# A Study of the Productivity of Twelve English Onset Phonaesthemes

Presented in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Michael Willett
Cardiff University, 2015

**Acknowledgments**

There are many people to whom I am indebted for their assistance in the production of this work.

To my parents, David and Ellen. What can I say except ‘thanks a million’. For never letting me settle for second best.

To Gerard O'Grady; a top supervisor and a top mate. Thanks for your guidance these many years, and thank you for answering so many emails late at night.

To Joanna Thornborrow, for planting the seeds of inspiration.

To Tom Bartlett, for being the voice of reason. Next time you're in Filling Station in bonnie Scotland, there's a meal and a pint on me.

To Lise Fontaine, for taking over the reins and for always having the right technological advice.

To Marta Wilczek-Watson, William Cheyne and Milla Rauma, for all your help with translation.

And for lending your voices, Marta Wilczek-Watson and Gabriella Bentley.

There are a great many other people to whom thanks is due, including, but not limited to; Cerys Allen, Sian Alsop, Robert Bluszcz, Amy Cannon, Wendy Cannon, Dr. Justine Coupland, Dr. Nikolas Coupland, Rebecca Craze, Amanda Evans, Dawn Harrington, Laura Johns, Heather Kerchal, Harriet Lloyd, Monika Owen, Rhian Rattray, Carol Rees, Dr. Ben Saunders, Elzbieta Slota, Matt Smith, Alysse Stehli, Amy Stephens, Erica Swain, Sara Sylvester, Carys Watkins, Piotr Wegorowski, Angela Williams and James Winters.

To Carly Rebecca Ellis, for the future.

Last but not least, this thesis is dedicated to three very special people:

First, to John Rupert Firth, linguist extraordinaire.

Second, to my fallen hero Riaz “Rico” Mohammed. You were brave, you ventured into the unknown and you kept going long after I'd have given up. Life is cruel, mate, and I'll never forget you.

Third, to my late grandfather Ken Nicholas, who passed away in the later stages of this work. You fought so bravely for so very long. Wherever you are, this one’s for you.

This work was supported by the Arts and Humanities Research Council (AHRC).

# Table of Contents

## List of Figures

## List of Tables



**Continues Overleaf**

## List of Appendices

Due to their length, the Appendices for this document are located on the attached CD-Rom (or at separate weblink if viewing document online). References to 'Appendix 1' in the document correspond to the .pdf file with the same name on the disc/website, and so on.

## Key to symbols and abbreviations

Consonants

| | |
|---|---|
| /ɹ/ | denotes a voiced alveolar approximant |
| /r/ | denotes an alveolar trill |
| /ɾ/ | denotes an alveolar tap |
| /l/ | denotes a voiced lateral approximant |
| /p/ | denotes a voiceless bilabial plosive |
| /b/ | denotes a voiced bilabial plosive |
| /t/ | denotes a voiceless alveolar plosive |
| /d/ | denotes a voiced alveolar plosive |
| /k/ | denotes a voiceless velar plosive |
| /g/ | denotes a voiced velar plosive |
| /m/ | denotes a bilabial nasal |
| /n/ | denotes an alveolar nasal |
| /ŋ/ | denotes a velar nasal |
| /f/ | denotes a voiceless labiodental fricative |
| /v/ | denotes a voiced labiodental fricative |
| /θ/ | denotes a voiceless dental fricative |
| /ð/ | denotes a voiced dental fricative |
| /s/ | denotes a voiceless alveolar fricative |
| /z/ | denotes a voiced alveolar fricative |
| /ʃ/ | denotes a voiceless palato-alveolar continuant |
| /ʒ/ | denotes a voiced palato-alveolar continuant |
| /tʃ/ | denotes a voiceless palato-alveolar affricate |
| /dʒ/ | denotes a voiced palato-alveolar affricate |
| /h/ | denotes a voiceless glottal continuant |
| /j/ | denotes a voiced palatal approximant |
| /ʁ/ | denotes an unvoiced uvular fricative |
| /χ/ | denotes a uvular fricative |
| /ʀ/ | denotes a uvular trill |

Vowels

| | |
|---|---|
| /i/ | denotes an unrounded high tense front vowel |
| /e/ | denotes an unrounded mid-high tense front vowel |
| /ɪ/ | denotes an unrounded high lax front vowel |
| /ɛ/ | denotes an unrounded mid front vowel |
| /œ/ | denotes a rounded mid front vowel |
| /æ/ | denotes an unrounded low front vowel |
| /ɶ/ | denotes a rounded low front vowel |
| /ɒ/ | denotes a rounded low back vowel |
| /ɔ/ | denotes a rounded mid back vowel |
| /u/ | denotes a rounded high tense back vowel |
| /ʊ/ | denotes a rounded high lax back vowel |
| /ɑ/ | denotes an unrounded low back vowel |
| /ʌ/ | denotes an unrounded mid back vowel |
| /ə/ | denotes an unrounded and unstressed mid/central vowel |
| /eɪ/ | denotes a diphthong moving from a mid front to a high front vowel |
| /aɪ/ | denotes a diphthong moving from a low front to a high front vowel |
| /ɔɪ/ | denotes a diphthong moving from a mid back to a high front vowel |
| /aʊ/ | denotes a diphthong moving from a low front to a high back vowel |
| /əʊ/ | denotes a diphthong moving from a mid vowel to a high back vowel |
| /ʊə/ | denotes a diphthong moving from a high back vowel to a mid vowel |
| /ɪə/ | denotes a diphthong moving from a high front vowel to a mid vowel |
| /ɛə/ | denotes a diphthong moving from a mid front to a mid central vowel |

Other Symbols

| | |
|---|---|
| “ ” | denotes the graphological representation of a word or sound |
| / / | denotes the phonemic transcription of a word |
| << >> | denotes the denotational and connotational meaning(s) of a word |
| : | denotes the lengthening of the preceding vowel |
| ˜ | denotes the nasalisation of a vowel or consonant not normally nasalised |

## ABSTRACT

This thesis studies twelve word-initial phonaesthemes identified in the vocabulary of English. Phonaesthemes (Firth, 1930; 1935) are phonotactic sequences that recur in multiple words with similar meanings. While several scholars have discussed these phonaesthemes (e.g. Sturtevant, 1947; Marchand, 1966; Bowles, 1995), no study has yet provided a detailed quantitative account of the extent to which they are present in the vocabulary of English or any other language. Moreover, no study has yet investigated whether these patterns are associations that speakers actually perceive, or what might influence their perception. In response to these gaps in the literature, this thesis begins with a quantitative study of the extent to which the twelve phonaesthemes are present in the vocabularies of three languages; English, French and Polish. This involves calculating and comparing the number of different words that exhibit each phonaestheme in the three languages. The study then investigates whether native speakers of English, French and Polish actually perceive the phonaesthemes. This is measured by studying the extent to which speakers productively associate the twelve phonotactic sequences with their respective meanings in the context of coined words. Finally, the study investigates a number of factors that could affect the extent to which the phonaesthemes are perceptible patterns. The findings indicate clear quantitative evidence of all twelve phonaesthemes in the vocabulary of English. Nine of the phonaesthemes are also evident in the vocabularies of French and Polish; however, eleven of the twelve patterns are more pervasive in English. In addition, the productive experiments find that all twelve phonaesthemes are perceived by English speakers; though some are more consistently perceived than others. It is clear from these experiments that each phonaestheme is more widely and perceived in English than in French or Polish. Finally, it appears that one key factor affecting the perceptibility and thus productivity of the phonaesthemes is the number of different meanings with which the phonotactic sequences recur.



**Keywords: phonaesthemes, sound symbolism, prosodies, connotations, productivity**

# CHAPTER 1: INTRODUCTION

## 1.1) The Context

Smooth down the avenue glitters the bicycle
Black-stockinged legs under navy blue serge
Home and Colonial, Star, International
Balancing bicycle leant on the verge
(Betjeman, 2006:69)

In the above extract taken from John Betjeman's poem 'Myfanwy', the use of the verb 'glitters' seems rather unusual. One might associate a bicycle with any number of disyllabic words that Betjeman could have chosen to preserve the meter of the poem – such as 'freewheels', 'trundles' or 'cruises'– but one would not typically associate a bicycle with the action of glittering. In the list of qualities that might be evoked when one encounters the word 'bicycle', the ability to emit or reflect light is likely to be rather less salient than the properties of having wheels or providing exercise. Moreover, 'glitter' is a verb I would more usually associate with entities that are [or appear to be] stationary, such as stars, lights and jewels. Therefore, it seems that Betjeman may have chosen this verb in order to imply some additional, unstated meaning. Specifically, this meaning could be the idea of <<light>>. Many scholars (e.g. Marchand, 1966; Knowles, 1987:Ch.5; Short, 1996:119-120; Sadowski, 2001) have observed that word-initial /gl/ tends to recur in English words that express a meaning related to <<light>> (as exemplified by 'gleam', 'glimmer', 'glisten', 'glare', 'glint', 'glow' and so on). It seems feasible that Betjeman's choice of verb might metaphorically identify the bicycle and its rider with light and perhaps heavenliness. This concurs with the fact that Betjeman openly declared his adoration of the poem's subject and rider of the bicycle, the real-life Myfanwy Piper (Wilson, 2006:122-125; Hancock, 2013:349-352).

A similar situation is seen in the opening stanza of Lewis Carroll's nonsense poem 'Jabberwocky', cited overleaf:

’Twas brillig, and the slithy toves
Did gyre and gimble in the wabe;
All mimsy were the borogoves,
And the mome raths outgrabe.
(Carroll, 2001:17)

Here, the expression ‘slithy toves’ seems to suggest a particular meaning. Many scholars (e.g. Firth, 1930:184-187; Crystal, 2002:124-126) have pointed out that word-initial /sl/ tends to recur with <<unpleasant>> or <<pejorative>> meanings in English (as in ‘slimy’, ‘sleazy’, ‘slag’, ‘sloven’, ‘slippery’ and so on). Consequently, even though this expression is coined by Carroll, and has no established meaning in English, it is not difficult to construe an attitudinal association from it. From Carroll’s use of the word ‘slithy’ as a modifier of ‘toves’, one may well infer that these entities are in some way unpleasant. Arguably, this inference may also extend more weakly to the context in which the poem’s narrative takes place.

In the above examples, we are dealing with the phenomenon of **phonaesthemes**, a term coined by J.R. Firth (1930). In this thesis, phonaesthemes are defined as prosodic forms (‘prosodies’) which acquire connotations by virtue of recurring in a number of different words that are experientially linked (cf. Chapter 2). Notably, Short (1996:119-121) argues that many authors of literary texts might use devices like phonaesthemes to create implicit meanings about characters or events (perhaps for such reasons as foreshadowing or irony); or to create areas of perceptual prominence in some parts of the text (cf. Mukařovský (1970) on foregrounding).

The /sl/-<<pejoration>> and /gl/-<<light>> associations cited above are by no means the only potential phonaesthemes evident in the vocabulary of English. Indeed, there are many more such patterns, both in syllable-initial (henceforth ‘onset’), syllable-medial (‘nucleus’) and syllable-final (‘coda’) positions (cf. Brinton, 2000:65) [1]. This thesis studies the following twelve phonaesthemes in the onsets of word-initial syllables (cf. § 2.6):

1. The association between /sl/ and <<pejoration>>;

[1] Many scholars conflate nucleus and coda prosodies into a larger unit, the syllable ‘rhyme’ or ‘rime’ (cf. Local, 1995; Halliday, 2009:70). For methodological convenience, this is the approach taken by the present study (cf. Chapter 5).

2. The association between /sm/ and <<pejoration>>;
3. The association between /sn/ and <<pejoration>>;
4. The association between /stɹ/ and <<straightness and stretching>>;
5. The association between /st/ and <<strength; stoicism>>;
6. The association between /gl/ and <<light>>;
7. The association between /gɹ/ and <<gripping>>;
8. The association between /tw/ and <<smallness >>;
9. The association between /sw/ and <<fast or strong movement>>;
10. The association between /skw/ and <<weakness and cuteness>>;
11. The association between /kl/ and <<clumsiness; thickness; stupidity>>;
12. The association between /kɹ/ and <<crookedness; curvilinearity>>.

It is not just in English that the presence of phonaesthemes has been claimed. These types of form-meaning recurrences have been observed in many different languages of the world. However, the specific prosodies and meaning associations involved tend to be different to those recurring in the vocabulary of English. For example, McGregor (1996) observed that coda /g/ tended to recur in verbal roots expressing <<causation>> and <<verbal/oral processes>> in the Aboriginal language Gooniyandi. Similarly, Abelin (1999) observed that onset /fj/ tended to recur in words expressing <<unpleasant>> meanings in Swedish (cf. § 3.2). Such relationships may or may not be perceived by speakers of the languages in whose vocabularies they are found; this is one of the key questions at the heart of this thesis.

The debate as to whether there is any sort of relationship or congruence between the phonetic structures of words and certain meanings is by no means a recent one. In the 'Cratylus' dialogue, Plato questioned whether language, when it was used to name things, was a product of convention ('thesis') or in some way reflected the nature of the things it named ('physis'). Throughout the Mediaeval and Renaissance periods, scholars such as Siger of Brabant and Boethius of Dacia speculated that mankind's original language – whatever this may have been – would have been the "perfect" tongue in which names were given to objects 'ab ipsa natura' (on the basis of their nature), and words were thus able to mirror the essences of things through their sound structures (Anderson, 1998:54-56). However, it was only in the seventeenth century that the foundations for the theory of phonaesthemes were truly laid, when John Wallis became the first scholar to reject the idea that phonetic structures

were prepossessed of some intrinsic meaning. Wallis (1653; cited by Genette, 1995:37-42) observed a number of similar sound structures among English words with similar meanings, and proposed that their presence in these words might give the sounds meaning (rather than that the sound structures were present in these words because they had some intrinsic, a-priori meaning). This is an important distinction to make: It is important to remember that although there are many theories of sound-meaning correlations in language (these are reviewed briefly in Chapter 2), phonaesthemes are a very specific type of relationship that neither imply nor necessitate that any phonetic structure is intrinsically meaningful. There is no natural or inherent correlation between the sound and meaning of any phonaestheme; the prosodies of phonaesthemes only acquire their meaning associations as a result of patterns in the vocabulary of a language. The importance of this distinction will become clearer when phonaesthemes are differentiated from other forms of 'sound-symbolism' (§ 2.4); and when the notion of arbitrariness is considered (§ 2.5).

### 1.2) Gaps in the Existing Phonaestheme Literature

Although a number of scholars have suggested possible phonaesthemes in English and other languages, or have speculated where such patterns may have originated, the phenomenon remains largely unexplored in the literature. Since the time of Firth's works in the 1930s, only seven scholars have studied English phonaesthemes in any detail: Sturtevant (1947), Marchand (1966), Bowles (1995), Sadowski (2001), Crystal (2002), Reay (2006) and Williams (2014). Notably, none of these accounts provide a comprehensive empirical measure of the extent to which any of the twelve aforementioned phonaesthemes are present in the vocabulary of English. With the exception of Bowles (1995) and Crystal (2002), no study has attempted to calculate the proportion of English words featuring any these prosodies which express (and do not express) their respective meaning associations. Furthermore, no study has yet revealed the extent to which the twelve prosodies recur with the meanings listed above in comparison to any other meanings in English. Indeed, many scholars simply cite a list of words demonstrating each phonaestheme, and a list of words to illustrate the presence of counterexamples. But this says nothing about the *extent* to which these patterns recur in the vocabulary of English. For example, it does not reveal how many English words with onset /gl/ convey a meaning related to <<light>>; or whether this represents a larger or smaller proportion than the number of /gl/ words that do not. It also says nothing about whether /gl/ recurs with a meaning of <<light>> more frequently than

with any other meaning in English; and if so, how much more frequently. In addition to the above, no study has yet investigated whether these twelve form-meaning patterns are unique to the vocabulary of English, or whether they might also be evident in the vocabularies of other languages. For instance, might onset /tw/ also recur with the meaning <<smallness>> in French? If so, to what extent? More or less frequently than with any other meaning? How much more or less frequently? It is the lack of such empirical data on these fronts that constitutes the first major gap in the existing literature.

Another important gap in the present literature concerns whether the twelve phonaesthemes cited above are actually perceived by speakers. As is discussed in Chapter 3, only a single study has as yet investigated whether speakers perceive any sort of links between the sounds and meanings of phonaesthemes in their language; Abelin's (1999) study in Swedish. No study has yet investigated whether the twelve phonaesthemes outlined above, which have been identified by scholars in the vocabulary of English, are patterns that are actually perceived by English speakers. Moreover, it is not clear whether such perception might also extend to speakers of other languages. Are these phonaesthemes patterns that ordinary speakers can even perceive? Or is their perception confined to linguists and those who explicitly study patterns in languages? Finally, if phonaesthemes are perceptible patterns, then what sorts of factors might affect their perceptibility? Could the perceptibility of a phonaestheme be dictated by the number of words in a language that reflect the sound-meaning association? Might it also be dictated by the frequency with which speakers encounter the pattern in day-to-day life? Both possibilities are investigated in this thesis (cf. Chapter 5); and a number of other factors considered in Chapter 6.

### 1.3) Research Questions Posed

These gaps in the existing literature form the basis of the five research questions addressed in this thesis, which are stated below:

Research Question 1:
To what extent are the twelve aforementioned phonaesthemes observable in the present-day vocabulary of English?

Research Question 2:
To what extent are the twelve phonaesthemes observable in the present-day vocabularies of languages other than English?

Research Question 3:

Do native English speakers appear to perceive the twelve phonaesthemes?

Research Question 4:
Do native speakers of languages other than English appear to perceive the twelve phonaesthemes?

Research Question 5:
What sorts of factors might affect the perception of the twelve phonaesthemes?

## 1.4) Aims and Scope of Thesis

### 1.4.1) Aims

In response to the gaps in the literature and the five research questions stated above, the aims of this thesis are as follows. The first aim is to conduct a detailed, quantitative study of the presence of the twelve phonaesthemes in the vocabulary of English. To accomplish this, I will study all of the words in the OED that contain the twelve onset prosodies. I will calculate the proportions of these words which do and do not convey the meanings outlined above. In addition, I will note the most common meanings expressed by words featuring each prosody. Together, these procedures will generate a more detailed quantitative picture than is currently available of the extent to which the twelve phonaesthemes are (or are not) evident in the vocabulary of English.

The second aim of this thesis is to compare the presence of these phonaesthemes in the vocabulary of English with their presence in the vocabularies of two other languages. This is purely exploratory. As outlined above, there is currently no literature to suggest whether these patterns are likely to be largely unique to the vocabulary of English, or whether they may also be evident in other languages. As such, the empirical measures outlined above

will be repeated in the vocabularies of French and Polish, using these languages simply for methodological ease and for the purpose of conducting an exploratory sample (cf. Chapter 4). Investigating the presence of these patterns in the vocabularies of other languages will also provide a useful point of reference against which to compare their presence in English, as is discussed further in Chapter 4.

While the above measures address several of the gaps in the existing literature, this study is primarily an investigation into speaker perception of phonaesthemes. The most notable gap in the existing literature concerns whether speakers do, in fact, associate the twelve prosodies with the meanings they recurrently express. In essence, the whole notion of phonaesthemes revolves around the idea that such patterns are actually perceived by speakers. Yet in English at least, there is no data to suggest that this is the case. Accordingly, the third aim of this thesis is to explore the extent to which native English speakers might perceive the twelve phonaesthemes outlined above. By 'perceive', I mean whether speakers feel or identify some sort of affinity between the meaning associations and the specific prosodies under test that is not present with other prosodies; and whether they feel or identify some sort of affinity between the prosodies and the specific meaning associations under test that is not present with other [different] meaning associations. Due to the complexities involved in measuring the presence and extent of such 'affinities', and the absence of existing literature in this area, there are no established ways of testing this phenomenon; no conventional methodologies or strategies that I can apply or adapt. As a result, the methodologies I employ are almost entirely novel, and are very much exploratory. In all, I use four types of experiments to explore the possible perception of these patterns (cf. Chapter 5). Some of these experiments ask speakers to attribute meanings to prosodies, while others ask speakers to attribute prosodies to meanings.

The perception of the twelve phonaesthemes is not just explored among English speakers. As outlined above, there is currently no literature to suggest whether the phonaesthemes might be perceived by non-English speakers; although as is noted in Chapter 3, Abelin's (1999) study suggests that Swedish speakers may make some degree of association between onset /sl/ and <<pejoration>>. As such, the fourth aim of the present study is to tentatively explore whether speakers of other languages might also perceive the twelve phonaesthemes identified above. To this end, the experiments performed are also conducted upon native speakers of French and Polish, in keeping with the vocabulary

investigations described above. As in the vocabulary-based experiments, investigating the extent to which speakers of other languages perceive the twelve phonaesthemes will also provide a useful point of reference against which to compare their perception in English, as is discussed further in Chapter 4.

The fifth and final aim of this thesis is one that is dependent on the findings generated by the preceding experiments. Because no scholars have as yet investigated speaker perception of the twelve phonaesthemes, it is not clear what sorts of factors are likely to affect their perception; if indeed these patterns are perceptible at all. As such, the fifth aim of this thesis is to investigate whether different phonaesthemes appear to be more perceptible than others, and, more importantly, to consider what sorts of factors could affect their perception. There are a number of factors that could feasibly affect the degree to which phonaesthemes are perceived: the frequency with which the prosodies convey their associated meanings in the vocabulary of a language (henceforth 'type frequencies'); as well as the frequency with which speakers encounter the phonaesthemes in everyday language use (henceforth 'token frequencies'; cf. § 4.3).

