre théorique

Les gros mots sont particulièrement difficiles à définir, d'abord parce que leur catégorisation semble évidente. La catégorie des gros mots n'est pas forcément définie explicitement dans la littérature qui l'aborde, et lorsqu'elle l'est, c'est parfois avec des définitions contradictoires. La seconde difficulté vient de l'expérience subjective et émotionnelle de ce qu'est un gros mot : tous les locuteurs et locutrices ne seront pas forcément d'accord sur le fait que tel mot est un gros mot. La littérature scientifique permet pourtant de dégager deux généralisations simples sur les gros mots. D'abord, ce sont des mots tabous,

autrement dit des mots socialement interdits. Cette interdiction est rarement formelle, mais elle est présente culturellement. La deuxième généralisation, c'est que les gros mots expriment, et provoquent chez les interlocuteurs qui les percoivent, une réaction émotionnelle. Cette réaction émotionnelle est plus forte avec les gros mots issus de la langue première que d'une langue seconde, et plus forte pour le gros mot lui-même que pour ses euphémismes. Elle peut se mesurer à partir de réactions physiques comme la sudation et l'augmentation du rythme cardiaque, ou des effets analgésiques. Elle est aussi plus forte que d'autres mots pourtant émotionnellement chargés. Cette réaction émotionnelle a des effets sur la concentration et la reconnaissance des gros mots : les gros mots sont plus facilement retenus et identifiés que d'autres items lexicaux. Elle explique aussi pourquoi les gros mots font partie des items linguistiques préservés chez certains patients aphasiques, ou chez les patients atteint du syndrome de Tourette présentant le symptôme de coprolalie, une tendance incontrôlée à prononcer des gros mots ou autres expressions tabous en contexte. Ces éléments suggèrent que ce que transmettent les gros mots, c'est bien un sens plus social et émotionnel, et non pas un quelconque sens qui serait référentiel. La littérature scientifique propose que ce sens plus social et émotionnel est associé inconsciemment à une forme. Parmi les études fondées sur des observations statistiques, Yardy (2010) propose que cette forme consiste en la présence de consonnes relativement non-sonores (les occlusives, fricatives non-voisées et affriquées), et que cette association est iconique et devrait donc se retrouver dans des langues non-apparentées et géographiquement distantes. Bergen (2016a: 49–64) propose que, en anglais tout du moins, cette forme consiste en des monosyllabiques fermés. Lev-Ari & McKay (2022) proposent que cette forme consiste en l'absence d'approximants. Chiang & Schlatter (ms.)

proposent que cette forme consiste en la présence de sons imitant des expressions hostiles chez les primates. Ces observations empiriques et statistiques, combinées à des remarques plus intuitives formulées auparavant (Wajnryb 2005, Hughes 2006, et Pinker 2007) semblent toutes recouper partiellement la proposition autour de la non-sonorité avancée par Yardy (2010). La plus éloignée est celle de Bergen (2016a: 49–64) concernant la monosyllabicité fermée, que nous vérifions dans notre première étude empirique. De ces éléments découle nos premières questions de recherche (QR1 et QR2). La dernière partie du Chapitre 3 établit un état des lieux des études sur le phonosymbolisme, autrement dit, sur les associations son-sens inconscientes observées empiriquement dans la langue. On observe des associations de ce type non seulement dans des items considérés habituellement comme périphériques dans le lexique, comme les interjections, les idéophones (Ibarretxe-Antuñano 2024: 490, 498; 2023: 324–326; Haiman 2018: 79–88), l'interprétation inconsciente des noms propres (Sidhu, Deschamps, Bourdage, & Pexman 2019: 1610), mais aussi dans des éléments plus habituels du lexique comme les noms communs (Blasi et al. 2016, Joo 2023). Les études sur le phonosymbolisme soulèvent la question de leurs implications pour deux piliers traditionnels de la linguistique: la double articulation de Martinet (2008 [1960]: 37–44, 1957) qui postule que les phonèmes sont des unités dépourvues de sens, et l'arbitraire du signe de Saussure (2005 [1916]:100–101, 155–157) qui postule que le sens d'un mot ne détermine pas sa forme et inversement. Ces considérations permettent d'arriver à notre dernière question de recherche (QR3). La deuxième partie du Chapitre 2 élabore la définition théorique des gros mots suivante : les gros mots sont des mots socialement interdits, particulièrement aux enfants, et plus précisément situés à l'extrêmité du registre