### 1.4.2) Scope of Thesis

As discussed above, the scope of this thesis is to investigate a sample of twelve English phonaesthemes in a sample of three different languages. This is because it would clearly be impossible to test every possible English phonaestheme in every language of the world. As previously mentioned, I chose to study the presence and speaker perception of phonaesthemes in French and Polish in addition to English. This enabled me to study a sample of languages with as few methodological difficulties as possible (see fuller discussion in § 4.1). As Firth's (1930; 1935) work implies that onset phonaesthemes may be the most prevalent type in English, it had been my original intention to study every proposed English onset phonaestheme across the three languages. However, on further study, I found that there were too many onset phonaesthemes in English to study in detail. I then contemplated studying the thirteen onset phonaesthemes posited by Firth (cf. § 2.6). However, the meaning associations for two of these phonaesthemes (/dɹ/ and /fl/) were not clearly defined or exemplified, as is discussed further in § 2.6. This meant that I could not accurately infer the precise nature of these sound-meaning correlations as intended by Firth. As such, I excluded these onsets from the study, and focused instead on the remaining eleven onset

phonaesthemes posited in Firth's works. To this I added a further phonaestheme; the relationship between word-initial /gl/ and <<light>>. While not a phonaestheme originally discussed by Firth, the consistent and widespread discussion of this correlation by subsequent scholars makes it impossible to ignore. In combination, the eleven phonaesthemes from the work of Firth and the /gl/-<<light>> correlation form the twelve relationships shown on pages 8-9.

In addition to having to restrict the study to a sample of twelve phonaesthemes in three different languages, it is worth discussing the way in which I operationalized and measured the 'perception' of phonaesthemes. The idea of 'perception' is particularly complex and difficult to measure empirically, because it is not possible to 'feel' or experience people's cognitive reactions to such phonetic and semantic stimuli, or the extent of any affinities they perceive between these stimuli; let alone quantify this data. As there are no established methodologies in the literature for measuring the extent to which speakers perceive affinities between certain sounds and certain meanings (see above), I chose to measure perception of the phonaesthemes indirectly, by means of linguistic productivity. 'Productivity' concerns the extent to which speakers apply existing patterns in the language analogously in new or unfamiliar linguistic situations. For example, if a native English speaker were to encounter an unfamiliar word featuring onset /sl/, the frequency with which /sl/ recurs in <<pejorative>> English words might lead the speaker to infer that this word conveyed an unpleasant meaning. The speaker would thus have used the /sl/ phonaestheme productively. I reasoned that if I could expose speakers to unfamiliar lexical items bearing each phonaestheme, I could explore the extent to which the twelve patterns were productive in the speakers' language use, and thereby reach some measure of quantification about the extent to which each phonaestheme was perceived. To this end, I used four different productive experiments, which I presented to English, French and Polish speakers. Two of the experiments provided speakers with coined words featuring the phonaesthemes, presented in the medium of speech. In the manner of language attitudes studies (e.g. Giles and Powesland, 1975), these experiments tested the meaning associations ascribed to each of these coined words (and thus the phonaesthemes; cf. § 5.2). The remaining experiments reversed the ordering of stimuli; presenting speakers with the twelve meaning associations identified above, and testing the prosodic associations ascribed to these meanings by speakers.

### 1.5) An Overview of the Thesis

Having contextualised the study and introduced its research questions, aims and scope, this section provides a chapter-by-chapter overview of the development of the thesis. This overview completes the present chapter.

Following this overview, the function of Chapter 2 is to provide a more detailed definition and conceptualisation of phonaesthemes. The point of departure is to explain the concepts of denotation, reference, sense and connotation. It is necessary to distinguish these primary components of meaning in order to fully understand phonaesthemes, since the meanings conveyed by phonaesthemes are entirely connotative. Following this, the second section of the chapter illustrates the concept of prosodic phonology. This discussion is vital in order for the reader to understand the way in which I conceptualise and discuss phonaesthemes in the present study. This section begins by defining the 'speech signal', then explains the two different views of the speech signal presented in the literature; the phonemic approach and the prosodic approach first proposed by Firth (1930; 1948). In doing so, it discusses how and why Firth's prosodic approach contrasts with the phonemic approach. The section concludes by discussing my reasons for adopting a prosodic view of the speech signal, and for treating phonaesthemes as prosodies (i.e. as gestalts) rather than as combinations of phonemes. Having outlined the necessary prerequisites, the third section of the chapter provides a detailed definition of phonaesthemes, beginning with the notion of 'phonetic habits' which led to Firth's (1930) use of the term for the first time. This very particular type of sound-meaning correlation is then differentiated from other types of 'sound-symbolism' in the fourth section of the chapter. This discussion is motivated by the fact that many of the terms referring to sound-meaning correlations are often used interchangeably in the literature. Following this, the fifth section of the chapter discusses how phonaesthemes, unlike several other sound-meaning correlations, are fully compatible with a view that the linguistic sign is arbitrary. If anything, they actually appear to provide further support for this idea. The chapter culminates in a sixth and final section, which discusses the support in the literature for the twelve phonaesthemes under study. As noted above, this coverage is relatively limited, and many of these patterns remain largely unexplored. The scholars to have discussed these phonaesthemes in English include Firth (1930; 1935), Sturtevant (1947), Marchand (1966), Bowles (1995), Sadowski (2001), Crystal (2002), Reay (2006) and Williams (2014).

The goal of Chapter 3 is to motivate the particular methodologies used to address the research questions. The chapter is divided into three subsections. The first discusses the evidence which suggests that phonaesthemes could become productive in language. In essence, this idea underpins all the methodologies which test speaker perception of phonaesthemes in this thesis (cf. 'Scope of Thesis', above). The chapter demonstrates how phonaesthemes are congruent with at least two views of language patterning; Usage-Based and Probabilistic Phonology. Within the frameworks of both models, it would be perfectly plausible for the phonaesthemic patterns of a language to emerge and become productive. In its second subsection, the chapter then critically examines the only study to have thus far investigated speakers' perceptions of phonaesthemes – Abelin's (1999) investigation in Swedish. Because there is no established methodology for the type of investigation conducted by the present thesis, Abelin's study is examined in order to assist with methodological cues and motivate the forms taken by my experiments. The most robust of Abelin's experiments are adapted and used to construct a pilot study (cf. Chapter 5). The chapter concludes with a third and final subsection which discusses my motivations and rationale for studying the presence and perception of the phonaesthemes in French and Polish alongside English. It weighs the advantages and disadvantages of conducting a broader study of the patterns across three languages against conducting a more detailed study of the patterns in English alone.

Chapter 4 presents the methodologies and findings of the vocabulary-based experiments used to address Research Questions 1, 2 and, in part, Research Question 5 (cf. § 1.3). The chapter comprises three subsections. The first subsection details the methodology and findings of the two-stage experiment used to address Research Questions 1 and 2. In its first stage, this experiment measures the phonaesthemes' 'type' frequencies in English, French and Polish; that is, the number and proportion of different words featuring each phonaestheme that convey its associated meaning shown in § 1.1. In its second stage, the experiment then investigates the most common meanings expressed by words featuring each of the twelve onsets. This is in order to investigate whether the meanings shown in § 1.1 are among those that recur most frequently with each onset. In combination, these processes generate a quantitative picture of the extent to which the twelve phonaesthemes are present in the vocabularies of English, French and Polish. The second subsection of the chapter then presents the methodology and (briefly) the findings of the experiment used to study the phonaesthemes' 'token' frequencies in English. The token frequencies provide a

representative measure of the extent to which speakers encounter the twelve phonaesthemes in everyday language use (see further discussion in § 4.3). In combination with the productive and type frequency data, the token frequencies of the phonaesthemes are used to address Research Question 5. I undertook this stage of the investigation in English only because there was no motivation to predict that the phonaesthemes would be present in the vocabularies of French or Polish, or perceived by French and Polish speakers (cf. Chapter 3.3). The discussion of whether it is the type or token frequency which most appears to affect the productivity (perception) of a phonaestheme is given at the close of Chapter 5, following the presentation of the productive data. The third and final subsection of the chapter summarises how the findings of the main vocabulary-based experiment address Research Questions 1 and 2.

Chapter 5, which is divided into five subsections, presents the methodologies and findings of the productive experiments used to address Research Questions 3-5 (cf. § 1.3). The forms taken by these experiments were informed by a small-scale pilot study. As such, the first section of the chapter sets out the methodologies of the five test types trialled in the pilot. It then critically examines their strengths and weaknesses in light of the main trends shown and the feedback provided by participants. It concludes by detailing the adaptations made to these experiments in preparation for the main study. The second subsection of the chapter then presents the methodologies and findings of the revised productive experiments implemented as the main study. These are presented on an experiment-by-experiment basis. The affordances and constraints of each test type are acknowledged critically. For each experiment, the findings from the English respondent cohort are presented first (these address Research Question 3); followed by those from the French and Polish cohorts (which address Research Question 4). Having examined each productive experiment in turn, the third subsection of the chapter calculates an overall measure of the productivity of the phonaesthemes in each language. This provides a cumulative picture of the most and least productive phonaesthemes in each language; something which is difficult to judge when analysing the findings of the individual experiments. It also enables a broad comparison of the productivity of the phonaesthemes across the three language cohorts. This three-way comparison is supplemented by statistical testing. The penultimate section of the chapter then compares the overall productivity of the phonaesthemes in English with their type and token frequencies (see above), in order to address Research Question 5. This analysis suggests that neither the type nor token frequencies of the phonaesthemes are wholly responsible for their

productivity; as such, other possible factors are considered in Chapter 6. The chapter culminates in a fifth and final subsection which summarises how the findings of the productive experiments – both individually and cumulatively – address Research Questions 3-5.

Chapter 6, which comprises two subsections, begins by providing a tentative discussion of possible factors that might have affected English speakers' perception (and thus productive use) of the phonaesthemes. This discussion is motivated by the fact that differences in productivity are apparent among the twelve phonaesthemes in English, and the fact that these differences do not appear to be wholly attributable to the phonaesthemes' type or token frequencies. The chapter proposes and investigates two further possible factors, labelled "multiple semantic recurrence" and "multiple phonetic recurrence". "Multiple semantic recurrence" is the label given to the idea that one [onset] prosody could recur with many [different] meanings; while "multiple phonetic recurrence" is the label given to the idea that one meaning could recur with multiple [onset] prosodies. Both factors are tested on a number of the phonaesthemes under study. The findings suggest that while both factors could have affected the productivity of phonaesthemes, multiple semantic recurrence seems to have been the most influential. Chapter six closes with a second subsection which discusses the limitations of the thesis and its findings.

Chapter 7 concludes the thesis and provides an outline of possible areas for further research in the future.

# CHAPTER 2: LITERATURE REVIEW

## 2.1) Defining 'Meaning': Denotation, Reference, Sense, Connotation

Before one can define or discuss phonaesthemes in any detail, one first needs to consider the way in which these phonetic forms are meaningful. Broadly, the literature identifies four different ways in which linguistic structures such as words, morphemes, prosodies and phrases can convey meaning. These are labelled denotation, reference, sense and connotation (Jackson and Ze Amvela, 2000:66-68). It is important to distinguish these types of meaning because phonaesthemes *only* create connotative meaning, as is discussed below (see also § 2.3).

The first component of linguistic meaning identified in the literature is denotation. In this thesis, denotation is understood as 'the relationship that holds between [a linguistic form] and persons, things, places, properties, processes and activities external to the language system' (Lyons, 1977:207; see also Crystal, 2003:170; Murphy and Koskela, 2010:56). The denotative meaning, or "denotatum", of a word or expression, is 'the class of objects, properties, etc., to which the expression…applies' (Lyons, 1977:207). For example, 'the denotatum of [the expression] 'cow' is a particular class of animals […] the denotatum of 'red' is a particular property' (ibid.). Denotation is a relationship that is presupposed and remains consistent each and every time a linguistic form is used, irrespective of the particular linguistic or situational context (Lyons, 1977:208). For instance, every time the word "cow" is written or uttered in English, the property <<bovine>> is presupposed; irrespective of any particular cow to which the speaker may be referring, or any implied meaning they may be seeking to create. Similarly, every use of the English verb "shout" presupposes the ideas <<uttering>> and <<loudly>> (OED: Online); a level of meaning that holds irrespective of any specific instance of shouting to which the speaker may be referring.

The second type of meaning pertaining to linguistic forms is reference. Frege (1892) was the first scholar to distinguish referential meaning from denotative meaning, and his work provides a detailed account of this distinction. For the purpose of this thesis, reference is understood as a type of meaning connected to the context of situation in which a phrase or expression is used (Lyons, 1977:207). Reference concerns the specific entity, process or quality referred to when a speaker or writer uses a particular expression in a particular

context (Jackson and Zé Amvela, 2000:66). For example, consider the English word ‘car’. The denotation of this word is <<wheeled conveyance>> (OED: Online). A speaker or writer may use this expression to talk about wheeled conveyances in general, such as in the utterance ‘most modern cars have four wheels’. However, they may also use this form in the context of referring to a specific car, such as in the utterance ‘have you seen John’s new car?’ In the first utterance, ‘car’ has no referent, inasmuch as there is no specific entity being identified by the use of the expression; just the concept of <<wheeled conveyances>>. But in the second utterance, the speaker or writer would be using the expression ‘car’ to identify a real-world entity; the specific vehicle owned and recently purchased by John. In this context, the expression ‘car’ has a referent.

The third strand of meaning pertaining to linguistic forms is sense. Sense refers to the network of meaning relationships between a word or expression and other words or expressions in the same language system (Rambaud, 2012:26). Synonymy, antonymy, hyponymy and meronymy are all examples of instances where linguistic forms are connected by a sense relationship (Jackson and Zé Amvela, 2000:106). To use a widely-cited example, there is a relationship of sense between the expressions “Venus”, “the morning star” and “the evening star” (cf. Frege, 1892:119; Nöth, 1995:93; Fontaine, 2008:20). Depending on the context, all three expressions can be synonymous; they can all be used to identify and talk about the planet Venus. While “Venus” is the astronomical name given to the body, “the morning star” is a term sometimes used in English to refer to the planet ‘when seen in the morning before sunrise’ (Frege, 1892:119). Similarly, “the evening star” is a term sometimes used to refer to the planet ‘when it appears in the heavens after sunset’ (ibid.). As such, there is a relationship of meaning between all three expressions: All three can have the same reference, so long as they are written or uttered in a context where the speaker is discussing <<planet Venus>>.

The fourth and final strand of meaning identified in relation to linguistic forms is connotation. First discussed by philosopher J.S.Mill (1843:37), connotations are implicit, associative ideas or qualities ascribed to words and other linguistic forms. Crucially, these associations are largely independent of the denotation, reference or sense of any item (Gill, 2011:62). Moreover, they are tacit associations made broadly by a society or culture, rather than idiosyncratic or emotional overtones perceived by individual speakers (Halliday and Yallop, 2007:30; Greene, Cushman, Cavanagh, Ramazani and Rouzer, 2012:298). To clarify,

consider the following example. The denotations of the English nouns “fragrance” and “odour” are broadly similar: both terms denote <<a property or stimulus perceived by the olfactory sense>> (cf. OED: Online). However, their connotations are less similar. Many English speakers would probably agree that “odour” carries associations of a more negative or unpleasant olfactory sensation than “fragrance”. These <<unpleasant>> or <<pejorative>> associations are connotations of “odour”; they are unstated ideas and qualities ascribed to this linguistic form. Of course, it is not just linguistic forms referring to objects, concepts or entities (i.e. nominal expressions) that can connote meaning. Just about any content-bearing linguistic form could be imbued with some connotational meaning. For example, Gill (2011:10) identifies some of the connotations associated with the adjective ‘red’ in English-speaking cultures; such as ‘blood, communism, fire, Santa Claus and so on’. Similarly, Hunston (1995) discusses the connotations associated with a number of English “report” verbs, including ‘acknowledge’, ‘insist’, ‘claim’ and ‘argue’ (cf. Thompson and Ye, 1991). Hunston (1995:133) suggests that each of these verbs connote disagreement ‘between the writer of the text and the one to whom the material is attributed’; an implicit, associative level of meaning beyond their denotations.

According to Backhouse (1992:297-298) and Partington (1998:65-66), connotations can be formed in a number of different ways. One of the ways is as a result of ‘what a [linguistic] item denotes within a culture’ (Partington, 1998:66). In other words, values or qualities may be ascribed to particular linguistic forms [by a culture or society] as a result of what these forms denote. To use Partington’s (ibid) example, the English noun “woman” denotes <<an adult female human>> (OED: Online). However, in English-speaking societies, there are particular qualities and attitudes that have been traditionally associated with adult female humans. According to Leech (1974:15), women have in the past been associated with such ideas as <<frailness>>, <<cowardliness>>, <<irrationality>> but also <<gentleness>>, <<compassion>> and <<sensitivity>>. As a result of these social stereotypes, the linguistic form denoting <<woman>> has acquired connotations of these ideas.

The second way in which connotations may be formed is through what both Backhouse (1992) and Partington (1998) label ‘social or situational’ circumstances. This term refers to connotations formed from the real-world contexts in which a word, expression or other linguistic item is commonly used. For example, if a linguistic form is frequently used (or perceived to be frequently used) by a certain social or cultural group, the form might acquire

associations (i.e. connotations) of this group (ibid). This is the idea at the heart of many variationist sociolinguistic studies. For instance, Labov (1966) famously argued that the use of post-vocalic /ɹ/ connoted prestige and social status in 1960s New York, as a result of the frequency with which it was used by speakers in higher socioeconomic classes. Similarly, if a particular linguistic structure is routinely used in a particular discursive context, the form might acquire connotations of this type of discourse. For instance, imagine that two individuals are romantically involved, and one member of the couple begins a conversation with the discourse marker "we need to talk". In a relevant context, this expression could connote <<I feel there is something wrong with our relationship>> or <<I am about to terminate our relationship>> to the speaker's interlocutor. This is because this particular discourse marker is perceived to be frequently used to preface a relationship break-up (Carey, 2011: Online).

In addition to the situational contexts in which linguistic items are found, connotations can also be formed from the particular *linguistic* contexts in which a form is encountered. It is connotations formed in this way which are particularly important to the theory of phonaesthemes, as is discussed below (cf. §2.3). To cite an example, Firth (1951a: 194-195) suggested that the word 'ass' frequently collocated with (i.e. occurred in combination with) a preceding pejorative adjective; most commonly 'silly, obstinate, stupid or awful'. As such, he argued that part of the meaning of the word 'ass' was its unpleasant or negative associations; at least in colloquial English. Firth argued that the meaning of any word was, at least in part, attributable to the items with which it collocated; and that to some extent, the meaning of a word could be known 'by the company it keeps' (Firth, 1957:11; see also Sinclair, 2004:28). In the same way, phonaesthemes also involve linguistic items acquiring connotations as a result of the linguistic contexts in which they are frequently found
(cf. § 2.3). The only difference is that the linguistic items in question are phonetic sequences (prosodies) rather than entire words; and that the linguistic contexts in which these forms recur are *within* words rather than in combination with other words. For example, the word-initial prosody /sl/ was cited in Chapter 1 as frequently recurring in words (i.e. linguistic contexts) expressing some kind of <<pejorative>> meaning. Therefore, just as the word 'ass' could acquire the connotation <<unpleasantness>> because of the linguistic contexts in which it is found, so too could the prosody /sl/. Similarly, the prosody /gl/ could acquire the connotation <<light>> because of the linguistic contexts in which it is found, and so on.

In sum, this section has outlined the four distinct types of 'meaning' that scholars discuss in relation to linguistic forms such as words, morphemes, prosodies and phrases. Denotation refers to 'the class of objects, properties, etc., to which [an] expression…applies' (Lyons, 1977:207), while reference is the act of identifying a specific entity, quality or process through the use of a linguistic expression in a particular context. Sense refers to the network of meaning relationships between a particular linguistic form and other forms in the language system (such as synonymy, antonymy, hyponymy and meronymy); while connotation is an unstated, implicit association ascribed to a word or other linguistic form by a particular society or community. Of these four types of meaning, it is only connotation that is important to the theory of phonaesthemes; specifically, connotations formed as a result of the linguistic contexts in which a form or expression is commonly found. This is because phonaesthemes involve particular linguistic forms acquiring associative ideas as a result of the words (i.e. linguistic contexts) in which they recur (cf. §2.3). For example, as discussed in Chapter 1, onset /sl/ is frequently found in [English] words expressing <<pejorative>> meanings. As a result, this linguistic form acquires an association of <<pejoration>> in English. But such prosodic forms clearly cannot denote or refer. Being phonetic structures below the level of words and morphemes, phonaesthemes do not designate a class of 'objects, properties, etc' (Lyons, 1977:207); and cannot be used to identify (refer to) a particular object, quality or process *in any context* unless they are combined with at least one other syllable prosody to form a complete word. Moreover, while there may be a relationship of sense between *some* of the words that exhibit a phonaestheme, this relation does not necessarily extend to all words bearing that particular prosody. For example, while many English /sl/ words express something <<pejorative>>, this is not true of all English /sl/ words. That is to say, there is no context in which all English /sl/ words are connected by a meaning of <<pejoration>>, or can all be used to refer to something unpleasant.

Having outlined the ways in which linguistic forms such as words, morphemes, prosodies and expressions can be 'meaningful', and discussed how and why phonaesthemes only convey meaning through connotation, the next stage in reaching a definition of phonaesthemes is to define and explain the term 'prosody'. This is the focus of the next section.

## 2.2) Defining Prosodies

By definition, a study of phonaesthemes is a study of meaningful phonetic units in the speech signal. However, there are differing views among scholars about how the speech signal should be analysed, and about what sorts of linguistic forms constitute meaningful phonetic units. When Firth used the term 'phonaestheme' for the first time, the established way of analysing the speech signal was the phonemic or 'segmental' approach. But Firth was dissatisfied with this perspective, and proposed an alternative view; prosodic phonology. It is only by first understanding the underpinnings of prosodic phonology that the conceptualisation of phonaesthemes used in this thesis can be fully understood. In this thesis, I adopt a prosodic rather than phonemic view of the speech signal, and treat phonaesthemes as prosodies. As such, the goal of the present section is to provide the reader with an overview of prosodic phonology. To this end, a brief definition of the term 'speech signal' is first given below. This is followed by a necessarily condensed outline of the phonemic ('segmental') view. Firthian prosodic phonology is then introduced and differentiated from the phonemic approach. The section culminates by pointing out why it is most appropriate to adopt a prosodic view of the speech signal in the present thesis, and to treat phonaesthemes as prosodies rather than as sequences of phonemes. In so doing, phonaesthemes are visually illustrated using diagrams adapted from Autosegmental Phonology.

### 2.2.1) The Two Components of the Speech Signal

Both the phonemic and prosodic view ultimately agree that the 'speech signal' is a concept that exists on two levels (Giegerich, 1992:31; Carroll, 2008:20). On the one hand, it exists at a 'concrete phonetic level' (Giegerich, 1992:31). This refers to the material, physical production of speech and its auditory perception. Speech is physically produced by simultaneously passing air through the vocal tract whilst producing gestures with the articulatory organs; the vocal folds, velum, oral cavity and tongue etc. The passage of air through the particular configurations of the vocal tract creates patterns of sound waves. These sound waves are carried through the air and detected by the ears of a listener. Each discrete auditory unit that the listener perceives from this continuum of sound waves is termed a phone (Carroll, 2008:20; Crystal, 2011). But in addition to this material, physical production and perception, the speech signal is also said to exist at a psychological level. This level is speakers' knowledge of the units of sound that are meaningful in their language. The

phonemic and prosodic approaches differ in the way in which they view this cognitive dimension of the speech signal. Put simply, the two theories propose a different analytic view of the meaning-making phonetic units in a language system (the forms that speakers interpret from the series of phones in physical speech, and cognitively process into words, expressions and sentences). Each of these views is now outlined in turn, with the primary focus on the view proposed by prosodic phonology. This is to ensure that the definition and conceptualisation of phonaesthemes used in the present thesis (discussed in § 2.3) can be fully understood.

### 2.2.2) Segmental Phonology - The Phonemic Approach

The phonemic or 'segmental' view argues that the psychological level of the speech signal is formed from a series of abstract models (Giegerich, 1992:31-32). Each of these abstract models, termed a **phoneme**, is an idealised acoustic representation of one of the smallest discrete phonetic segments that make 'contrasts in meaning' within a language (Crystal, 2011). Consider the word-initial sound segments in the English expressions "bit" /bɪt/ and "pit" /pɪt/. These sound segments signal a contrast in meaning between two words that are otherwise phonetically identical; a 'minimal pair' (cf. Cruttenden, 2008:41-42). This is termed contrastive distribution (Hayes, 2009:20). Because /b/ and /p/ signal a difference in meaning between these otherwise phonetically-identical items, they represent distinct phonemes of English; discrete phonetic segments that make contrasts of meaning. However, there would be no such contrast in meaning if the initial sound segment in the English word "pit" /pɪt/ was uttered with or without aspiration (i.e. as [pʰɪt] or [pɪt]). That is to say, uttering the initial sound segment in "pit" with or without aspiration would not result in a contrast of meaning. Therefore, [pʰ] and [p] are not in contrastive distribution, and are not different phonemes of English, even though they are phonetically distinct (cf. Ladefoged and Johnson, 2010:72). In the phonemic view of the speech signal, each word in the language system is processed as a unique combination of phonemes, excepting cases of homophony.

### 2.2.3) Prosodic Phonology

Although the phonemic view is dominant in much linguistic theory (Crystal, 2011), Firth (1948:122-123) criticised the idea that the [continuous] speech signal could be segmented into such minimal, discrete sound units. As O'Grady (2013:55) notes, 'the

cornerstone of Firth's approach was his belief that the speech signal does not easily lend itself to analyses in terms of discrete segments. Rather, he postulated that some features had the potential to spread across segments'. Firth observed that many of the meaningful auditory structures in the speech signal could actually extend across the phone(s) realising several – or indeed very many – phonemes. Cases in point include intonation, vowel harmony, assimilation and nasalisation. For example, in the system of intonation, a particular tone movement may only extend across the realisation of a single phoneme. However, it may also extend across the phones realising several phonemes, several lexemes, or even an entire clause; a speaker may produce an utterance with a single tone movement that begins with the first syllable and ends with the final syllable (cf. Halliday, 1970: Ch.1; see also Figure 1, below). Another example is vowel harmony (Clark, Yallop and Fletcher, 2007:407). While no longer a feature of English, vowel harmony can be found in many languages of the world. Vowel harmony is a phenomenon whereby the first vowel in a word constrains any possible succeeding vowels; only those that 'agree' with the first vowel in certain phonetic features may be used. For instance, in Turkish, 'a vowel other than the first in a word may be low [and] unrounded or high; [but all] other features are…taken from the first vowel' (ibid). By its very definition, then, vowel harmony is a system that extends across the realisation of several phonemes. Because many of the meaningful auditory structures in language had the potential to spread across the realisations of several 'phonemes', Firth argued that the segmental approach did not provide the most appropriate view of the speech signal. Instead, Firth suggested that it would be more appropriate to conceptualise the speech signal as a sum of auditory structures that could extend in linear sequence to any length across the utterance (O'Grady, 2013:55). Firth termed these auditory structures **prosodies** (Firth, 1948:122-123; Firth, 1951b:226-227).

In this thesis, as noted above, I conceptualise the twelve phonaesthemes as prosodies rather than as combinations of phonemes. My reason for adopting this view is that there are both semantic and phonetic grounds for treating the phonaesthemes as gestalts. Semantically, each phonaestheme is a unit of [connotational] meaning that extends across the realisation of several phonemes. Ten of the twelve phonaesthemes (/sl/, /sm/, /sn/, /kl/, /kɹ/, /st/, /tw/, /gl/, /gɹ/ and /sw/) extend across the phones realising two phonemes; while the /skw/ and /stɹ/ phonaesthemes extend across the phones realising three phonemes. As was shown in Chapter 1, it is these phonetic sequences *in their entirety* that are associated with their particular connotations (see further discussion in §§ 2.3; 2.5). For instance, it is the onset sequence /sl/

as a whole that is associated with <<pejoration>>; not just one or other of the constituent consonants. Similarly, it is the onset sequence /sw/ as a whole that is associated with <<fast or strong movement>>; not /s/ or /w/ in isolation. As such, conceptualising the phonaesthemes as prosodies (rather than combinations of phonemes) clarifies that their connotations extend over the entire phonetic sequence; and avoids any implication that the meaning associations might be attributable to one or other of the constituent consonants.

In addition to semantic grounds, there are also phonetic grounds for conceptualising the phonaesthemes as prosodies. In all twelve phonaesthemes, there are certain phonetic features which extend across the entirety of the onsets by way of assimilation. This suggests that it is more appropriate to treat the phonaesthemes as single auditory units of meaning. To demonstrate this point, a visual representation is required. The theory of Autosegmental Phonology (Goldsmith, 1979) uses a system of notation which can be used to show the spread of phonetic features across the phonaesthemes. The central idea behind Autosegmental Phonology is that although the articulatory gestures used to produce phones do not 'start and finish all at the same instant' (ibid:413), speakers interpret the physical speech signal as if this were the case (cf. Geigerich, 1992:32). Goldsmith (1979:212) observed this phenomenon in utterances made by speakers of Igbo, an African tone language. He saw that the phonation of lexical tones – pitch movements within or across a syllable that are essential to the meaning of a word – did not necessarily begin and end with the phonation of discrete vowel phonemes. This in itself supports Firth's criticisms of the segmental approach: For Igbo speakers, lexical tone is clearly a meaningful auditory structure in the speech signal; yet Goldsmith observed that it could extend across the boundaries of discrete phonemes. To illustrate this spreading of phonetic features across segments, Goldsmith used a system of two-tiered diagrams. These diagrams illustrated the consonant and vowel segments of words on the upper tier, and phonetic features on the lower tier. The spreading of these features across the consonant and vowel segments was indicated by connecting lines. Figure 1, overleaf, shows an example of this tiered representation:

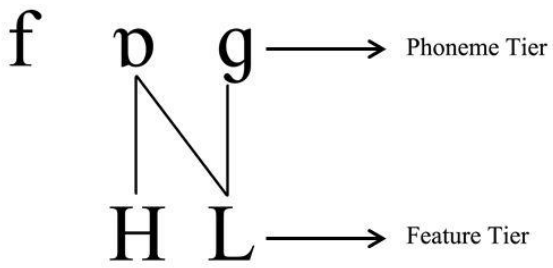



Figure 1: Autosegmental diagram of a falling tone across the English monosyllable /fɒg/ (adapted from Clements and Hume, 1995:247).

Figure 1, above, shows a hypothetical example in which a pitch (tone) movement spreads across two phonemes in the English monosyllable "fog" /fɒg/. The tone movement begins with a high pitch (H) on the first voiced phoneme (the vowel /ɒ/). This is indicated by the vertical line connecting the /ɒ/ in the phoneme tier with the H in the feature tier. Throughout the phonation of the /ɒ/ vowel, the pitch falls. This falling tone movement extends into the phonation of the succeeding voiced /g/. By the time the phonation of /g/ is complete, the pitch is low (L). The low pitch of the coda /g/ is indicated by the vertical line connecting the /g/ in the phoneme tier with the L in the feature tier. The spread of the falling tone across both phonemes is indicated by the diagonal line connecting the low tone in the feature tier (L) with both the /ɒ/ and /g/ phonemes.

This form of tiered diagramming can be used to show the presence of phonetic features spread across all twelve of the onsets under study; thus demonstrating the phonetic grounds for conceptualising the phonaesthemes as gestalts rather than as sequences of phonemes. For instance, Figure 2 (overleaf) shows that in the consonant cluster /sl/, the voicelessness of /s/ (i.e. the feature [-voice]) spreads from the /s/ into the following /l/ by way of progressive assimilation (Cruttenden, 2008:214; 297). This is indicated by the dotted line. This progressive assimilation causes the (usually-voiced) /l/ consonant to become devoiced (Cruttenden, 2008:214). This is indicated by the strike through the phonetic feature [+voice] annotated under /l/:

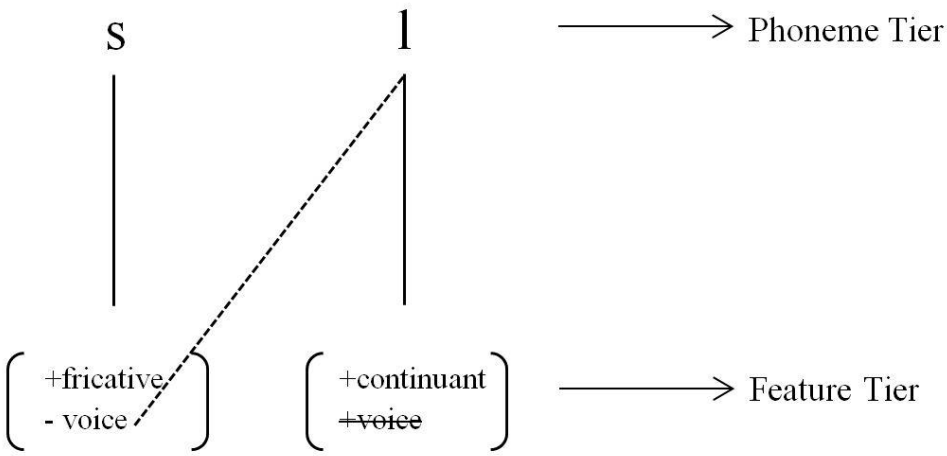



Figure 2: The spread of [-voice] across the cluster /sl/

Because the voicelessness of /s/ spreads across the onset into /l/, and causes /l/ to lose the voicing it would have had if produced at the beginning of a cluster or as a single onset, it is appropriate to view the cluster as a single phonetic unit extending in linear sequence (i.e. a prosody). The same is true for the remaining eleven phonaesthemes. In each case, one or more phonetic feature spreads across the entirety of the onset sequence, making it more appropriate to consider the phonaestheme as a gestalt. Table 1, below, provides a visual diagram and descriptive summary of the phonetic features spreading across the remaining eleven phonaesthemes.

Table 1: Phonetic features spreading across onsets /sm/, /sn/, /st/, /stɹ/, /sw/, /skw/, /tw/, /kl/, /kɹ/, /gl/ and /gɹ/

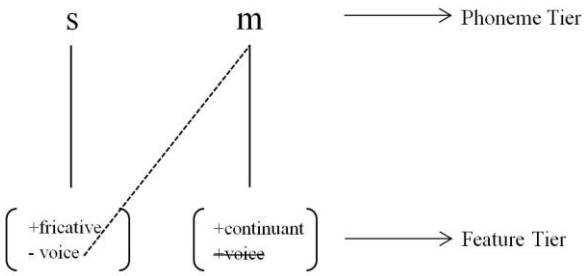
| **Phonaestheme** | **Diagram showing phonetic feature(s) spreading across the phonaestheme** | **Explanation** |
|---|---|---|
| /sm/ |  | Voiceless /s/ causes devoicing of succeeding /m/ (Cruttenden, 2008:209) |

Table Continues Overleaf

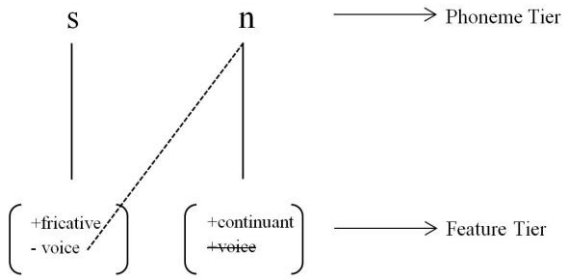
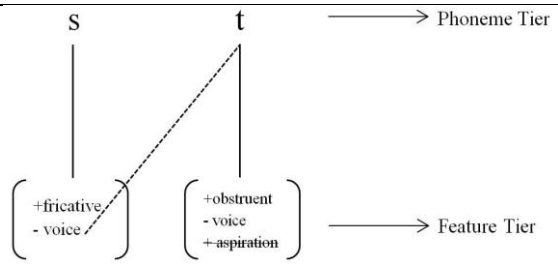
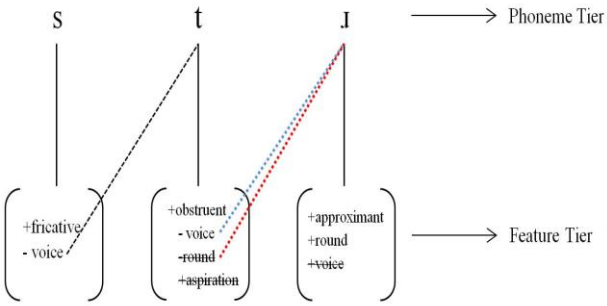
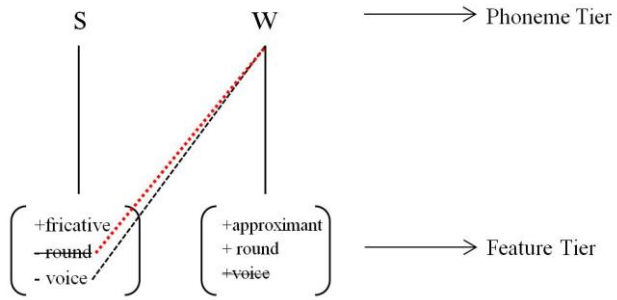
| Phonaestheme | Diagram showing phonetic feature(s) spreading across the phonaestheme | Explanation |
|---|---|---|
| /sn/ |  | Voiceless /s/ causes devoicing of succeeding /n/ (ibid:210) |
| /st/ |  | Voiceless /s/ causes deaspiration of succeeding /t/ (ibid:45; 171) |
| /stɹ/ |  | 1. Voiceless /s/ causes deaspiration of succeeding /t/ (ibid), shown by black line.<br>2. Voiceless /t/ causes devoicing of succeeding /ɹ/ (ibid:219), shown by blue line.<br>3. Anticipation of /ɹ/ causes rounding of preceding /t/ (Hoffmann, 1971), shown by red line. |
| /sw/ |  | 1. Voiceless /s/ causes devoicing of succeeding /w/ (Cruttenden, 2008:229), shown by black line.<br>2. Anticipation of /w/ causes rounding of preceding /s/ (Knight, 2012:135-136), shown by red line. |

Table Continues Overleaf

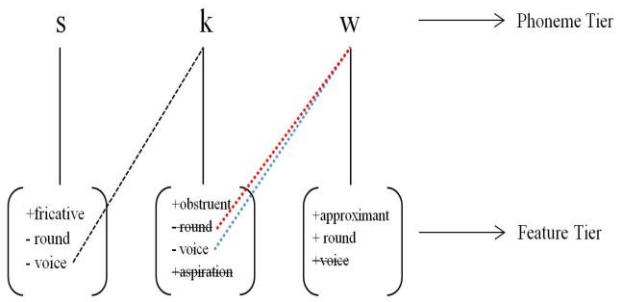
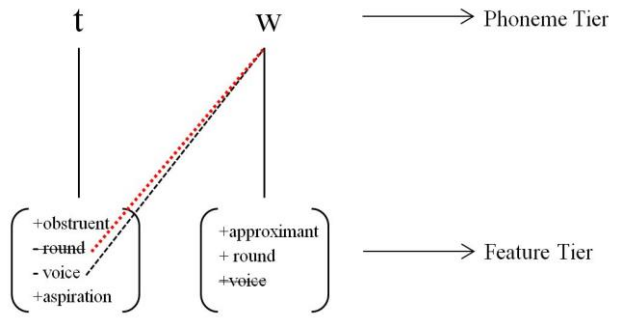
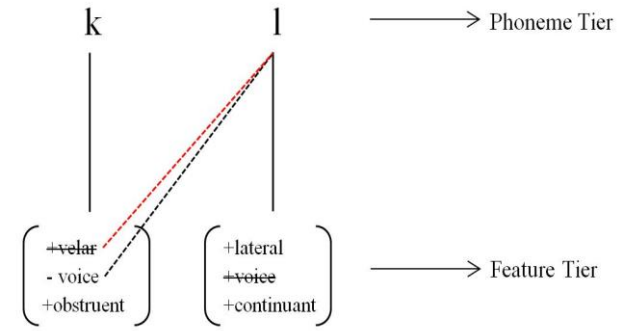
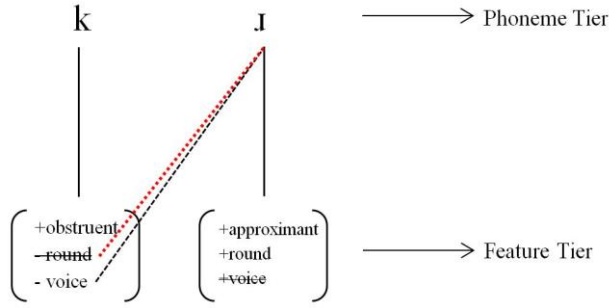
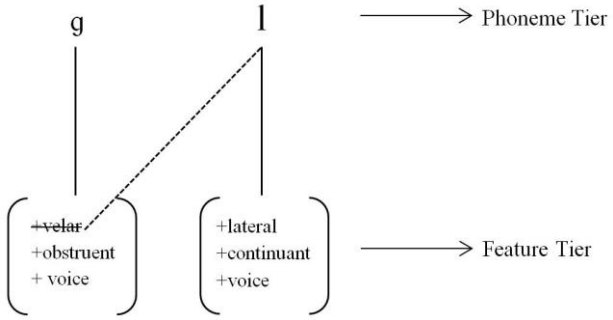
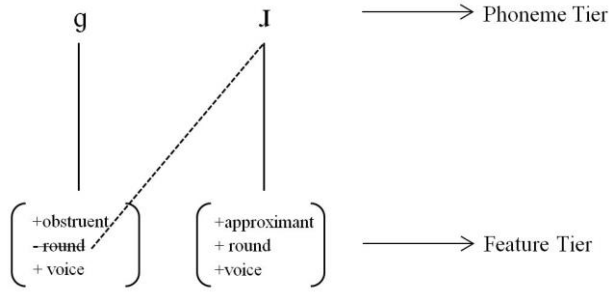
| Phonaestheme | Diagram showing phonetic feature(s) spreading across the phonaestheme | Explanation |
|---|---|---|
| /skw/ |  | 1. Voiceless /s/ causes deaspiration of succeeding /k/ (Cruttenden, 2008:175), shown by black line.<br>2. Voiceless /k/ causes devoicing of succeeding /w/ (ibid:228-9), shown by blue line.<br>3. Anticipation of /w/ causes rounding of preceding /k/ (Knight, 2012:135-136), shown by red line. |
| /tw/ |  | 1. Voiceless /t/ causes devoicing of succeeding /w/ (Cruttenden, 2008:228-9), shown by black line.<br>2. Anticipation of /w/ causes rounding of preceding /t/ (Knight, 2012:135-136), shown by red line. |
| /kl/ |  | 1. Voiceless /k/ causes devoicing of succeeding /l/ (Cruttenden, 2008: 214), shown by black line.<br>2. Anticipation of /l/ causes fronting (lateral release) of preceding /g/ (Cruttenden, 2008:175), shown by red line. |

Table Continues Overleaf

| Phonaestheme | Diagram showing phonetic feature(s) spreading across the phonaestheme | Explanation |
|---|---|---|
| /kɹ/ |  | 1. Voiceless /k/ causes devoicing of succeeding /ɹ/ (ibid:219), shown by black line.<br>2. Anticipation of /ɹ/ causes rounding of preceding /k/ (Hoffmann, 1971), shown by red line. |
| /gl/ |  | Anticipation of /l/ causes fronting (lateral release) of preceding /g/ (Cruttenden, 2008:175) |
| /gɹ/ |  | Anticipation of /ɹ/ causes rounding of preceding /g/ (Hoffmann, 1971) |

In sum, this section has discussed two theories; segmental phonology and prosodic phonology. Both theories ultimately agree that the speech signal is a binary construct, comprising the physical realisations of speech sounds and speakers' knowledge of the meaningful phonetic units in their language. However, the theories present differing views about how to analyse and interpret the meaningful phonetic units of a language. The phonemic ('segmental') approach argues that they are the minimal discrete phonetic units capable of creating meaning. But Firth (1948) argued that the speech signal did not lend itself to such discrete segmentation (O'Grady, 2013:55). He observed that the meaningful auditory structures in a language often extended beyond the boundaries of the phones realising these individual 'phonemes'. As such, he argued that it would be more appropriate to view the

meaningful phonetic units of the speech signal as auditory units that could extend to any length across the utterance. Firth termed these structures prosodies. The twelve phonaesthemes tested in the present study are treated as prosodies rather than sequences of phonemes. There are both semantic and phonetic grounds for this conceptualisation. Semantically, it is the phonetic sequences as gestalts that scholars have associated with particular connotations; not one or other of the phonaesthemes' constituent consonants in isolation. Phonetically, there are certain phonetic features which extend across the twelve phonaesthemes by way of assimilation. For example, in the /sl/ phonaestheme, the voicelessness of /s/ spreads into the following /l/, causing devoicing. This spreading of such features across the onset sequences makes it more appropriate to view the clusters as a single phonetic units extending in linear sequence (i.e. as prosodies); rather than as sequences of discrete phonemes.