informel; ils expriment et transmettent des émotions fortes, au point de pouvoir provoquer le rire, la gêne, ou des réactions physiques mesurables plus fortes qu'avec d'autres items linguistiques; ils forment une catégorie prototypique dans l'esprit des locuteurs, souvent désignée par un terme commun comme *swear words* en anglais ou *gros mots* en français; enfin, ce n'est pas leur sens référentiel tabou mais le mot lui-même qui est interdit, de sorte qu'il est toujours possible d'exprimer le même sens référentiel avec un mot ou une expression alternative qui n'est pas un gros mot, ainsi la référence à un sujet tabou n'est pas un critère efficace pour détecter les gros mots dans une langue et définir la nature des gros mots dans la langue d'un point de vue synchronique.

3. Études sur les gros mots

Notre première étude empirique consiste d'abord à reproduire l'étude de Bergen (2016a: 49–64) sur les gros mots de l'anglais, en les comparant aux mots les plus fréquents du lexique anglais trouvables sur la base de données en ligne MRC Psycholinguistic Database. L'analyse statistique de ces données montre que la tendance à la monosyllabicité fermée proposée par Bergen (2016a) est en effet observable dans les gros mots de l'anglais. Ensuite, nous conduisons l'analyse des gros mots du français, collectés empiriquement par un questionnaire en ligne, pour les comparer avec les mots les plus fréquents du français disponibles sur la base de données en ligne Lexique. La tendance proposée par Bergen ne s'applique pas au français. En revanche, la tendance de non-sonorité proposée par Yardy s'applique également aux gros mots de l'anglais et du français, ce qui répond à notre première question de recherche (QR1). Les données récoltées pour le français permettent aussi de confirmer que

la référence à un sujet tabou comme la religion, la sexualité, ou les déchets du corp humain, ne définit pas synchroniquement les gros mots : de nombreux gros mots très prototypiques du français, comme *bâtard*, *bordel*, *con*, *enfoiré*, *foutre*, ou *putain* ont perdu ou sont en train de perdre toute référence à un domaine sémantique tabou – en l'occurrence, toute référence à la sexualité. Cette définition pose les bases de la discussion d'une autre de nos questions de recherche (QR2b). La deuxième étude tente de répondre à la question de la correspondance de cette tendance à un schéma inconscient chez les locuteurs et locutrices (QR2a). Elle consiste à analyser les mots utilisés comme gros mots dans des univers de fictions issus d'œuvres anglophones et francophones. Pour la fiction anglophone, nous utilisons les données issues d'une page du dictionnaire collaboratif en ligne Wiktionary qui récolte les gros mots utilisés dans une série d'œuvres de différents genres. Pour la fiction francophone, nous utilisons les données récoltées sur une page de l'encyclopédie en ligne Wikipédia sur les gros mots du Capitain Haddock, dans la série de bande-dessinée *Les Aventures de Tintin* publiée entre les années 1930 et les années 1970. Comme pour les gros mots authentiques, nous comparons ces gros mots fictionnels à leurs lexiques respectifs et observons aussi parmi ces mots la tendance à contenir plus de consonnes non-sonores, ce qui tend à confirmer la présence d'une association son-sens inconsciente chez les locuteurs et locutrices créant ou sélectionnant ces gros mots fictionnels – bien que la tendance n'atteigne pas le seuil de significativité pour les gros mots du Capitaine Haddock en français. Pour pallier aux limites de cette étude concentrée exclusivement sur la production des auteurs et autrices de fiction, nous élaborons une dernière étude, expérimentale cette fois-ci, où nous demandons à trente-deux locutrices et locuteurs anglophones et trente-deux locutrices et

locuteurs francophones d'inventer spontanément, ludiquement, des mots qui seraient issus d'une langue alien, c'est-à-dire extra-terrestre comme on en trouve dans les œuvres de science-fiction, à l'aide d'une série de questions thématiques illustrées. Ces productions créatives comprennent une moitié de gros mots et une moitié de non-gros mots afin d'offrir un point de comparaison. Là encore, l'analyse statistique montre une tendance significative des gros mots à contenir plus de consonnes non-sonores que les non-gros mots. Ces résultats suggèrent que les participantes et participants à notre expérience utilisent inconsciemment une association son-sens pour créer des gros mots.