Having outlined the necessary theoretical approaches – the notion of connotation and the notion of prosodic phonology – the next section provides a more detailed definition of phonaesthemes than has thus far been possible. The point of departure for this is to consider Firth's (1930) first use of the term.

### 2.3) Defining Phonaesthemes

Firth's first use of the term 'phonaestheme' is found in a discussion of 'phonetic habits' (Firth, 1930:180-186). For Firth, 'phonetic habits' were any kind of phonological behaviours or patterns learned or identified by a particular linguistic community by way of frequent exposure (1930:180). One type of phonetic habit Firth posited was the potential for certain syllable prosodies to recur in particular contexts of experience, and the connotations that could arise from speakers repeatedly experiencing this pattern. Firth reasoned that if a linguistic community was habitually exposed to a certain prosody co-occurring in a particular context of experience, the prosody could acquire an association with this context of experience within the community.

Firth demonstrated this type of phonetic habit through the /sl/-<<pejoration>> correlation in English. To begin, he considered the English adjective "slack" /slæk/. Firth pointed out that the denotation of this linguistic item – the quality of being relaxed or idle (OED: Online) – remained constant throughout all its derivational and inflectional forms;

'*slacks, slacked, slacking, slacken, slackness, slackish, slackly, slackest, slacker, slackers*' (1930:183). As such, he argued that the sequence /s/, /l/, /æ/ and /k/ functioned as a habitual phonetic cue, identifying lexical items as being related to this denotation. However, in addition, Firth (ibid:184) argued that the onset prosody /sl/ simultaneously marked the word as belonging to 'a much bigger group of habits…the sl phonaestheme'. It is here that the term 'phonaestheme' is seen used for the first time; a term defined as a **'cumulative suggestive value'** (ibid) of a phonetic sequence within a speech community – in other words, a connotation (cf. § 2.1). In the case of the onset prosody /sl/, Firth argued that this connotation was one of <<pejoration>>; on the basis that 'we hear and learn to make the sounds at present in what we may describe as pejorative contexts of experience' (ibid:185). Put simply, Firth's claim was that many English /sl/ words express meanings that are culturally <<unpleasant>> in some way, which seems to imbue onset /sl/ with a connotation of <<pejoration>> or <<unpleasantness>>. Firth listed over thirty-five examples in support of this claim, including *slack, slouch, slush, sludge, slime, slosh, slash, sloppy, sluggard, slattern, slut, slang, sly, slither, sloth, sleet, slip, slipshod, slit, slay, sleek, slant, slovenly, slab, slap, slough, slum, slump, slobber, slaver* and *slate* (ibid). To this list one could feasibly add *slag, slam, slander, slapdash, slapper, sleaze, slob* and *slurry* from contemporary English use.

It is worth noting that some of the exemplifying words provided by Firth appear to be more prototypically <<pejorative>> than others. Prototypicality (Rosch, 1970; Rosch and Mervis, 1975; Aitchison, 2012:69) is the idea that some lexical items are better exemplars of a particular semantic category or concept than others. For example, I would argue that "slut" is a clearer exemplar of a <<pejorative>> /sl/ word than "slack"; though both are in Firth's list of supporting words. This is because it seems difficult to imagine any non-<<pejorative>> context in which one could use the expression "slut"; the very denotation of this expression – 'a woman of dirty, slovenly, or untidy habits […] a foul slattern' (OED: Online) is <<pejorative>>. By contrast, "slack" may be used in the context of referring to the tension of a rope; or in a maritime context, referring to the conditions of a tide. However, even though some of Firth's supporting words are more prototypically <<pejorative>> than others, the important point is that each are related to the broad conceptual idea of <<pejoration>> in at least one of their senses. For example, "slack" can be used pejoratively to refer to a careless attitude or lack of effort, in addition to the two other contexts described above. Similarly, a "sloth" may refer to a creature of the rainforest, or a lazy and inactive person; while "slate" might geological mineral or an intense criticism. In reality, a similar

situation is seen with each of the phonaesthemes under study: Although the exemplifying words provided by scholars are all broadly related to a particular context of experience in at least one sense[2], some are more central (prototypical) exemplars of this context of experience than others. (For a full list of the exemplifying words provided by scholars for the twelve phonaesthemes, see Table 3 and Table 5).

To clarify, then; a phonaestheme is where a phonetic sequence frequently recurs in a particular context of experience within a speech community. As a result, the phonetic sequence acquires a connotation of this context within the community. For example, as a result of onset /sl/ habitually recurring in English words expressing broadly <<pejorative>> meanings, this prosody seems to have acquired a connotation of <<pejoration>> in English (Firth, 1930; Sturtevant, 1947; Bowles, 1995). Similarly, as a result of onset /gl/ habitually recurring in English words expressing meanings related to <<light>>, the prosody seems to have acquired a connotation of <<light>> in English (Marchand, 1966; Bowles, 1995; Sadowski, 2001), and so on. However, it is essential to note that there is nothing inherently meaningful about such sounds. The fact that onset /sl/ recurs in a number of words related to a <<pejorative>> context of experience in English does not entail that onset /sl/ is intrinsically <<pejorative>>, or that it would be associated with <<pejoration>> in non-English-speaking communities. In the words of Firth (1930:184-185), 'it just so happens that we hear and learn to make [such] sounds at present in [particular] contexts'. The same is true for all the phonaesthemes studied in the present thesis, and indeed any phonaestheme in any language of the world. It just happens that the twelve onset prosodies listed on pages 8-9 presently recur with particular contexts of experience (i.e. with the expression of particular meanings) in the vocabulary of English, and thus may have acquired connotations of these contexts in English-speaking communities (Firth, 1930:184-185). The importance of this point will become clear in § 2.4, below, when phonaesthemes are differentiated from other types of sound-meaning correlations.

Because phonaesthemes do not entail that any phonetic sequence is intrinsically meaningful, counterexamples to these patterns can (and do) exist without challenging the theory itself. For instance, while there is a clear recurrence between onset /sl/ and <<pejoration>> in English, there are also many English words bearing onset /sl/ that do not

[2] With the notable exception of the words provided by Firth in support of the /fl/ phonaestheme, as discussed further in § 2.6.1.

express <<pejorative>> meanings (e.g. ‘sleep’ and ‘sleigh’). Similarly, there are many English words bearing onset /stɹ/ that do not express ideas related to <<straightness and stretching>> (e.g. ‘strumpet’ and ‘strong’). Indeed, there are counterexamples to all twelve phonaesthemes under test (cf. Table 3). Firth acknowledges this possibility throughout his works, making it very clear that a phonaesthemic pattern need not necessarily extend to all words bearing the same phonetic form. Nevertheless, despite the possibility and indeed presence of such counterexamples, Firth suggests that any phonaestheme is likely to be ‘felt and observed’ (1935:45) throughout the speech community in whose language it is present. This is clearly evident from his use of the collective pronoun: ‘We are appreciably affected by [some] initial and final phone groups not ordinarily recognised as having any function’ (Firth, 1930:184). Notably, however, Firth does not indicate that he ever put this assertion to the test; whether in English or any other language. As outlined in Chapter 1, the present study furthers the work of Firth by seeking to do just this.

Before discussing the difference between phonaesthemes and other types of sound-meaning correlations discussed in the literature, it is worth expanding a little on the idea of recurrence. There is no indication in any of Firth’s works as to precisely how recurrent a form-meaning pattern would need to be in order to constitute a phonaestheme; nor whether a particular prosody could belong to two phonaesthemes if it recurred with more than one context of experience. The inference arising from this is that any instance where a prosody occurs twice or more with a particular meaning is potentially a phonaestheme. However, the more frequently a prosody recurs with one particular meaning, or the greater the degree to which a prosody recurs with one meaning relative to any other, the more likely it is that the prosody and meaning will acquire an association and a phonaestheme will be formed (cf. Usage-Based and Probabilistic theories, § 3.2). This is supported by Firth’s claim that ‘the more consistently similar sounds function in situations having similar affective aspect, the clearer [the phonaesthemic function]’(1930:185). Of course, conceptualising phonaesthemes in this way leaves open the possibility that certain prosodies could feasibly belong to multiple phonaesthemes. This possibility is important to the discussion given in Chapter 6, where the ideas termed “multiple semantic recurrence” and “multiple phonetic recurrence” are introduced as possible factors affecting the productivity of phonaesthemes. In essence, both these ideas revolve around the possibility that there is not necessarily an exclusive phonaesthemic relationship between one prosody and one meaning. One prosody could recur with multiple [different] meanings, just as one meaning could recur with multiple [different]

prosodies.

Having defined phonaesthemes in detail, the next section distinguishes this highly specific type of sound-meaning correlation from others cited in the literature. This discussion is vital for the reader, because although there are many types of sound-meaning correlation in human language, the terms referring to these relationships are often poorly defined and used interchangeably by scholars. Following this, § 2.5 discusses how phonaesthemes, unlike several of these other theories, are not contrary to the idea that the linguistic sign is arbitrary. This is largely because the theory of phonaesthemes does not dictate that any phonetic sequence is intrinsically meaningful. Indeed, if anything, phonaesthemes appear to provide further support for the idea of arbitrariness in language.

**2.4) Phonaesthemes and Other Sound-Meaning Correlations**

The idea that the phonetic structures of words are meaningful is a field which has historically provoked a large amount of discussion. Much of this has resulted in a disparate collection of terms; including terms for words whose sound structures are in some way 'meaningful', terms for specific types of meaning conveyed by segments or prosodies, and terms for the general notion of sound-sense correlations. Such terms include 'sound-symbolic', 'universal', 'natural', 'iconic', 'ideophonic', 'kinaesthetic', 'mimetic', 'imitative', 'onomatopoeic', 'phonaesthetic' and 'euphonic', as well as the 'phonaesthemic' relationships investigated in the present study. Not only are these terms often poorly defined, but they are often used interchangeably by scholars, causing inevitable difficulties for anyone seeking to make sense of where and how phonaesthemes fit into the general field of enquiry. For instance, Firth explicitly refuted the idea that phonaesthemes were 'sound-symbolic'; a concept he described as 'fallacy' (1951a:194). However, it appears that Firth may have interpreted the term 'sound-symbolic' as meaning an 'absolute universal' sound-meaning correlation (see definition below). In contrast, many contemporary scholars (e.g. Hinton, Nichols and Ohala, 2006) simply use the term 'sound-symbolic' as a label for any kind of sound-meaning correlation in any language of the world. Moreover, whereas some scholars (e.g. Jespersen, 1922a/b; Tabakowska, 1999) use the term 'iconic' to denote a specific type of sound-meaning correlation, other scholars (e.g. Anderson, 1998; Fischer, 1999:123) use the term to mean 'sound-symbolic' in the general sense used by Hinton et.al (2006). Clouding the issue further is the notion of phonaesthetics, whose similarity of name makes confusion with

phonaesthemes almost inevitable. To simplify the tangle of terminology for the reader, this section provides a short review of the other main types of sound-meaning relationships (besides phonaesthemes) cited in the literature. It also briefly summarises the relationship between phonaesthemes and these other theories; showing how phonaesthemes fit in the general field of enquiry.

To begin, Table 2 (overleaf) gives a brief overview of the way in which I interpret the other (i.e. non-phonaesthemic) types of sound-meaning relationships cited in the literature. Each of these terms is subsequently discussed in more detail.

Table 2: An overview of non-phonaesthemic sound-meaning correlations

| Sound-Meaning Relationship | Definition |
|---|---|
| **Sound-Symbolism** (cf. Hinton et.al, 2006; Abelin, 2009; Shinohara and Kawahara, 2010) | Any type of correlation between a linguistic sound and meaning |
| **Absolute Universal** (cf. Finegan, 2012) | A correlation between a linguistic sound and meaning shared by – and evident in – all languages worldwide |
| **Universal Tendency** (cf. Finegan, 2012) | A correlation between a linguistic sound and meaning that is very widespread throughout the world, but not necessarily shared by all languages worldwide |
| **Natural** (cf. Hinton and Bolinger, 2003) | A correlation between a linguistic sound and meaning that is dictated by humans' biology. As a result, such correlations should be shared by – and evident in – all languages worldwide. |
| **Iconism/Iconic** (cf. Fischer, 1999; Tabakowska, 1999; Radden and Panther, 2004) | A conceived similarity between the phonetic qualities or features of linguistic sound and a meaning, as perceived by a group of cognizant human beings |
| **Kinaesthesia** (cf. Anderson, 1998) | A correlation between the gestures required to produce a linguistic sound and the meaning it tends to express |
| **Ideophone/Ideophonic** (cf. Doke, 1935; McGregor, 2001; Voeltz and Kilian-Hatz, 2001; Dingemanse, 2008) | An iconic representation of an idea through linguistic sounds (i.e. an iconic word). The iconism could be as a result of the phonetic qualities of the constituent sounds, or as a result of the way in which the sounds are combined, or the pitch used in uttering the word. |
| **Mimetic** (cf. Hinton et.al, 2006) | When speakers attempt to represent a non-linguistic sound using the phonetic resources of their own language |
| **Imitative** (cf. Meier, 1999; Attridge, 2004; De Cuypere, 2008) | When speakers attempt to represent a non-linguistic sound using any sound producible by the human vocal tract |
| **Onomatopoeia/Onomatopoeic** (cf. Anderson, 1998; Hinton et.al, 2006) | Any representation of non-linguistic sounds through the medium of speech (whether mimetic or imitative sound-symbolism) |
| **Phonaesthetic** (cf. Crystal, 1995a; Crystal, 1995b; Crystal, 2006) | A beautiful or aesthetically-pleasing linguistic sound (the study of which is termed phonaesthetics) |
| **Euphonic/Euphony** (cf. Day, 1867; Myers-Shaffer, 2001) | The literary terms for phonaesthetic/phonaesthetics |

While some scholars, such as Fischer (1999:123) use the term **'sound-symbolic'** synonymously with 'iconicity' (defined below), many contemporary scholars (e.g. Hinton et.al, 2006; Abelin, 2009:9; Shinohara and Kawahara, 2010) use this label as an umbrella term for any instance of sound-meaning correlation in human language. This is the stance adopted by the present study. By this conceptualisation, phonaesthemes can be termed 'sound symbolic' relationships, given that they involve some sort of correlation between phonetic structure and meaning. Similarly, onomatopoeia, iconism, kinaesthesia and phonaesthetics are also sound-symbolic relationships.

In the context of sound-symbolic correlations, '**universal**' relationships are divisible into two types (cf. Finegan, 2012:248). '**Absolute universals**' are patterns found in all languages worldwide, without exception. Absolute universals are generally seen as a consequence of positing '**natural**' sound symbolic relationships; innate correspondences between sound and meaning dictated by, or arising from, humans' biological nature (Hinton and Bolinger, 2003:113). At the present time, there does not appear to be any evidence of any specific sound-meaning correlations that are absolutely universal, even though the potential for sound-symbolism may well be common to all human languages (Brown, 1958:139). Even Joseph Greenberg, 'successful in discovering [many] similarities among the languages of the world' throughout the twentieth century (Sereno, 2006:263), did not appear to have found any specific sound-meaning correlations that were shared by all languages worldwide (*ibid*). Nevertheless, there are some sound-meaning associations that are very widespread throughout human language (such as the iconism of the vowel [i] discussed below), even if they are not absolutely universal. As Firth (1935:45) notes, 'there are…certain very widespread correlations between sound and sense in a vague general sort of way'. Such relationships are known as '**universal tendencies**' (Finegan, 2012:248); patterns that hold in a very widespread way for a large number of languages, but which are not generalisable to all languages worldwide.

**Iconicity**, or **iconism** (Radden and Panther, 2004:16), refers to a specific type of sound-symbolic relationship. Iconism is a process of some conceived correlation between the phonetic qualities or features of a linguistic form and some worldly content 'as perceived by a cognizant human being' (Tabakowska, 1999:410). This form of relationship takes its name from the notion of icons (Peirce, 1873:140-150). An icon is any type of sign (i.e. any image, sound or gesture) which is 'in one or more respects, the same' as the concept or object it

represents (Lechte, 2008:172). For example, a photographic image is an exact representation of the scene or subjects it depicts. It is of course impossible for any linguistic form to be fully 'iconic', since, unlike the medium of photography, language does not allow for the reproduction of exact likenesses of worldly content (Fischer, 1999:124). However, a segment or prosody becomes iconic when a thinking human being decides that its phonetic features or qualities are *in some way* like some worldly content; that there is some sort of conceptual link between the two. For example, Fischer (1999:123-124) discusses how the sound [ʃː] may be used to represent the wind. The iconism here is not that [ʃː] sounds like the wind; this would be an onomatopoeic association (see below). Rather, the phonetic features of the sound and the worldly content <<wind>> are existentially alike: both are physical rushes of air. Some scholars (e.g. Magnus, 1999; 2001) argue that iconic correlations have a natural basis in human language, and are thus perceived by all humans. Grammont (1901), Jespersen (1922a; 1922b) and Sapir (1929) believed that this was true of the vowel [i]. The scholars observed that [i] recurred in words expressing <<smallness>> and related ideas throughout many languages of the world, and speculated that this reflected an absolute universal iconism (Anderson, 1998:93; cf. kinaesthesia, below). However, it is not necessarily the case that iconism has any natural or absolute universal basis. Many conceived likenesses between a sound and some worldly content are actually specific to particular linguistic communities (termed 'conventional' relationships; cf. discussion of arbitrariness, § 2.5). Such is the case with the [i] vowel. While the iconism posited by Grammont, Jespersen and Sapir appears to be very widespread in Indo-European language communities, scholars such as Kim (1977) and Diffloth (2006) have since demonstrated that it is not an absolute universal, since the vowel actually recurs with the meaning <<largeness>> in many Asian languages.

**Kinaesthesia** is the term used to refer to 'correspondences of meaning and the physical attributes of articulation' (Anderson, 1998:167). In kinaesthetic sound-symbolism, the gestures required to articulate one or more of the sounds in a word are seen to correlate with the word's meaning. This type of association appears to be one of the earliest recognised forms of sound-symbolism, having been discussed by Plato in the Cratylus dialogue. Throughout the narrative, Plato posited a number of kinaesthetic relationships; such as a correlation between the vowel /ɒ/ and the meaning <<roundness>> (cf. 2009 translation, p.132). He reasoned that the lip rounding required to produce this sound explained why the vowel tended to appear with a meaning of <<roundness>>; such as in "goggulon" (round) and "omicron" (the name for the grapheme "o") (ibid). Two millennia later, a number of

scholars speculated that there could be a kinaesthetic basis to the iconism of [i] claimed by Grammont (1901), Jespersen (1922a; 1922b) and Sapir (1929). Bolinger and Sears (1981:161) and Fónagy (2000:342-343) speculated that 'the smallness and thinness [associations] of /i/ could be related to the narrow channel between the tongue and hard palate in /i/ sounds' (Fónagy, 2000:342-343). A different view was proposed by Pinker (1995:167) and Roach (2001:42-43); who suggested that [i] may have acquired an iconic association with <<smallness>> due to the fact that producing the vowel requires the vocal cords to be contracted to their shortest and thinnest size. Both views seem questionable, however, as none of the scholars proposed a mechanism by which speakers could become conscious of the various articulatory gestures. As such, these explanations could simply be epiphenomena; attempts to see causal correlations where there are none (Hurford, 2012:673). For it is not necessarily the case that [i] is produced with contracted vocal cords and a narrow gap in the oral cavity *in order* to express <<smallness>>. Production of the vowel is only possible through these (and other) gestures, but the fact that the vowel tends to recur with the meaning <<smallness>> does not entail that any of these gestures are meaningful. Indeed, there are equally feasible kinaesthetic correlations between the gestures of [ɑ] and <<smallness>>, yet [ɑ] recurs with the meaning <<largeness>> in many languages of the world (Jespersen, 1922a; Sapir, 1929). These correlations include the fact that [ɑ] is produced with the tongue in a very low position in the mouth, making a very small – if any – intrusion into the oral cavity (Kreidler, 2004:48-55); and the fact that [ɑ] is one of the vowels requiring the smallest amount of tension in the vocal cords (ibid).

Like iconism and kinaesthesia, **ideophones** are specific types of sound-meaning correlations. Ideophones are iconic representations of ideas such as 'manner, colour, sound, smell, action, state or intensity' through linguistic sounds (Doke, 1935:118). Noted for being particularly common to African and Aboriginal languages (cf. Doke, 1935; McGregor, 2001; Voeltz and Kilian-Hatz, 2001; Dingemanse, 2008), ideophones are words whose phonetic structure, either in part or as a whole, is conceived [by the speech community] to bear a likeness to some property of that which they name or describe. Unlike iconic relationships, this likeness need not be between the worldly content and the phonetic qualities of the segments or syllable prosodies in the word. In ideophones, the iconism could be between the worldly content and features such as volume (intensity), or the particular order in which the sounds of the word are combined. For example, Dingemanse (2008) discusses the word pronounced /vɛlɛvɛlɛ/, in the Siwu language of Ghana. The meaning of this word is <<a

dizzy, giddy feeling in the body>>. The ideophonic relationship between the phonetic structure and meaning of this word is not difficult to see. Here, it is not the phonetic qualities or features of the segments that are 'like' the sensation of spinning rapidly, but rather the ordering of the segments. The phonetic structure of the word comprises a repeating cycle of a voiced fricative, short mid-front vowel, liquid and the same short mid-front vowel. (There is no intrinsic meaning, of course, to the particular segments used in this ideophone; the idea of cyclic repetition could just as easily be conveyed by a sequence such as /sɒɹɒsɒɹɒ/ or /lɪmɪlɪmɪ/.)

'**Mimetic**' is the label given to words or prosodies where speakers attempt to represent non-linguistic, environmental sounds through the phonetic resources available in their language system(s). Such environmental sounds may include those made by animals, sirens, machines and so forth (Hinton et.al, 2006:3). Because these sounds use the phonetic resources within speakers' language systems, 'they are only approximate…imitations of certain sounds' (Saussure, 1966:69), and are largely specific to individual languages. The bark of a dog, for example, is represented by different linguistic sounds in French ('ouaoua') and English ('woof-woof') (*ibid*; see also Shipley, 1977:4 and Durkin, 2009:127). Though it is a difference rarely elucidated throughout the literature, 'mimetic' sound-symbolism is distinct from **'imitative'** sound-symbolism. Imitative sound symbolism is where a speaker attempts to faithfully reproduce a non-linguistic sound (Meier, 1999:140), and may break the phonotactic rules of their language system to do so. In these cases, the resulting sound may be unlike any extant forms in the speaker's native language; the only limiting factor to imitative sound symbolism being the constraints of the human vocal tract (see discussion by Attridge, 2004:136 and Cuypere, 2008:108). In modern literary and linguistic theory, both mimetic and imitative forms of sound-symbolism are grouped together under the label '**onomatopoeia**'.

To draw one final distinction in terminology, it is notable that the study of phonaesthemes is quite different to the study of **phonaesthetics**, despite the similarities in name. Phonaesthetics refers to the idea of aesthetically-pleasing linguistic sounds; an idea more widely termed '**euphony**' (Day, 1867:222-224; Myers-Shaffer, 2000:171). The term 'phonaesthetic' is first seen in the correspondence of J R R Tolkien. In a letter to a correspondent (cited in Carpenter, Tolkien and Tolkien, 1981:175), Tolkien discusses how he used the phonetic forms that appeared 'most beautiful' to him when coining words for the Elvish language in *The Hobbit*, *The Lord of the Rings* and *The Silmarillion*. He recalls how

the Quenya dialect of Elvish 'might be said to be composed on a Latin basis with two other (main) ingredients that happen to give me 'phonaesthetic' pleasure: Finnish and Greek' (ibid). Much of the scholarly activity in this area is constituted by the research of Crystal (1995a; 1995b; 2006). For example, in 2006, Crystal studied the phonetic structures of the words perceived as 'most beautiful' by English speakers, citing the top fifty results of a survey undertaken by the British Council in 2004. From these results, Crystal attempted to generalise a set of phonetic characteristics that may be regarded by English speakers as "phonaesthetic" (i.e. aesthetically pleasing). This argument is somewhat contentious, though, as Jones (2003) notes: 'When we come to words which are said to be particularly beautiful, we are…on dangerous ground. People who assert that certain words are beautiful are almost invariably thinking of the meaning of the words and not of the sound' (cf. Gorrell, 1994:193). For the purpose of the present study, it is not important to discuss the intricacies of this particular debate; it is sufficient to appreciate that phonaesthemes and phonaesthetics are two very different types of sound-symbolic relationship, despite their similarity in name.

Before moving-on to discuss how phonaesthemes are compatible with the idea that the linguistic sign is arbitrary, it is worth summarising where and how the theory fits into the study of sound-meaning correlations at large. Phonaesthemes are a type of sound-symbolic relationship; they involve associations between linguistic sounds and meanings. Phonaesthemes do not necessitate that any linguistic sound is 'naturally' (intrinsically) meaningful to humans; and would seem unlikely to be absolutely universal associations (i.e. correspondences between sound and meaning perceived worldwide). This is because phonaesthemes are formed from patterns in the phonetic forms used to express particular meanings, and different languages tend to express the same meanings using different phonetic forms (Saussure, 1966:67-68; cf. § 2.5, below). However, this does not discount the possibility of phonaesthemes being universal tendencies; associations perceived on a very large scale but not universally. In essence, this is one of the possibilities that the present study tentatively explores (cf. Chapter 1; Chapter 4; Chapter 5). Unlike iconic or ideophonic sound-symbolic correlations, there need be no conceived similarities between the phonetic qualities, features or structure of a phonaestheme and the meanings with which it recurs. Moreover, there need be no conceived similarity between any of the gestures involved in producing a phonaestheme and the meaning with which it recurs, unlike kinaesthetic sound symbolism. Phonaesthemes are neither mimetic nor imitative, as they do not involve any attempt to reproduce sounds from the external environment (whether within the phonetic resources of a

speaker's language system or otherwise). Finally, phonaesthemes are a form of sound-symbolism distinct from phonaesthetics, because although some phonaesthemes may be perceived as 'beautiful' sounds by individuals or cultures, this need not be the case. It is largely irrelevant whether a phonaestheme is perceived to be an 'ugly' or 'beautiful' sound; what matters are the meanings with which the phonetic structure recurs. (Nevertheless, if a particular phonetic structure were to recur with a meaning that was culturally judged to be <<beautiful>>, there would of course be no reason why a <<beauty>> phonaestheme could not emerge.)

### 2.5) Phonaesthemes and Arbitrariness

It is clear, then, that phonaesthemes represent a type of sound-symbolic relationship distinct from others in the literature. It is particularly important to remember that phonaesthemes neither necessitate nor imply that any linguistic sound is innately meaningful in any way. Therefore, unlike 'absolute universal' or 'natural' theories of sound-symbolism, phonaesthemes are fully compatible with a view that the linguistic sign is arbitrary. The goal of the present section is to demonstrate this. The section begins by defining and explaining the idea that the linguistic sign is arbitrary. This is followed by a brief explanation of how 'absolute universal' and 'natural' theories of sound-symbolism are incongruent with this view; and how phonaesthemes, by contrast, are not. The section concludes by discussing how the general principle of arbitrariness does not discount the possibility of partially motivated sound-meaning relationships emerging over time within linguistic communities.

In order to appreciate the idea that the linguistic sign is arbitrary, one must first recognise that language is a semiotic system. A semiotic system is a system of signs, wherein 'meaning is created and meanings exchanged' (Halliday, 2003:2). In the early twentieth century, Saussure (1916/1966) proposed a model of the sign systems of human language whose principles still underpin modern linguistic theory (Burling, 2000:307; Wintle, 2002:467; Hinton et.al, 2006:1). Saussure saw language as a combination of two mutually-influencing systems. These he labelled 'la langue' and 'la parole'. 'La langue' referred to the 'words, syntax, rules, conventions and meanings' of language (Sanders, 2004:148; Strinati, 2004:81); the knowledge of a particular language system shared by individuals in a given speech community (Brown, 2013:34-35). 'La parole' referred to the specific instances of language use which expressed the meanings created in 'la langue'; such as the production of

speech acts or written texts (ibid). For Saussure, every sign in a language system – every entity capable of realising and exchanging meaning – embodied the union between langue and parole. Linguistic signs were seen as combinations of an idea or meaning (the ‘signified’) and a cognitive model of the [usually phonetic] form used to express this idea in the speech community (the ‘signifier’) (Saussure, 1966:65-70).

Saussure represented the duality of linguistic signs through an illustration similar to Figure 3, below. In this diagram, the upper level of the circle (sign) is a representation of the ‘signified’; the meaning or idea being conveyed (in this case, the idea <<star>>). The lower level is a representation of the ‘signifier’; the form used to express this meaning in a particular speech community (in this case, a representation of the phonological signifier for <<star>> in Welsh). The bordering circle illustrates the idea that both components are bound together forming one linguistic sign, while the dual-arrows refer to the two-way links between meanings and their forms of expression; the idea that the concept and meaning are ‘united…by an associative bond’ within a speech community, and that ‘each recalls the other’ (Saussure, 1966:65-66):

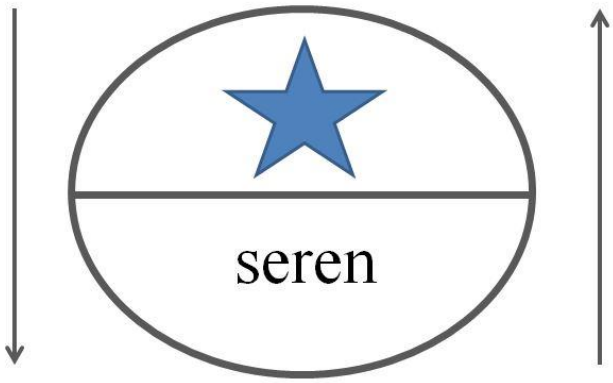



Figure 3: The Duality of the Linguistic Sign: Signified and Signifier (adapted from Saussure, 1966:67).

For any investigation concerning sound-symbolic relationships, the most significant aspect of Saussure’s theory is his view on how the signified and signifier relate to each other within the linguistic sign. Saussure famously asserted that the 'bond between the signifier and the signified is **arbitrary**' (1966:67, author's emphasis). In other words, there is nothing intrinsic about any meaning which necessitates that its form of expression must use a

particular phonetic structure. Nor is there anything intrinsic about any phonetic structure which dictates that it must – or should – be used in expressing a particular meaning (cf. Blank and Koch, 1999:21). According to Saussure, no phonetic form has any intrinsic meaning. A phonetic form is only meaningful through convention; when a linguistic community agrees that this form is to be used in the expression of a particular idea. As such, meanings and ideas have *'no natural connection'* with their forms of expression (Saussure, 1966:69, author's emphasis). To illustrate this, consider the concept <<sister>>. Saussure argued that there was nothing about the meaning <<sister>> which dictated that its phonological form of expression [in French] had to use the string of sounds /sœʁ/. In essence, this is proven by the fact that the same idea is expressed by very different phonetic sequences in different languages (e.g. /ˈsɪstə/ in English; /ˈɪrmə/ in Portuguese; /æˈdɛlˌfiː/ in Greek; /χwaɪr/ in Welsh, and so on). Indeed, Saussure argues that <<sister>> 'could be represented equally by just any other [form]' (ibid:67-68). In the words of Shakespeare; 'that which we call a rose by any other name would smell as sweet' (Shakespeare, 1599/2000:59).

Although Saussure's claims remain widely accepted in modern linguistic theory (cf. Hockett, 1960; Hinton et.al, 2006:1), many theories of sound-symbolism are seen to be in conflict with the idea that the linguistic sign is arbitrary (Ahlner and Zlatev, 2010:298). This is particularly true of theories claiming the existence of natural and absolute universal sound-meaning correlations (e.g. Magnus, 1999; 2001). By definition, such theories argue that there *is* a natural connection between certain meanings and their forms of expression, and that certain phonetic forms are *not* simply meaningful by virtue of convention. Consider the claims made by Magnus (1999). Magnus argued that every phoneme in every language of the world had a natural and universal iconism; a perceived likeness to some worldly content intrinsically dictated by human nature. For example, she argued that /s/ had a natural, absolute universal iconism of <<serpent>>. The basis for this claim was that ideas related to <<serpent>> were purportedly evident in 46% of English monomorphemic words bearing onset /s/ in her 'active vocabulary' (Magnus, 1999: Kindle Edition). There are a number of logical flaws in this argument; Magnus seemed to assume that the patterns in the vocabulary of English were representative of all languages worldwide, and her choice to study only the words in her 'active vocabulary' said nothing of the English words that were not in her active vocabulary. Nevertheless, Magnus made similar claims for a large number of [English] phonemes, and in her later (2001) study, claimed to have found speaker evidence of this natural, universal iconism; although the speaker cohort sampled were almost entirely English

speakers.

Unlike claims of natural and absolute universal sound-symbolism, phonaesthemes are fully compatible with the idea that the linguistic sign is arbitrary. This is because phonaesthemes involve prosodies acquiring connotations within a linguistic community as a result of their recurrence in particular lexical contexts (see discussion in § 2.3). As stated above, this neither implies nor necessitates that any linguistic sound is intrinsically (i.e. naturally) or universally meaningful in any way. In fact, the prosodies and meanings co-occurring in phonaesthemes actually *support* the idea that the linguistic sign is arbitrary. This is because there does not appear to be any natural reason why the prosodies in phonaesthemes would have been chosen to express the meanings with which they recur. For example, there does not appear to be any natural reason why onset /sl/ recurs with <<pejoration>> in English. As noted by Firth (1930:184-185), 'there is nothing inherently pejorative' about the sounds /sl/. A <<pejorative>> phonaestheme could just as easily have been formed with any other prosody, such as onset /pɹ/, nucleus /æ/ or coda /ʃ/. The same can be said for all twelve phonaesthemes under study: There does not appear to be anything inherently <<light>>-like about the prosody /gl/; nor anything inherently <<straight>> or <<stretchy>> about /stɹ/, and so on.

The fact that the prosodies of phonaesthemes do not appear to bear any natural correspondences to the meanings with which they recur (and thus appear to support the idea of linguistic arbitrariness), leads several scholars to question how the necessary conditions for phonaesthemes could have formed. Both Allott (1995) and Fischer (1999:129) question how it is possible to explain why certain prosodies recur in certain lexical contexts without recourse to some natural connection between form and meaning. One possible explanation could be that [originally] arbitrary sound-meaning relationships might develop diachronically into motivated patterns of form within a linguistic community. Imagine a native English-speaking community needed to coin a new word to express a meaning related to <<strength>>. It is possible or even likely (Bauer, 1983:294; Hassler, 2005:160) that the phonetic structure of this new word could be formed by analogy to existing English words with similar meanings. For example, the existing, arbitrary presence of onset /st/ in English words like "strong" and "steadfast" could be used analogously. In this situation, the relationship between the sound and meaning of the new word would have been partly *motivated* (i.e. influenced) by an existing form in the language (Fischer and Nänny, 2001:2;

Nöth, 2001:17-28; see further discussion below). Such formal analogies to existing [arbitrary] relationships could have occurred and recurred many times during the evolution of a language, and might explain the relatively consistent presence of certain prosodies in certain lexical contexts (Grace, 2005); a phenomenon that Bolinger (1965:245) and Wales (1990:339) term phonaesthemic 'constellations'. The underlying principle of arbitrariness would not conflict with the development of these partially motivated patterns of form, because there would still be no natural correlation between any of the prosodies and meanings in question (Bolinger, 1968:242, Hurford, 2012:121). This would also explain why there are almost certainly likely to be counterexamples to phonaesthemic patterns; instances where the form and meaning do not recur. The lack of a natural connection between the phonetic form and any given meaning would mean that the form could have been freely used in words expressing other meanings as well.

Notably, even Saussure (1966:131) discussed the possibility of motivated formal patterns emerging within a language system over time. He suggested that some forms within a language system may be chosen to express particular meanings because of the forms taken by existing signs within the system; even though these existing forms are arbitrary (i.e. do not bear any natural relationship to the meanings they express). By way of an example, Saussure (ibid) cited the French forms used to denote <<nineteen>> ("dix-neuf") and <<twenty>> ("vingt"). He argued that the form of "dix-neuf" was motivated, whereas the form of "vingt" was not. This is because the form of "dix-neuf" /ˈdi:z ˈnœf/ is the sum of the phonological forms denoting <<ten>> and <<nine>>. The meaning of this sign (<<ten plus nine>>) had therefore clearly motivated a formal analogy to existing signs within the language system[3]. But "vingt" (<<twenty>>) is not motivated; its form contains no analogy to either of the existing forms denoting <<two>> ("deux") or <<ten>> ("dix") in French. The same is true of the form used to express <<eleven>> in English: /ɪˈlɛvən/. The meaning of this sign is <<ten plus one>>, yet its form contains no analogy to either of the existing English forms denoting <<one>> or <<ten>>. It is therefore not motivated.

In sum, the present section has discussed how phonaesthemes, unlike natural and absolute universal theories of sound-symbolism (cf. § 2.4), are not contrary to Saussure's claim that the linguistic sign is arbitrary. In the early twentieth century, Saussure (1916/1966)

[3] A similar situation is evident in the English form 'nineteen'; a compound whose meaning motivates a formal analogy to the existing expressions 'nine' and 'ten'.

proposed a dyadic model of language which is still widely-accepted in linguistic theory today. In Saussure's model, every linguistic sign consisted of a union between an idea or meaning and the phonetic form used to express this meaning in a given linguistic community. Crucially, Saussure saw the union between ideas and their forms of expression as being arbitrary. He saw no natural reason why any phonetic form would be used above any other in expressing a given idea. For Saussure, no linguistic form was intrinsically meaningful. Meaning was only formed when a linguistic community agreed that a particular phonetic sequence would be used to express a particular idea. Theories of 'natural' and 'absolute universal' sound-symbolism (e.g. Magnus, 1999; 2001) appear to be incongruent with this view, because they argue that natural correspondences of form to meaning *do* exist in human language. However, phonaesthemes are fully compatible with the idea that the linguistic sign is arbitrary. This is because phonaesthemes neither necessitate nor imply that any linguistic form is intrinsically meaningful. A phonaestheme simply involves the recurrence of a particular phonetic form in a particular lexical context within a given linguistic community, leading to the form acquiring a connotation of this context within that community. Indeed, phonaesthemes actually support the idea of arbitrariness, because there does not appear to be any natural reason why the phonetic forms of phonaesthemes would have been chosen to express the meanings with which they recur. For example, there does not appear to be any natural reason why onset /sl/ recurs with the meaning <<pejoration>> in English. This leads some scholars (e.g. Allot, 1995; Fischer, 1999) to question why certain phonetic forms recur in certain lexical contexts if there is no natural correspondence between the two. One possible explanation could be the analogous use of existing arbitrary form-meaning correlations throughout the history of a language. This is a possibility that even Saussure (1966:131) considers; noting that the absence of any natural connection between form and meaning does not discount the possibility of arbitrary relationships motivating the forms of other linguistic signs by analogy.

### 2.6) Discussion in the Literature of the Twelve Phonaesthemes Under Test

Having now distinguished phonaesthemes from other types of sound-symbolism, and explained how phonaesthemes are fully compatible with a view that the linguistic sign is arbitrary, this final section in the present chapter reviews the support in the literature for the twelve patterns under test. As noted above (§ 1.2), there is by no means extensive support or discussion of these patterns in the literature (although the /gl/-<<light>> phonaestheme has

been awarded more substantial coverage than the other eleven patterns, as discussed below). The majority of these associations are discussed by just two to three scholars in the field, while several are only identifiable in the work of Firth. Notably, the limited existing works to have studied these patterns are entirely vocabulary-based. Using dictionaries, thesauri and in some cases their own intuition, the authors of these works examined the existing vocabulary of English and formulated lists of words to demonstrate the presence of one or more phonaesthemes in the language. Many of these word lists are at best anecdotal. Only Bowles (1995), Sadowski (2001) and Crystal (2002) attempted to quantify the presence of any phonaesthemes in English. With varying degrees of detail, each of these scholars calculated the number and proportion of English words featuring specific onset prosodies which also expressed particular meanings. Bowles (1995) studied a variety of phonaesthemic patterns in detail, a number of which are explored further in the present thesis. Sadowski (2001) restricted his enquiry to the /gl/-<<light>> phonaestheme, while Crystal (2002) studied only the /sl/-<<pejoration>> phonaestheme. However, no scholar has yet investigated the extent to which any of the twelve prosodies recur with the meanings under test *relative to any other meanings*; either in English or any other language. Moreover, as noted in Chapter 1, there is currently no evidence to suggest whether English speakers (or indeed speakers of other languages) might be aware of these patterns, and certainly no established methodologies to implement or draw-on in order to address this point (cf. § 1.4.1). It is for this reason that the present study remains very much exploratory; and that, in addition to tentatively investigating speaker perception of the phonaesthemes in English, French and Polish, it also reviews the vocabularies of these languages systematically and thoroughly, so as to investigate more fully the extent of any empirical evidence for these patterns.