4. Discussion

Le Chapitre 4 offre une discussion des implications de l'existence de cette association son-sens inconsciente dans les gros mots de l'anglais et du français. En particulier, on tente d'expliquer comment cette association a pu intervenir historiquement et biaiser le choix collectif d'interdire tel mot et d'en permettre tel autre, bien que les deux aient le même sens référentiel. Cette association suggère que même des conventions sociolinguistiques très conscientes comme l'interdiction des gros mots peuvent être influencées non seulement par des logiques sociales comme la distinction et les paniques morales (McEnery 2006: 226–227) mais aussi par l'interprétation contextuelle inconsciente des sons dans un mot. Les parties suivantes de la discussion montrent comment les données récoltées pour le français permettent de comprendre les gros mots comme une catégorie fondamentalement détachée de tout sens référentiel, et qui n'englobe pas toutes les insultes qui visent un groupe social dominé (en anglais, cette catégorie d'insutes est appelée *slurs* – elle comprend les insultes racistes, insultes homophobes, etc.). Nous proposons un sens commun à tous

les gros mots dérivé de la littérature, “intrusion dans l’espace personnel de l’interlocuteur ou interlocutrice”, et un sens plus simple d’après notre propre réflexion, qui serait tout simplement la “rupture du tabou”. Enfin, nous développons une réflexion sur les implications des associations son-sens telles que celles proposées dans cette thèse ou décrites dans la littérature existante, pour ces deux piliers classiques de la linguistique que sont la double articulation de Martinet et l’arbitraire du signe de Saussure. Notre point principal est que ces deux notions ne sont en réalité pas incompatibles avec l’existences d’associations son-sens décrites dans la littérature sur le phonosymbolisme, tant qu’on admet que (i) ces deux notions décrivent comment la langue transmet du sens référentiel d’une manière tout à la fois consciente, compositionnelle, et déterministe indépendamment du contexte, et (ii) la langue peut aussi transmettre du sens non-référentiel et/ou de manière inconsciente, et/ou non-compositionnelle, et/ou probabiliste et contextuelle, et c’est typiquement de cette manière-là que les associations phonosymboliques transmettent du sens, qui par ailleurs n’est pas nécessairement référentiel, comme dans l’association que nous étudions sur les gros mot. Nous argumentons aussi pour une redéfinition plus spécifique de l’arbitraire du signe, qui ne se concentre pas sur les mots et leurs sens mais s’applique à toute association forme-sens – que cette forme soit un son ou groupe de sons, un mot, ou une forme plus abstraite et syntaxique. Comprise ainsi, la notion d’arbitraire décrit utilement des associations son-sens qui font partie de la connaissance linguistique inconsciente des locuteurs et locutrices, mais ne sont pas pour autant motivées.

5. Conclusion

Cette thèse tente de contribuer à la recherche sur une question générale, qui est d'expliquer pourquoi les gros mots deviennent des gros mots dans les communautés linguistiques. À partir de l'état de l'art, nous élaborons trois questions de recherche plus spécifiques sur les gros mots de l'anglais et du français :

1. Y a-t-il une tendance dans les gros mots à être plus souvent des monosyllabiques fermés, ou à contenir des consonnes non-sonores, comparés aux autres mots du lexique ?
2. (a) Si cette tendance est statistiquement établie dans les gros mots, correspond-elle alors à un schéma, une association forme-sens cognitivement réelle, inconsciente ? (b) Si oui, quel est le sens impliqué dans cette association ?
3. Si de telles associations forme-sens existent, quelles sont les implications théoriques pour l'idée saussurienne (généralement acceptée) que la forme linguistique d'un mot est arbitraire, et pour l'idée (tout aussi acceptée) que les phonèmes n'ont pas de sens ?