This section takes as its starting point the two publications in which Firth (1930; 1935) discussed onset phonaesthemes. In all, he identified thirteen word-initial phonaesthemic patterns in the vocabulary of English. Eleven of these are studied in the present thesis (cf. § 1.2.2). These include the /sl/, /sm/, /sn/, /sw/, /st/, /stɹ/, /kl/, /kɹ/, /gɹ/, /skw/ and /tw/ phonaesthemes. The two remaining phonaesthemes, /fl/ and /dɹ/, are not studied in the present thesis, because they are very weakly-supported (and, in the case of /fl/, contentious) associations, as is discussed below. Following the review of Firth's works, the section then discusses the limited number of subsequent works in the field, including Sturtevant (1947), Marchand (1966), Bowles (1995), Sadowski (2001), Crystal (2002), Reay (2006) and Williams (2014). Several of these contributions provide valuable support for one

or more of Firth's eleven phonaesthemes under test. Many also discuss the twelfth phonaestheme under test; the association between /gl/ and <<light>>. While not a phonaestheme identified by Firth, the extent to which this pattern is cited in the literature – at least by comparison to any other phonaestheme – suggests it might be a widely-perceived association in English, and thus worthy of research. The section concludes by summarising the various scholars to have discussed, supported and cited each of the phonaesthemes explored in this thesis.

### 2.6.1) Firth (1930; 1935)

Over the course of two publications, Firth (1930; 1935) hypothesised thirteen onset phonaesthemes in English. These are summarised in Table 3, overleaf. From left to right, the columns in this table show: (i) the thirteen onset prosodies; (ii) the meanings with which they recur in English; (iii) a sample of words that exhibit each phonaestheme according to Firth, and (iv) a random sample of words that I have chosen to illustrate counterexamples to each phonaestheme. In some cases (indicated by purple cells), Firth explicitly described the meaning with which an onset recurred. Such was the case with the /sl/-<<pejorative>> phonaestheme (cf. Firth, 1930:184-186). In other cases (indicated by pink cells), Firth simply provided a list of words which he claimed demonstrated the phonaestheme; leaving the reader to infer the recurring meaning. In the majority of cases, this is relatively straightforward. However, the lack of an explicit description of the meaning association causes problems in interpreting the /fl/ phonaestheme, as is discussed below.

Table 3: Firth's English Phonaesthemes

| **Prosody (publication in which cited)** | **Associated Meaning** | **English words cited by Firth that exhibit this tendency** | **Counterexamples in English** |
|---|---|---|---|
| /sl/ (1930:184-185) | Pejoration | slack, slouch, slush, sludge, slime, slosh, slash, sloppy, sluggard, slattern, slut, slang, sly, slither, slink, slip, slipshod, slit, slay, slant, slovenly, slap, slough, slum, slump, slobber, slaver, slur, slog, slate | Sleep, slim, slender, |
| /sm/ (1930:185) (1935:44) | Pejoration | Smut, smudge, smash, smatter, smirch, smirk, smug | Smile, smooch |
| /sn/ (1930:185) (1935:39) | Pejoration | Sneak, snatch, snip, snap, snag, snub | Snow, snooker |
| /sw/ (1930:186) | Fast, or strong [predominantly downward] movement | Swill, swish, swim, sweep, swipe, swoop | Sweat, sweet |
| /stɹ/ (1930:185) (1935:44) | Straightness and/or stretching | Straight, stresses, strains, strength, stretched, stripe, stride, strive, struggle, streak, string, strenuous, stream, strip, strap | Strumpet, strategy |
| /st/ (1930:187) (1935:44) | Strength (possibly metaphorically extending to stillness and/or stoicism) | Stand, stiff, staff, steep; stick, still, stud, stump, stem, stalk, stoke, stare, stay, stain | Stoop, status, stupid, stupor |
| /dɹ/ (1930:187) | Hanging or falling downwards | Drip, drop, droop | Drive, dream, druid |
| /fl/ (1935:39) | (Unclear) | Flick, flake, fluke | (Unknown - meaning cannot be inferred) |
| /kl/ (1935:44) | Clumsiness or stupidity; also thickness (of density) | Clay, clot, clod, cloy, clog, clinker, clump, clumsy, cling, clench, clinch, clamp, clasp | Clear, clue, clip |

Table Continues Overleaf

| Prosody (publication in which cited) | Associated Meaning | English words cited by Firth that exhibit this tendency | Counterexamples in English |
|---|---|---|---|
| /gɹ/ (1935:44) | Gripping | Grip, grasp, grab, grope, grapple | Green, grass, grand, great |
| /kɹ/ (1935:44) | Crookedness | Crank, cross, criss-cross, crick, crab, cramp, crumple, crag, crook, crazy, crimp, cringe, cripple | Crumb, create |
| /skw/ (1935:44) | Weakness; possibly extending to cuteness | Squeeze, squelch, squirm, squirt, squib, squeal, squid, squander, squeamish | Squint |
| /tw/ (1930:186) | Smallness | Twitch, twist, twine, tweak, twinge, twinkle, twiddle, twaddle | Twang, twelve |

As noted above, it appears that Firth may have been less convinced about the /dɹ/ and /fl/ phonaesthemes than the other associations he posited. Firth provided only three exemplar words in support of these phonaesthemes; markedly fewer than for any other onset, as shown in Table 1 (above). If, after studying the vocabulary of English, Firth could find just three words in support of these phonaesthemes, it seems unlikely that these associations would be widely-perceived in English. Moreover, it is questionable whether the three words that Firth provides to demonstrate the /fl/ phonaestheme (‘flake’, ‘fluke’ and ‘flick’) actually have any meaning in common. It has been subsequently suggested that onset /fl/ recurs in words expressing <<movement>> in English (as typified by words like ‘fly’, ‘flee’, ‘flit’, ‘flicker’ and ‘flounce’); as well as in other Germanic languages (cf. Sturtevant, 1947; Knowles, 1987: Ch.5; Liberman, 1990; Abelin, 1999; Allan, 2001:136). But while this interpretation could explain ‘flick’ and even ‘flake’, it could not explain Firth’s claim that ‘fluke’ exhibits the /fl/ phonaestheme. As such, it seems that Firth may have envisaged a different meaning connection among /fl/ words; one which is not recoverable from his choice of exemplifying words and lack of an explicit description of the phonaestheme. As a result of Firth’s less convincing evidence for the /dɹ/ and [particularly] /fl/ phonaesthemes in comparison to the other associations he posited, I decided to exclude these phonaesthemes from the present study, and focus instead on testing the remaining eleven. These include the /sl/, /sm/, /sn/, /sw/, /st/, /stɹ/, /kl/, /kɹ/, /gɹ/, /skw/ and /tw/ phonaesthemes, as shown above.

### 2.6.2) Sturtevant (1947)

After Firth's work in the 1930s, the next scholar to have discussed phonaesthemic patterns in English was Sturtevant (1947:111). Using a numbered diagram (removed for purposes of copyright), Sturtevant posited the following word-initial phonaesthemes in the vocabulary of English (cf. Table 4, below):

Table 4: Sturtevant's English Onset Phonaesthemes

| **Onset Prosody** | **Semantic association** | **Examples of English words exhibiting this pattern** |
|---|---|---|
| /b/ | Explosive action | bash, biff, bang, boom, bong |
| /kl/ | [Sudden] loud noise | clatter, clamour, clang, clunk |
| /kɹ/ | Sudden, short loud noise | crack, crash, crunch |
| /fl/ | Fast movement [mostly of light] | flicker, flitter, flash, flare, flame, flounce |
| /sl/ | Pejoration/unpleasantness | slash, slap, slip, slide, slop, slush |
| /gl/ | Light | glare, gleam, glow, glimmer, gloom |

From his diagram, it appears that Sturtevant observed the same correlation between /sl/ and <<pejoration>> previously cited by Firth: Sturtevant provides a list of /sl/ words which are all broadly <<pejorative>>. These include 'slash', 'slap', 'slide', 'slip', 'slop' and 'slush'. In addition, Sturtevant's diagram also provides support for the /gl/-<<light>> phonaestheme; he clearly shows the recurrence of onset /gl/ in words related to <<light>> (cf. 'glare', 'gleam', 'glow', 'gloom' and 'glimmer'). However, while Sturtevant's account supports the /sl/ and /gl/ phonaesthemes under test, his diagram also implies different connotations for onsets /kl/ and /kɹ/ than those discussed tested in this thesis. Whereas Firth's /kl/ words suggest a meaning of <<clumsiness, thickness or stupidity>>, the /kl/ words cited by Sturtevant imply that /kl/ also recurs with a meaning of <<reverberating loud sound>>. And whereas Firth claims that /kɹ/ recurs in words expressing <<crookedness>>, the words cited by Sturtevant imply that /kɹ/ also recurs with a meaning of <<sudden, short loud sound>>. These differences of opinion between the scholars suggest that onsets /kl/ and /kɹ/ might recur with multiple [different] meanings in the vocabulary of English (cf. discussion of 'bi-

phonaesthemes' below).

### 2.6.3) Marchand (1966)

Following Sturtevant's work, phonaesthemes received little discussion in the literature until the 1960s, when Marchand (1966:315-335) proposed a large inventory of onset and rime phonaesthemes in English. So extensive was this inventory that it is beyond the scope of the present section to review every pattern he discussed (for this reason, a summary table of Marchand's proposed onset phonaesthemes is provided in Appendix 1[4]). However, although his account was extensive, Marchand did not attempt to quantify the presence of any phonaestheme in the vocabulary of English. Rather, he simply provided lists of words to illustrate the patterns he proposed, and a number of counter-examples to demonstrate that these patterns did not extend to all words bearing the particular prosodies. Nevertheless, within his account, Marchand cited four of the same phonaesthemes tested in the present study. These included:

1. the /st/-<<strength and stoicism>> phonaestheme, as typified by 'stand, 'step', 'steady' and 'stead';
2. the /kɹ/-<<crookedness; curvilinearity>> phonaestheme, as typified by 'creak', 'crook' and 'crank';
3. the /sw/-<<fast or strong movement>> phonaestheme, as typified by 'swing', 'sweep' and 'swipe';
4. the /tw/-<<smallness>> phonaestheme, as typified by 'twitter', 'twiddle' and 'tweet'.

In addition, Marchand also posited different meaning associations for six of the onsets under test. These included:

1. /stɹ/, which Marchand associated with the meaning <<stepping and striding>>, rather than the <<straightness and stretching>> under test. This he exemplified through words such as 'stride', 'straddle' and 'stroll'.

[4] All Appendices of this thesis are included as .pdf files on the attached CD-Rom.

2. /skw/, which Marchand associated with the meaning <<discordant sounds>>, rather than the <<weakness; cuteness>> under test. This he exemplified through words such as ‘squeak’ and ‘squall’.
3. /sn/, which Marchand associated with the meaning <<facial gestures>>, rather than the <<pejoration>> under test. This he exemplified through words such as ‘snarl’, ‘sniff’ and ‘snigger’.
4. /sl/, which Marchand associated with the meaning <<falling and sliding movements>>, rather than the <<pejoration>> under test. This he exemplified through words such as ‘slide’, ‘slip’, ‘slouch’ and ‘slump’.
5. /kl/, which Marchand, in common with Sturtevant (1947), associated with the meaning <<loud sound>>, rather than the <<clumsiness, thickness and stupidity>> under test. This he exemplified through words such as ‘clatter’, ‘clamour’ and ‘clang’.
6. /gɹ/, which Marchand associated with the meaning <<growling>>, rather than the <<gripping>> under test. This he exemplified through words such as ‘growl’, ‘grunt’ and ‘grumble’.

In common with Sturtevant’s (1947) claims regarding /kl/ and /kɹ/, these differences of opinion suggest that /stɹ/, /skw/, /sn/, /sl/ and /gɹ/ might recur with other meanings in the vocabulary of English besides those discussed above (cf. ‘bi-phonaesthemes’, below; see also Chapter 6).

### 2.6.4) Bowles (1995)

A number of the phonaesthemic patterns under test were given a limited amount of empirical support in the work of Bowles (1995). Bowles’ study was arguably the most thorough investigation into English phonaesthemes at the time of writing this thesis. Using the Longman Active Study Dictionary of English, Bowles undertook a vocabulary-based investigation into the meanings most frequently conveyed by words featuring twenty-two different onset prosodies and fifty-two different rime prosodies. The findings of this investigation appeared to show evidence for nine of the twelve onset phonaesthemes under test. These included:

1. /sw/-<<fast or strong movement>> (63% of all /sw/ words sampled);
2. /sm/-<<pejoration>> (50% of all /sm/ words sampled);

3. /sl/-<<pejoration>> (49% of all /sl/ words sampled);
4. /stɹ/-<<straightness and stretching>> (48% of /stɹ/ words sampled);
5. /gl/-<<light>> (46% of /gl/ words sampled);
6. /gɹ/-<<gripping>> (44% of /gɹ/ words sampled);
7. /kɹ/-<<crookedness; curvilinearity>> (43% of /kɹ/ words sampled);
8. /tw/-<<smallness>> (35% of /tw/ words sampled);
9. /st/-<<strength and stoicism>> (26% of /st/ words sampled).

Over a quarter of the words in the dictionary bearing each of these onsets also conveyed the meanings under test in the present study. For two of the onsets (/sw/ and /sm/), the proportion of words conveying the meanings tested in the present study was fifty percent or more. While such findings do not guarantee or even imply the possibility of speaker perception, they do at least suggest some degree of empirical (quantifiable) evidence in the vocabulary of English for a number of the phonaesthemes under study. However, even these detailed findings do not provide a complete picture of the extent to which the phonaesthemes are present in the vocabulary of English. For instance, Bowles' findings do not reveal how the above proportions compare to any other meanings with which these onsets might have recurred (cf. Chapter 1). By way of an example, consider /tw/. Bowles found that thirty-five percent of /tw/ words sampled recurred with <<smallness>>. But how did this compare to any other meaning(s) which recurred with onset /tw/? How much more frequently did /tw/ recur with <<smallness>> than with any other meaning?

In addition to the above findings, Bowles also posited different meaning associations for three of the onsets under test. First, he observed that onset /skw/ correlated with <<brief, unexpected movement>> (56% of /skw/ words sampled), rather than the meaning <<weakness and cuteness>> under test. In common with Sturtevant and Marchand, he also observed that onset /kl/ correlated with the meaning <<loud sound>>, rather than the association of <<clumsiness, thickness or stupidity>> under test; and that onset /sn/ tended to correlate with the meaning <<nasal/facial gestures>> (the same correlation observed previously by Marchand). However, Bowles appears to have been the first scholar to explicitly discuss the possibility that some prosodies might recur with multiple [different] meanings, and thus acquire two or more different connotations within a speech community (cf. § 2.3). He observed that <<loud sound>> was not the only recurrent meaning among words bearing onset /kl/; and that <<nasal/facial gestures>> was not the only recurrent

meaning among words bearing onset /sn/. Rather, he noted that /kl/ also recurred with the meaning <<closing movement>> (as in 'clam', 'clamp', 'clap', 'clasp' and 'clench'); and that /sn/ also recurred with the meaning <<pejoration>> as tested in the present study. Bowles referred to such prosodies as 'bi-phonaesthemes' (1995:103-104; cf. discussion of 'multiple semantic recurrence' in Chapter 6).

### 2.6.5) Crystal (2002)

As part of a larger discussion of verbal art and "language play", Crystal (2002:124-126) provided a brief discussion of the /sl/-<<pejoration>> phonaestheme in English. Using a similar method to Bowles (1995, discussed above), Crystal analysed the /sl/ head lexemes in an unspecified English dictionary, and produced a chart illustrating which of these lexemes seemed to connote a negative association in his opinion as a native English speaker, and which did not (i.e. either connoted a relatively netural or positive association). Using this chart, Crystal then counted and compared the number of /sl/ lexemes with negative and neutral/positive connotations. He argued that 'there are at least 41 [/sl/] words which have at least one sense with 'negative' associations, and 27 which have none' (ibid:124). As a result, Crystal he suggested that English /sl/ words 'are twice as likely' to carry negative than non-negative connotations (ibid.).

It is difficult to determine the extent to which these claims are valid or reliable, since Crystal omitted a number of important details about the methodology of his study. First, he did not specify which particular dictionary he sampled. This makes it difficult to determine how comprehensive his list of /sl/ lexemes might be. For instance, if Crystal had sampled a secondary school English dictionary, the number of /sl/ head lexemes would likely have been smaller (and thus less representative) than if he had sampled the full OED. In addition, Crystal did not provide any details about how he interpreted negative or non-negative connotations. On what criteria did he judge whether an item carried a negative connotation or not? It is for this reason that the present study provides the list of criteria on which I judged whether each phonaestheme was exhibited (or not) in the vocabulary of English, French and Polish (see discussion in Chapter 4.1.1). This ensures that the investigation could be replicated in the future if required. Nevertheless, despite these limitations, Crystal's findings indicate a degree of further support for the /sl/-<<pejoration>> phonaestheme in English, and

do not appear to disagree with Bowles' (1995) finding that roughly half the English /sl/ words he sampled expressed something <<pejorative>> or <<unpleasant>>.

### 2.6.6) Reay (2006)

One of the most recent discussions of English phonaesthemes can be found in the work of Reay (2006:893-901). Using schematic similar to the one used by Sturtevant (1947), Reay hypothesised seven onset and five rime phonaesthemes in English. These are summarised in Table 5, below:

Table 5: Reay's (2006) English Onset Phonaesthemes

| Onset Prosody | Semantic association |
|---|---|
| /sk/ | 'One who shirks work or duty' (scab, skiver, scoundrel etc.) |
| /sk/ | 'Light movement' (scamper, scurry, scoot etc.) |
| /sl/ | 'Sliding movement' (slither, slide, slip etc.) |
| /d/ | 'Dull, stupid' (dopey, daft, drippy etc.) |
| /sw/ | 'Movement through air or water' (swing, sweep, swoop etc.) |
| /kl/ | 'Clinging action' (cling, clamp, clasp etc.) |

Three of these onset phonaesthemes are of note to the present study. First, it is clear that Reay supports the association between /sw/ and <<movement>> tested in this thesis. The present author expressed the recurring meaning of English /sw/ words as <<fast or strong movement>> from the exemplar words provided by Firth (see Table 1); these included 'swill', 'swish', 'swim', 'sweep', 'swipe' and 'swoop'. Reay expresses the association slightly differently as <<movement through air or water>>, but it is clear that the two inferences are congruent: both involve the connotation of <<movement>>, and by definition, movement requires an elastic medium (such as water or air) in order to take place. Also evident in Reay's diagram are /kl/-<<clinging action>> and /sl/-<<sliding movements>> phonaesthemes, which present different meaning associations for /kl/ and /sl/ than the <<clumsiness, thickness and stupidity>> and <<pejorative>> connotations tested in the present study.

### 2.6.7) Sadowski (2001) and Williams (2014) on the /gl/-<<light>> phonaestheme

Before summarising the scholarly support for the twelve phonaesthemes explored in this thesis, it is worth noting that the /gl/-<<light>> phonaestheme has been discussed to a markedly greater extent than the other eleven patterns under test; both in the sound-symbolism literature and by linguists more generally. Onset /gl/ is by far the most widely-cited of all English phonaesthemes, and the correlation that is most frequently used to demonstrate the concept. In addition to its discussion by Sturtevant (1947) and Bowles (1995), this pattern is also cited in the work of Bolinger (1950:117-118), Knowles (1987: Ch.5), Short (1996:119-120), Bauer (1998), Sadowski (1999; 2001), Pearce (2007:172), Liberman (2009:36) and Williams (2014). The most detailed accounts of the phonaestheme are given by Sadowski (2001) and Williams 2014. Williams (2014) documented the /gl/-<<light>> words present in Middle English vocabulary, and analysed their use and significance in the contemporary Christian poem 'Pearl'; a presence first discussed by Sadowski (1999) some fifteen years earlier. However, it is Sadowski's (2001) article that is most relevant to the present thesis, as it provides the most extensive account of the /gl/ phonaestheme in modern-day English. Sadowski compared the number of /gl/ root words expressing meanings related to <<light>> in Old English and Modern English. His findings (2001:78) indicate that forty percent of the /gl/ root-words in the unabridged 1989 OED expressed a meaning related to <<light>>; a figure that appears roughly congruent with the forty-six percent observed in the Longman dictionary by Bowles (1995; see above). This figure included words expressing physical illumination and brightness; words expressing ideas which presuppose the presence of light (such as sight and reflection); words expressing the absence of light; and words expressing ideas of metaphorical illumination (i.e. joy or splendour). These are illustrated in Table 6, overleaf:

Table 6: Examples of English /gl/ words expressing meanings related to <<light>>, as observed by Sadowski (2001:76)

| **Meaning related to <<light>>** | **Examples of English /gl/ words observed by Sadowski which express this meaning** |
|---|---|
| Physical light; illumination; brightness | glare, gleam, glimmer, glint, glisten, glitter, glow |
| Ideas which presuppose the presence of light (sight; reflection) | glance, glare, glower, glossy |
| Absence of light; darkness | gloaming; gloom |
| Joy/Splendour | glad, glee, glamour, glory |

### 2.6.8) Summary of Scholarly Support for the Twelve Phonaesthemes Under Test

To conclude the review of the scholarly support for the twelve phonaesthemic patterns under investigation, Table 7, overleaf, summarises the various works to have discussed these particular correlations. The table shows that the /gl/-<<light>> phonaestheme is clearly the most frequently cited throughout the literature, while the /kl/-<<clumsiness, thickness and stupidity>> and /skw/-<<weakness; cuteness>> phonaesthemes are the least frequently-cited, being only identifiable in the work of Firth. This is a point to which the discussion returns in Chapters 4, 5 and 6.

Table 7: A summary of the scholarly support for the twelve phonaesthemes under test

| PROSODY | ASSOCIATED MEANING | SCHOLARS TO CITE PHONAESTHEME |
|---|---|---|
| /gl/ | Light | Sturtevant (1947); Bolinger (1950); Knowles (1987); Bowles (1995); Short (1996); Bauer (1998); Sadowski (1999; 2001); Pearce (2007); Liberman (2009); Williams (2014) |
| /sw/ | Fast/strong movement | Firth (1930); Marchand (1966); Bowles (1995); Reay (2006) |
| /sl/ | Pejoration | Firth (1930); Sturtevant (1947); Bowles (1995); Crystal (2002) |
| /st/ | Strength and stamina | Firth (1930; 1935); Marchand (1966); Bowles (1995) |
| /kɪ/ | Crookedness | Firth (1935); Marchand (1966); Bowles (1995) |
| /tw/ | Smallness | Firth (1930); Marchand (1966); Bowles (1995) |
| /stɪ/ | Straightness; stretching | Firth (1930; 1935); Bowles (1995) |
| /sm/ | Pejoration | Firth (1930); Bowles (1995) |
| /sn/ | Pejoration | Firth (1930); Bowles (1995) |
| /gɪ/ | Gripping | Firth (1930); Bowles (1995) |
| /kl/ | Clumsiness; thickness; stupidity | Firth (1935) |
| /skw/ | Weakness; possibly cuteness | Firth (1935) |

## 2.7) Literature Review: Conclusion

The purpose of this chapter was to define and conceptualise phonaesthemes. The point of departure for this was to distinguish the different types of meaning that may be conveyed by linguistic forms: denotation, connotation, sense and reference. It was necessary to distinguish these different types of meaning in order for the reader to fully understand phonaesthemes, since the meanings conveyed by phonaesthemes are entirely connotative. Following this, the chapter then defined and explained the concept of prosodic phonology. This discussion was also vital for the reader, since the conceptualisation of phonaesthemes in the present thesis

presupposes a prosodic view of the speech signal. Having outlined the necessary prerequisites, the third section of the chapter gave a detailed definition of phonaesthemes as first outlined by Firth (1930). This very particular type of sound-meaning correlation was then differentiated from other types of 'sound-symbolism' in the literature; including 'universal', 'natural', 'iconic', 'kinaesthetic', 'ideophonic', 'onomatopoeic' and 'phonaesthetic' (euphonic) correlations. Following this, the fifth section of the chapter discussed how phonaesthemes, unlike several other types of sound-meaning correlations, are not contrary to Saussure's (1916/1966) idea that the linguistic sign is arbitrary. If anything, they actually appear to provide further support for this idea. The chapter culminated in a sixth and final section which discussed the limited support in the existing literature for the twelve phonaesthemes explored in the present study. The scholars to have discussed one or more of these patterns include Firth (1930; 1935), Sturtevant (1947), Marchand (1966), Bowles (1995), Sadowski (2001), Reay (2006) and Williams (2014).

The next chapter in this thesis motivates the particular methodologies used to address the research questions posed in Chapter 1. It begins by discussing the evidence which suggests that phonaesthemes could become productive in language. In essence, the idea underpinning all the methodologies in this thesis is that phonaesthemes are potentially perceptible associations which can become productive. The chapter demonstrates how phonaesthemes are congruent with at least two views about the productivity of linguistic patterns; Usage-Based and Probabilistic Phonology. Within the frameworks of both models, it would be perfectly plausible for the phonaesthemic patterns of a language to emerge and become productive. The chapter then critically examines the only study to have thus far investigated speakers' perceptions of phonaesthemes – Abelin's (1999) investigation in Swedish. Because there is no established methodology for the type of investigation conducted by the present thesis, Abelin's study is examined in order to assist with methodological cues. The most robust of Abelin's experiments are adapted and used to construct a pilot study (cf. Chapter 5). The chapter concludes by discussing my motivations and rationale for studying the presence and perception of the phonaesthemes in French and Polish alongside English.

# CHAPTER 3: MOTIVATING THE METHODOLOGIES

## 3.1) Theoretical Support for the Emergence and Perception of Phonaesthemes

As noted in Chapter 1, the main focus of this study is the extent to which the twelve phonaesthemes listed on pages 8-9 are (or are not) perceived by native and non-native English speakers. This is tested by measuring the extent to which speakers use the phonaesthemes productively; the idea being that productive use indicates perception of a phonaestheme (cf. 'Scope of Thesis', Chapter 1). However, the idea that the phonaesthemes may or may not be perceived presupposes that phonaesthemes are in fact perceptible patterns in language. Yet with there being so few studies which have thus far investigated the phenomenon, this is not necessarily an assumption that can automatically be made. To this end, the present chapter considers whether (and if so, how) it would be possible for speakers to acquire awareness of phonaesthemic patterns in their language, and thus how phonaesthemes could emerge as productive patterns.

It will be remembered from § 2.3 that phonaesthemes involve phonotactic sequences (prosodies) acquiring connotations within a speech community, by virtue of recurring in a range of words that express broadly similar meanings. This very definition suggests that phonaesthemes can emerge as perceptible patterns within a language without any deliberate analysis of its vocabulary; that is, through some sort of process that is either automatic or not fully conscious. Notably, there are at least two theories – Usage-Based Phonology (Bybee, 1985; 2001) and Probabilistic Phonology (Pierrehumbert, 2003) –which would allow speakers to acquire awareness of phonaesthemes in this manner. In essence, both theories posit that humans are predisposed towards the recognition of patterns in their language system. Usage-Based Phonology suggests that this ability is a direct consequence of the way in which the lexicon is structured, while Probabilistic Phonology argues that the recognition of linguistic patterns is a mechanism cultivated from infancy to assist in processing language accurately and efficiently. Each of these theories are now discussed in turn, and considered in relation to phonaesthemes.

### 3.1.1) Usage-Based Phonology and the Perception of Phonaesthemes

Usage-Based Phonology was a theory first proposed by Bybee (1985; 2001) in response to what she saw as the shortcomings of a strongly generativist view of language (cf. Chomsky and Halle, 1968; Kenstowicz and Kisseberth 1979; Halle and Mohanan, 1985; Pinker, 1999). The central view of generative approaches is that 'language is a formal system that uses a finite number of rules to generate [an infinite number of] sentences in order to communicate' (Shanker, 2002:127). The infinite number of potential morphosyntactic combinations – particularly in complex languages like English – suggests that humans' mental lexicons are unlikely to be sufficiently large to store every individual morpheme or lexical item as a discrete entry with its own unique set of phonological, semantic and syntactic rules (Kenstowicz, 1994:59-60). Therefore, strongly generativist views posit that most morphemes are stored as abstract, underlying forms in the lexicon, and that speakers produce ('generate') all the possible lexical items and combinations in their language by applying a series of regular phonological and syntactic rules to these underlying forms (Kenstowicz, 1994:59-60; Pinker, 1999:136; Burdett, 2002:59). Only those words that cannot be realised by the application of these regular rules are stored as full entries with their own set of phonological, morphological and syntactic properties (*ibid*). For example, the plural forms of most English nouns are realised by the addition of the /s/ suffix, or an allophone of this phoneme, depending on the rime of the word-final syllable (Lardiere, 2006:64). The fact that this phonological rule is so widely applicable within English means that there is no need to store the plural forms of most nouns; such storage would be redundant and a waste of the limited space in the lexicon. Therefore, only nouns where the underlying morpheme <<plural>> cannot be generated through the default rule of 'apply /s/ suffix or allophone depending on phonetic context' are stored in their entirety (such as 'children', 'mice' and 'larvae').

While many scholars – most notably Chomsky – have defended a generative view of human language, Bybee (2001:20-28) argued that a strongly generativist perspective was problematic. She argued that it was not clear that any part of language was stored in the form of rules; and that regular linguistic patterns need not be stored separately or differently in the lexicon from irregular patterns. Instead, she suggested that the lexicon could be organised as networked patterns of regularities and irregularities, learned by way of exposure to language in use. Bybee's proposed model of this process – the Usage-Based model – can be

summarised as follows. When a speaker experiences a linguistic item (phonetic sequence, morpheme, word etc.) for the first time, a series of representations of the item's phonological, semantic and syntactic properties is created and stored. Each time the item is subsequently encountered, these representations are 'activated' (ibid:20); that is, processed with some degree of consciousness (Schmidt, 2005:293). Through repeated exposure to the words in their language (and thus repeated activation of these words' phonological, semantic and syntactic representations), speakers are able to identify and learn the regularities and irregularities in the phonology, semantics and syntax of their language system. For example, 'a child [learns] that *–ed* marks the regular past tense of English [by] learning a number of verbs containing this suffix, such as *played, spilled, talked,* and so on' (Bybee, 2001:20). Humans are viewed as being predisposed towards recognising such regularities and irregularities, because in Bybee's model, these patterns constitute the very basis on which the lexicon is organised. The lexicon itself is seen as a vast network of inter-related patterns, where linguistic items are stored multi-dimensionally. In this space, the phonological, semantic and syntactic representations of words simultaneously belong to various groups of phonological, semantic and syntactic patterns. Bybee referred to these groups of patterns as 'schemas' (ibid:21). Every regularity and irregularity that a speaker identifies in the language constitutes one schema in the lexicon. The greater the frequency with which a schema is activated (i.e. the more frequently a speaker encounters a particular phonological, semantic or syntactic pattern), the more it is reinforced as a pattern of organisation, and the easier to access it becomes (ibid:28).

Figures 4 and 5, overleaf, provide illustrations of two different schemas. Figure 4 shows a selection of English words associated by a phonological pattern; the onset sequence [s] + [l]. Figure 5 shows a selection of English words associated by a semantic and phonological pattern; the meaning <<plural>> and the feature [z] in rime position. (Neither diagram shows the entirety of each schema – for example, there are many more English words that are also related to the semantics of <<plural>> – but the diagrams are purely for illustrative purposes):

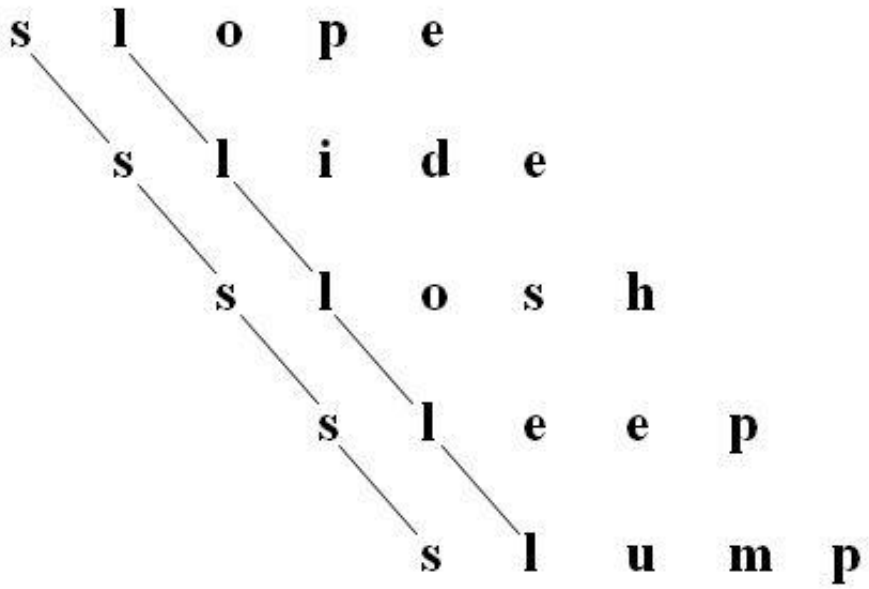



Figure 4: Phonological schema of lexical items with onset sequence [s] + [l] (adapted from Bybee, 2001:22).

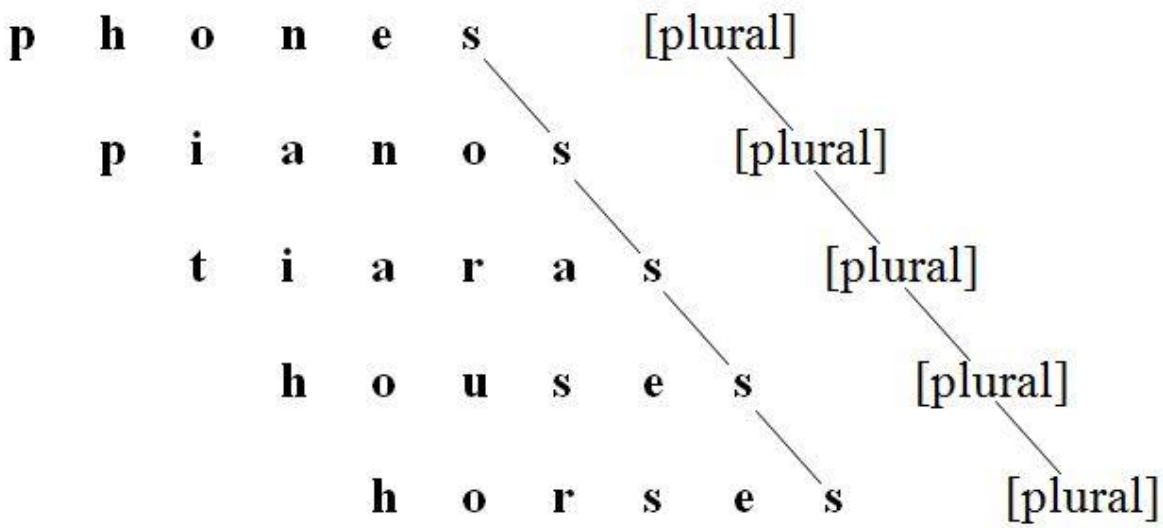



Figure 5: Semantic schema of English plural words with final [z] (adapted from Bybee, 2001:23).

The connecting lines in the diagrams show 'identity relations' (*ibid*), which are the specific phonological, semantic or syntactic components common to the linguistic items in a schema.

For example, in Figure 4, the bundle of phonological features[s] + [l] is shared by each of the words depicted (and many more besides); and thus forms one pattern of phonological organisation in the lexicon. Similarly, in Figure 5, the semantic feature <<plural>> and phonological feature [z] (one of the regular allophones realising plural forms in English) are shared by each of the words depicted. These form one semantic and one phonological pattern of organisation in the lexicon. (Naturally, there are a number of irregular English words which would be members of the <<plural>> semantic schema but not of the [z]-final phonological schema, such as 'children' and 'mice'. As noted above, neither diagram is intended to represent the entirety of each schema.)

In Bybee's view of the lexicon, the activation of any one of the items in a schema 'spreads to other items' to which it is connected (*ibid*). For example, if a speaker experienced the word 'slide' (as in Figure 4, above), the phonological representations of 'sleep', 'slump', 'slosh' and 'slope' (and any other word with this bundle of features) could also become activated, assuming these lexical items had been experienced previously by the speaker. Similarly, if a speaker experienced the word 'pianos' (as in Figure 5, above), the representations of other plural words could also become activated; as could any word that was semantically-related to the idea of <<piano>> or <<music>>. Of course, every word in a speaker's lexicon belongs to its own set of various phonological, semantic and syntactic schemas. So while 'slide' might activate 'slump' through one phonological schema (cf. Figure 4), it may also belong to another phonological schema which did not include 'slump'; such as words with medial [aɪ] (e.g. 'slime' and 'slight' etc.). It might also activate certain semantic schemas which did not include any of the other words in the [s] + [l] schema; perhaps <<falling>> or <<injury>>. And while 'pianos' might activate 'phones' and other English plural forms (cf. Figure 5), it might also simultaneously activate semantically-related words such as 'violin', 'orchestra' and 'music'; as well as certain phonological schemas like words with initial [p] (e.g. profit, person, panda, police etc.).

Bybee's Usage-Based model provides one theory of how phonaesthemes could emerge as perceptible (and in turn, productive) patterns. For if the lexicon is indeed organised as Bybee suggests, it provides an organisational system through which speakers can identify patterns in the phonology and semantics of their language. In essence, a phonaestheme is simply the repeated simultaneous activation of a particular phonological and a particular semantic pattern (schema). Speakers are attuned to recognising phonological and semantic

schemas, because such patterns form the very basis on which the lexicon is organised. Indeed, the applicability of the Usage-Based model to the perception of phonaesthemes is a point not lost on Bergen (2004:305). Using an adaptation of Bybee's diagrams, Bergen illustrates the simultaneous activation of phonological and semantic schemas in a sample of words exhibiting the /gl/-<<light>> phonaestheme:

IMAGE REMOVED FOR PURPOSE OF COPYRIGHT

Figure 6: The simultaneous activation of schemas in the /gl/-<<light>> phonaestheme (Bergen, 2004:305).

### 3.1.2) Probabilistic Phonology and the Perception of Phonaesthemes

Probabilistic Phonology (Pierrehumbert, 2001) is another theory which could account for the emergence of phonaesthemes as perceptible patterns within a language. Like other Probabilistic models of language (see, for example, Seager, 1983; Ellis, 2002; Jurafsky, 2003), the underlying idea of this theory is that of **stochastics** (Zuraw, 2000: xiv; Pierrehumbert, 2001; Blevins, 2007:153-154). Stochastics is the notion that, from an early age, speakers are able to assess the numerical distributions of features and patterns in the language they experience; calculating the features and patterns which are frequently and infrequently encountered, and in what contexts. These assessments of numerical distribution are used to form assumptions of probability about linguistic forms and patterns, which help speakers to process linguistic encounters accurately and efficiently, and are also thought to assist the rate at which young speakers acquire language (Auer and Luce, 2005:610; Fazly, Alishahi and Stevenson, 2010:1026).

Probabilistic Phonology, as the name suggests, applies the principle of stochastics to a phonological context. Its central idea is that speakers make assumptions of probability about phonological forms and patterns as a result of the frequency with which they encounter these forms; and in what contexts of experience (phonological, lexical, situational and so on). Speakers are thought to use these assumptions of probability for a variety of purposes. Chief

among these is to assist in interpreting meaningful sounds (and thus lexical items) from the continuous sound waves of an interlocutor's speech signal (Carr, 2008:166; Currie Hall, 2011: Online). In addition, it has also been suggested that speakers use these assumptions of probability in judging whether unfamiliar phonological sequences are likely to belong to their own language system (Pierrehumbert, 2003:188-189).

To demonstrate how probabilistic assumptions about phonological forms can assist speakers in interpreting the speech signal, consider the following example. Imagine that two British English speakers were discussing a mutual friend. One of the interactants recalled that the mutual friend used to own a large motorcycle; to which his interlocutor replied:

'Oh yeah. I remember that. It was fast'.

Now imagine that the first interactant was not able to distinguish whether the phones constituting the onset of the final word realised a /f/ or its voiced counterpart /v/ (and thus whether the final word in the utterance was 'fast' or 'vast'). The interactant could use a number of probability judgments about the phonological forms /f/ and /v/, and about the sequences [fæst] and [væst][5], to try to infer which consonant the ambiguous phones were most likely to have realised. For instance, they might calculate the likelihood that they realised /f/ or /v/ based on the frequency with which they experienced the sequences [fæst] and [væst] in the context of discussing motorcycles. This might lead them to infer that [fæst] was a more likely choice, because in their linguistic experience, this phonetic sequence collocated more frequently than [væst] with a semantics of <<motorcycles>> or <<transport>>. Likewise, the interactant could also calculate the probability of the phone having realised /f/ or /v/ based on the number of times they had previously experienced the speaker utter the sequences [fæst] and [væst]. For instance, if they had heard the speaker utter [væst] more frequently than [fæst] during previous interactions, the hearer might infer that the ambiguous phone was more likely to have realised /v/. And the frequency with which the hearer encountered both phonological sequences in day-to-day language could also be used to form a probabilistic inference. If the hearer tended to encounter [fæst] more frequently than [væst] in their day-to-day experience of language, they may infer a greater probability that the ambiguous phone realised /f/ (cf. Carr, 2008:166).

[5] Square brackets are used to indicate a sequence of phones rather than phonemes.

The idea that speakers form assumptions of probability about phonological forms from the language they experience, and apply these assumptions in subsequent linguistic encounters, means that Probabilistic Phonology offers another possible view of how phonaesthemes could emerge as perceptible patterns. Every time a speaker encounters a phonaestheme in language, they are by definition encountering a particular phonological form recurring in a particular experiential context (i.e. with the expression of a given meaning). So each exposure to a phonaestheme increases the frequency with which the speaker encounters the phonological form and meaning together; and thus adds to the degree of probability calculated that the two will co-occur in subsequent linguistic experiences. For example, each time a speaker encounters a /gl/-<<light>> word in language, it adds to the likelihood that /gl/ will be found in a context of expressing <<light>> in subsequent linguistic experiences. And each time a speaker encounters a /stɹ/-<<straightness or stretching>> word in the language, it adds to the likelihood that /stɹ/ will be found in a context of expressing <<straightness and stretching>> in subsequent linguistic experiences. This echoes Bybee's (2001:28) claim that the greater the frequency with which a schema is activated (i.e. the more frequently a speaker encounters a particular phonological, semantic or syntactic pattern), the more it is reinforced as a pattern of organisation in the lexicon. It also suggests that the phonaesthemes most likely to emerge as perceptible patterns could be those most frequently encountered, or those with the greatest number of words exhibiting the association in the vocabulary of the language (cf. discussion of 'type' and 'token' frequencies; Chapters 4, 5 and 6).