Une première étude empirique sur les gros mots de l'anglais a permis de répondre à la première question de recherche (QR1) : les gros mots de l'anglais et du français tendent à contenir significativement plus de consonnes non-sonores, c'est-à-dire plus d'occlusives, de fricatives sourdes, ou d'affriquées. Notre deuxième étude empirique apporte une première réponse à notre deuxième question de recherche (RQ2a) : les auteurs et autrices de fiction tendent à sélectionner plus de consonnes non-sonores lorsqu'ils suivent leur intuition créative et cherchent des mots pour servir de gros mots dans leurs univers fictionnels. Notre troisième étude dépasse les limites de la précédente

en demandant à des anglophones et francophones d'inventer expérimentalement des mots fictionnels aliens, dont une moitié de gros mots et une moitié de non-gros mots. Ces mots inventés suivent la même tendance : les gros mots inventés lors de cette expérience tendent à contenir plus de consonnes non-sonores que les non-gros mots inventés. L'association inconsciente son-sens expliquant ces résultats doit impliquer un sens émotionnel commun aux gros mots, qui pourrait être "intrusion dans l'espace personnel" ou "rupture du tabou". Ce type d'association son-sens invite à reconsidérer les notions de double articulation et d'arbitraire (RQ3), dont nous considérons qu'elles ne sont pas fondamentalement incompatibles avec le phonosymbolisme. Simplement, la double articulation et l'arbitraire d'une part, et le phonosymbolisme d'autre part décrivent deux façons possibles pour la langue de transmettre du sens, et l'existence d'une façon n'exclut pas l'existence de l'autre. En termes de développements ultérieurs à cette recherche, une analyse statistique plus élaborée permettrait d'affiner les résultats et de détecter des tendances plus systématiques et différenciées selon les types de phonèmes, en particulier si différents types de voyelles sont plus ou moins présentes dans les gros mots. Notre protocole expérimental permet aussi d'étudier à quel point les participantes et participants peuvent être créatifs lorsqu'ils tentent de créer des mots à partir de rien: par exemple, suivent-ils les règles phonotactiques de leur langue ou s'en écartent-ils ? Notre définition des gros mots, qui n'inclut pas toutes les insultes interdites, invite à considérer le chevauchement et la tension qui existe entre le tabou sur ces insultes et le tabou sur les gros mots : le chevauchement entre ces deux catégories ne crée-t-il pas de l'ambiguité et des dilemmes éthiques pour les locuteurs et locutrices qui sont habitués à utiliser la force émotionnelle cathartique de ces gros mots ? Ces considérations invitent

aussi à prendre en compte que l'interprétation inconsciente des sons peut biaiser des conventions sociales fortes comme l'interdiction des gros mots. On peut se demander quelles autres conventions sociales ou sociolinguistiques sont influencées par des biais de cette nature, et notamment les différences de registre.

6. Organisation de la thèse

Le Chapitre 1 de cette thèse présente la question de recherche générale et les questions de recherche spécifiques issues de l'état de l'art, puis décrit brièvement la structure de la thèse.

Le Chapitre 2 est divisé en deux parties, une donnant l'état de l'art et le cadre théorique. L'état de l'art est divisé en trois parties, l'une sur la recherche sur les gros mots, l'une sur la recherche sur des associations phonosymboliques sur les gros mots et le phonosymbolisme, et l'une sur les études sur le phonosymbolisme. Le cadre théorique est élaboré à partir de l'état de l'art et vise à définir les gros mots d'un point de vue translinguistique et synchronique, en vue de répondre à l'une des questions de recherche sur le sens des gros mots impliqué dans l'association son-sens dont nous faisons l'hypothèse.

L'existence statistique de cette association est confirmée dans la première étude du Chapitre 3, qui décrit les méthodes, données, et résultats de trois études empiriques sur les gros mots de l'anglais et du français, comparés au reste du lexique de la langue correspondante. L'existence cognitive de cette association dans l'esprit des locuteurs et locutrices de l'anglais et du français est confirmée dans la deuxième et la troisième étude de ce chapitre. La deuxième porte sur des gros mots fictionnels obtenus sur des ressources ouvertes en ligne,

la troisième porte sur des gros mots crées par des participants lors d'un protocole expérimental que nous avons conduit.

Le Chapitre 4 inclut la discussion de nos résultats et leurs implications pour l'interdiction des gros mots, pour la définition des gros mots d'un point de vue translinguistique et synchronique, et pour les notions de double articulation et d'arbitraire du signe.

Le Chapitre 5 présente la conclusion à cette thèse avec un résumé et des considérations sur les perspectives futures de recherche ouvertes par la présente thèse.