In sum, this section has shown that there are at least two theories which could account for the emergence of phonaesthemes as perceptible patterns within a language: Usage-Based Phonology and Probabilistic Phonology. As has been discussed, both models suggest that humans are predisposed towards the recognition of patterns in their language system. Usage-Based Phonology suggests that this ability is a direct consequence of the way in which the lexicon is structured, while Probabilistic Phonology argues that the recognition of linguistic patterns is a mechanism cultivated from infancy to assist in processing language accurately and efficiently. Given that phonaesthemes could emerge in a similar way to the patterns discussed by these models, it seems appropriate to use methodologies which anticipate that phonaesthemes can be perceived by speakers. The specifics of these methodologies are detailed in Chapter 5 (cf. Overview of Thesis, § 1.5). The next section in the present chapter critically discusses the only study to have thus far investigated speakers' perceptions of

phonaesthemes; Abelin's (1999) investigation in Swedish. Because there is no established procedure for the type of investigation conducted by the present thesis, Abelin's study is reviewed in order to assist with methodological cues. The most academically robust of Abelin's experiments are used to shape the methodologies taken by the experiments in the pilot and main studies presented in Chapter 5.

### 3.2) Studying Phonaesthemes Productively: Abelin's (1999) Swedish Study

Abelin's (1999) study addressed two research questions. The first was to determine whether there was evidence of any [principally onset] phonaesthemes in the present-day vocabulary of Swedish. The second was to investigate whether Swedish and non-Swedish speakers appeared to be aware of any of these correlations. To address these research questions, Abelin used a two-stage methodology. First, in the manner of many of the studies reviewed in Chapter 2, she examined the existing vocabulary of Swedish, recording any prosodies that recurred twice or more with the expression of a particular meaning (excluding cases of inflection and syntactic derivation; cf. p.102). Abelin's methodology suggests that this involved searching for common keywords in the definitions of lexical items sharing the same onset prosodies (Abelin, 1999:71-72). As is discussed in § 4.2.1, I considered this particular strategy too restrictive in the present study. This is because it might lead a researcher to overlook one or more exemplars of a phonaestheme; possibly very many. For instance, many of the prototypically <<pejorative>> words listed by Firth in support of the /sl/ phonaestheme do not feature the word "pejorative" in their definitions; or even the word "unpleasant" (cf. OED: Online). Cases in point include "sly", "slurry" and "slog". Clearly, then, one cannot totally determine whether or not a linguistic item expresses a particular meaning simply by searching for particular words in its definition. In Abelin's study, this strategy was designed for economy of time; Abelin's thesis involved sampling a much larger proportion of vocabulary than the present study. However, the downside to this strategy is that Abelin may have overlooked a number of exemplifying words for one or more phonaesthemes; or even some phonaesthemic patterns entirely. To minimise these possible oversights in the present study, I designed an alternative method of studying phonaesthemes in the existing vocabularies of languages. This is discussed in more detail in § 4.2.1.

Having determined a list of possible phonaesthemes in the vocabulary of Swedish, Abelin then conducted a number of experiments to test whether Swedish and non-Swedish

speakers productively associated the prosodies and meanings of these potential phonaesthemes; the idea being that productive association would indicate perception of a phonaestheme (cf. §§ 1.4; 3.1, above). This basic two-stage data collection process is also used in the present study (cf. Chapter 4).

In reviewing the methodologies of Abelin's productive experiments, it is worth first summarising the phonaesthemes she posited in the vocabulary of Swedish. This will simplify the discussion of each experiment; and in turn the critique of their strengths and weaknesses. The onset phonaesthemes proposed by Abelin in the vocabulary of Swedish are summarised in Table 8, overleaf. (In her study, Abelin also carried out several experiments to test other sound-symbolic correlations. For example, in one experiment, she compared the phonetic sequences used by speakers of different languages to express interjections like pain, happiness and interest. However, as it is only relevant to the present study to review the experiments used by Abelin to test phonaesthemes, it is only these experiments that are reviewed in this thesis.)

Table 8: Onset phonaesthemes identified in the vocabulary of Swedish (Abelin, 1999)

| **Onset Prosody** | **Meaning(s) with which this prosody recurs in the vocabulary of Swedish** *(exemplifying words provided by Abelin, with translations)* |
|---|---|
| /bl/ | light *(e.g. "blänka" – glitter)* |
| /gl/ | light *(e.g. "glänsa" – shine)* |
| /fj/ | pejoration *(e.g. "fjollig" – foolish)* |
| /sn/ | pejoration *(e.g. "snaskig" – smutty)* |
| /pj/ | pejoration *(e.g. "pjåkig" – mawkish)* |
| /sl/ | pejoration *(e.g. "slampa" – slut)* |
| /fn/ | pejoration *(e.g. "fnurra" – grumpy)*<br>sound/noise *(e.g. "fnysa" – snort)*<br>dryness *(e.g. "fnöske" – tinder)* |
| /gn/ | pejoration *(e.g. "gnidare" – miser)*<br>speech/singing *(e.g. "gnabb" – bickering)* |
| /skr/ | speech/singing *(e.g. "skrål" – bawling)*<br>destruction *(e.g. "skrot" – scrap)*<br>roughness *(e.g. "skrovlig" – rough)* |
| /fr/ | roughness *(no examples provided)* |
| /kr/ | roughness *(e.g. "krås" – ruffle)*<br>crookedness *(e.g. "kringelikrokar" – many bends in different directions)* |
| /br/ | sound/noise *(e.g. "braka" – crash)* |
| /kl/ | sound/noise *(e.g. "klappra" – clatter)* |
| /kv/ | sound/noise *(e.g. "kvida" – whimper)* |
| /kn/ | sound/noise *(e.g. "knaka" – creak)*<br>roundness *(e.g. "knorr" – curl)* |
| /sp/ | length/thinness *(e.g. "spenslig" – slender)* |
| /str/ | length/thinness *(e.g. "strimla" – strip)* |
| /spr/ | length/thinness *(e.g. "spröjs" – window-bar)*<br>separation *(e.g. "sprida" – spread)* |
| /tr/ | length/thinness *(e.g. "tråd" – thread)*<br>bad mood *(e.g. "tradig" – boredom/boring)* |
| /vr/ | bad mood *(e.g. "vrövel" – silly talk)* |
| /gr/ | hollowness *(e.g. "grav" – tomb)* |
| /sk/ | hardness *(e.g. "skal" – shell)* |
| /skv/ | wetness *(e.g. "skval" – gush)* |
| /mj/ | fine-grained substances *(e.g. "mjäla" – silt)* |
| /fl/ | movement *(e.g. "fladdra" – flutter)* |

Having identified these possible phonaesthemes in the vocabulary of Swedish, Abelin then devised a number of experiments to test whether the relationships were productive

among Swedish and non-Swedish speakers. Table 9, overleaf, summarises the aims, method and respondents sampled in each of these experiments. The methodologies and findings of each of these experiments are then critically reviewed, in order to identify methodological strengths which may be used or adapted in the present study; as well as methodological weaknesses which must be avoided or rectified. A general methodological criticism is also identified throughout Abelin's study vis-à-vis the presentation of coined words. This is discussed in its own subsection following the reviews of each experiment.

Table 9: A Summary of Abelin's (1999) Productive Experiments

| Experiment | Aim | Respondents | Method |
|---|---|---|---|
| Thesaurus Experiment | To investigate whether English and Swedish speakers associated certain concepts with similar word onsets | - 3 English speakers;<br>- 3 Swedish speakers | 1. Abelin chose three ideas at random (ideas selected were <<roughness>>, <<smoothness>> & <<stupidity>>).<br>2. Then chose a selection of existing English and Swedish words that were synonymous with these meanings<br>3. Asked English speakers which of the English synonyms sounded like they most accurately expressed each concept.<br>4. Asked Swedish speakers which of the Swedish synonyms sounded like they most accurately expressed each concept |
| Cross-Language Interpretation of Swedish Phonaesthemes Test | To investigate whether a selection of phonaesthemes identified in the vocabulary of Swedish were productive among non-Swedish speakers | - 1 English speaker;<br>- 1 Arabic speaker;<br>- 1 Dutch speaker;<br>- 1 German speaker;<br>- 1 Ibo speaker;<br>- 1 Spanish speaker<br><br>Table Continues | 1. Abelin read aloud one word featuring each of the onsets /fj/, /sl/, /pj/, /fl/, /kl/, /skr/, /skv/, /str/, /tr/, /gl/, /bl/ and /gr/. These were all onsets identified as recurring with particular meanings in the vocabulary of Swedish (cf. table 8, above). The /pj/ and /bl/ words were coined; the remainder were all existing lexical items in Swedish<br>2. Then asked respondents to write down the meaning they felt each word would most likely express<br>3. Analysed responses to investigate whether non-Swedish speakers associated any of the words (thus onsets) with the recurrent meanings identified in the vocabulary of Swedish (cf. table 8).<br><br>Overleaf |

| Experiment | Aim | Respondents | Method |
|---|---|---|---|
| Test 1a<br>(open-ended) | To investigate whether a selection of phonaesthemes identified in the vocabulary of Swedish were productive among Swedish speakers | - 14 Swedish speakers | 1. Abelin selected six meanings. These included <<pejoration>>, <<crookedness>>, <<bad mood>>, <<dryness>>, <<wetness>> and <<roughness>>. Each meaning was found to recur with one or more onsets in the vocabulary of Swedish (cf. table 8).<br>2. Then provided respondents with a short (written) definition related to each of the meanings<br>3. Asked respondents to coin (write down) any word of their choosing to name or describe each definition |
| Test 2a<br>(open-ended) | To investigate whether a selection of phonaesthemes identified in the vocabulary of Swedish were productive among Swedish speakers | - 14 Swedish speakers<br><br>Table Continues | 1. Abelin presented respondents with six disyllabic coined words, all in written form. Each word featured an onset associated with one of the six meanings tested in 1a, above (cf. table 8). Word onsets used included /pj/ (<<pejoration>>); /kr/ (<<crookedness>>); /vr/ (<<bad mood>>); /fn/ (<<dryness>>); /skv/ (<<wetness>>) and /skr/ (<<roughness>>). First syllable bore the onset under test. The same final syllable was used in all six words; the 'semantically-neutral' /ɪg/ (Abelin, 1999:220). Words coined were "skvatig", "fnotig", "vratig", "krötig", "pjotig" and "skratig".<br>2. Abelin then asked respondents to write down the meaning they felt would be best expressed by each of the coined words.<br><br>Overleaf |

| Experiment | Aim | Respondents | Method |
|---|---|---|---|
| Test 1b (constrained choice) | To investigate whether a selection of phonaesthemes identified in the vocabulary of Swedish were productive among Swedish speakers | - 15 Swedish speakers | 1. In each question, Abelin provided respondents with a single definition related to one of the meanings shown in Table 8 (above). All meanings from the table appear to have been tested, but Abelin did not discuss the findings for the meanings <<light>>, <<fine-grained substances>> or <<sound/noise>>.<br>2. In each question, Abelin also provided respondents with three coined words (in written form). Words were coined following the same criteria as in Test 2a. One of the three words in each question contained the onset that recurred with the meaning (definition) provided.<br>3. In each question, Abelin asked respondents to choose which of the three words they thought was most appropriate to express the definition provided. |
| Test 2b (constrained choice) | To investigate whether a selection of phonaesthemes identified in the vocabulary of Swedish were productive among Swedish speakers | - 15 Swedish speakers | 1. In each question, Abelin provided respondents with a single word bearing one of the onset phonaesthemes shown in Table 8 (above). Words were provided in written form, and were coined following the same criteria as in Test 2a. Abelin appears to have tested each of the onsets shown in Table 8, although she only discussed the results for 14 of the onsets.<br>2. In each question, Abelin also provided respondents with a series of definitions. One of these definitions related to the meaning associated with the onset of the coined word.<br>3. In each question, Abelin asked respondents to choose which of the definitions they thought was most appropriate for the coined word provided. |

### 3.2.1) The Thesaurus and Cross-Language Experiment

Of all Abelin's experiments, the Thesaurus and Cross-Language Phonaesthemes tests have the most substantial methodological weaknesses. The main shortcoming of both experiments concerns the number of respondents sampled. In the Thesaurus experiment, Abelin surveyed only three English and three Swedish speakers; and in the Cross-Language Phonaesthemes test, Abelin sampled only a single speaker from each of the six languages under study (cf. Table 9, above). The sample sizes in these experiments make it impossible to draw any conclusions about the productivity and perception of phonaesthemes from the results, since it is inappropriate to assume that such small numbers of respondents will offer representative cross-sections of speakers of their respective languages. Indeed, in the Cross-Language Phonaesthemes test, the response of a single speaker could have skewed the data for an entire language cohort by a hundred percent. This could have totally determined whether a phonaestheme was deemed to be productive or non-productive in a particular language. Similarly, in the Thesaurus experiment, the response of a single speaker could have affected the inferred productivity of a phonaestheme in English or Swedish by a third. It is here that the first methodological cue is taken for the present study: the number of respondents sampled must be large enough to offer a degree of representativity and generalisability. As is discussed in Chapter 5, the present study samples thirty speakers of English, French and Polish; double the number of the largest respondent cohort sampled by Abelin in any experiment. This ensures greater generalizability of findings over Abelin's study, and increases the confidence with which conclusions can be drawn from the data.

It is not just in terms of the sample size that the Thesaurus and Cross-Language Phonaestheme experiments are problematic. Both tests encounter further methodological issues caused by the use of pre-existing words. In the Thesaurus experiment, Abelin asked English speakers to select words from lists of pre-existing English synonyms; and asked Swedish speakers to select words from lists of pre-existing Swedish synonyms (cf. Table 9, above). However, Abelin did not appear to have considered the potential problems that the use of these pre-existing words may have caused. Since both cohorts of speakers would have been familiar with the meanings of most if not all of the words supplied, there would have been nothing preventing English and Swedish speakers from simply selecting the word whose [existing] meaning they felt was most closely synonymous with the definition provided; rather than basing their judgment on the sounds of the words in any way. As was discussed in

§ 2.4 (under ‘phonaesthetics’), it is extremely difficult if not impossible to separate sound from meaning in this way; any connotations ascribed to the sound structure of a familiar (known) word will almost certainly be affected by speakers’ knowledge of what the word means. A more meaningful approach could have been to provide the respondents with coined words, which would have been free from pre-existing semantic associations. The informants would then have had to formulate their responses based on the sound structures of the words alone. A similar situation is seen in the Cross-Language Phonaesthemes test. Abelin did not clarify why she coined words to test the /pj/ and /bl/ phonaesthemes but used existing words in Swedish to test the remaining associations. Here again, it would arguably have been more appropriate to once again provide respondents with exclusively coined words. This would eliminate the possible counter-argument that any or all of the respondents might have been familiar with the meanings of one or more of the existing Swedish words. As a result of these methodological shortcomings, it would be practical to ensure that all words given to respondents in the present study are coined; except in an experiment which specifically sets out to test the meaning effects of existing words in a language (cf. Pilot Study test Type 5, § 5.1.5). Using coined words would reduce any potential meaning effects caused by speakers’ knowledge of existing words; and would ensure, as far as possible, that respondents ascribed connotations to linguistic forms as a result of their sound structures alone.

### 3.2.2) Tests ‘1a’, ‘1b’, ‘2a’ and ‘2b’

#### 3.2.2.1) Sample Sizes

In the remaining experiments in her thesis (tests ‘1a’, ‘1b’, ‘2a’ and ‘2b’) Abelin surveyed a larger number of respondents than in the earlier tests. As such, these experiments are not open to the same degree of criticism as the ‘Thesaurus’ and ‘Cross-Language’ experiments with regards sample sizes. The number of respondents sampled in these experiments was still smaller than the cohort sizes used in the present study (as noted above), and were still too small to have generated widely generalizable or fully representative findings, or findings suitable for statistical testing. Nevertheless, these later tests represented a substantial improvement over the first two experiments in this respect. However, there was still an issue with sample size in these later experiments, in that Abelin surveyed fourteen respondents in tests ‘1a’ and ‘1b’, but fifteen respondents (the same cohort plus a further informant) in tests ‘2a’ and ‘2b’. While a larger number of respondents increases the

representativity and generalizability of findings, sampling unequal numbers of respondents makes it more difficult to compare the productivity of phonaesthemes across these experiments. Abelin circumvented this difficulty by discussing her results in terms of majority or minority trends. However, this sort of general description is not sufficiently precise for the present study, as noted in § 1.2. Not only are the aims of the present thesis more quantitatively grounded than those of Abelin's study; but I also perform a number of statistical tests on sections of my data (as discussed in Chapter 5). Like all statistical tests, these calculations ideally rely on the samples (i.e. numbers of respondents) being of equal size in order to produce accurate comparisons. For these reasons, the present study surveys the same thirty speakers of each language in every experiment. This ensures that the productivity of phonaesthemes can be accurately compared across different experiments, as well as between different language cohorts in the same experiment.

#### 3.2.2.2) Tests '1a' and '2a'

The methodologies of tests '1a' and '2a' exhibit both strengths and weaknesses. To discuss these, it is necessary to briefly review the main findings of both experiments, beginning with those of test 1a. To this end, table 10, below, summarises all the onset prosodies that recurred (i.e. that were used more than once) in the words coined by speakers for the definitions in test 1a:

Table 10: Summary of recurring onset prosodies in Abelin's (1999) Test '1a'

| **Meaning** | **Onset prosody(s) recurring with this meaning in the vocabulary of Swedish** | **Onset prosody(s) recurrently used in words for this meaning in Test 1a (ratio of respondents to use this onset)** |
|---|---|---|
| <<pejoration>> | /pj/, /fn/, /fj/ | /fl/ (3/14), /fj/ (2/14) |
| <<crookedness>> | /kr/ | /kr/ (2/14), /vr/ (2/14), /sl/ (2/14), /pl/ (2/14), /sjl/ (2/14) |
| <<bad mood>> | /vr/, /tr/, /gr/ | /vr/ (2/14) |
| <<dryness>> | /fn/ | /fn/ (4/14), /kr/ (3/14) |
| <<wetness>> | /sl/, /sv/, /pl/, /bl/ | /sv/ (2/14), /spl/ (2/14) |
| <<roughness>> | /skr/, /fr/, /kr/ | /fl/ (2/14) |

Abelin suggested that these results appeared to show the productive use of a number of the phonaesthemes identified in the vocabulary of Swedish (cf. Table 8). This was because:

- Four of the fourteen native Swedish speakers coined words with onset /fn/ in response to the <<dryness>> definition;
- Two speakers coined words with onset /fj/ in response to the <<pejorative>> definition;
- Two speakers coined words with onset /kr/ in response to the <<crookedness>> definition;
- Two speakers coined words with onset /vr/ in response to the <<bad mood>> definition, and
- Two speakers coined words with onset /sv/ in response to the <<wetness>> definition.

Taken at-large, these results suggest that this style of experiment could be worth adapting in the present study. It appears that the experiment provided speakers with a context which facilitated their recognition and productive use of the phonaesthemes *to some extent*. However, it is worth noting that the findings might not be quite as revealing as Abelin implies. In a sample size this small, it is by no means clear whether these results indicate evidence that the patterns were perceived. The fact that four speakers (29%) coined words with onset /fn/ seems more convincing, but the fact that only two speakers (14%) coined words with onsets /fj/, /kr/, /vr/ and /sv/ makes it impossible to suggest that these findings indicate perception of the phonaesthemes. In a sample size of fourteen speakers, it is possible that two respondents could have chosen the same onset by chance. (It is, of course, also possible that four speakers could have used onset /fn/ by chance in response to <<dryness>>; although the fact that nearly a third of the cohort chose this onset suggests that this was *somewhat* less likely.) As such, claiming that the results for /fj/, /kr/, /vr/ and /sv/ indicated some perception of these phonaesthemes could be risking a 'type 1' error; claiming that a result indicates evidence of a relationship where none exists (McKillup, 2012:130). In Abelin's study, as in the present investigation, the lack of previous evidence against which to compare such findings could heighten the risk of making a type 1 error unless steps were taken to avoid this. It is here that two key methodological cues can be taken for the present study: Primarily, the number of respondents sampled has to be as large as is possible and practical, in order to ensure that any findings are valid and did not simply occur by chance; and sufficiently diverse to constitute a representative cross-section of the given population(s). It is for this reason that each speaker cohort I sampled in the present study was double the size of Abelin's largest groups, at thirty informants apiece, and included a diversity of ages,

genders and socioeconomic backgrounds wherever possible. This point is discussed further in the next section of the present chapter, when I weigh the motivations for studying three languages against the numbers of respondents sampled. The second methodological cue that can be taken for the present study is that one cannot claim any given result 'proves' that a phonaestheme is perceived (or not). It is impossible to know or even suggest how many speakers would need to be sampled, or how many times the patterns would need to be used productively in order to 'prove' that the phonaesthemes were perceived or not; particularly in view of the lack of existing studies into the perception of these patterns. In an exploratory study such as this, the analysis of data will need to focus on numerical tendencies and trends, making appropriately tentative remarks as to what these findings appear to indicate. For instance, in Chapter 5, I compare the productivity of the phonaesthemes within experiments, across experiments, within each language cohort and across the three language cohorts, but the focus of these comparisons is on whether the patterns appeared to have been more or less productive in each case. Naturally, the greater the number of respondents who apply a phonaestheme productively, the more convincing the evidence is likely to be that the pattern has been perceived, providing the cohort is sufficiently large as to be representative. But comparing in this way, and describing the general trends in the data, is the limit of what can be attempted at this stage. I return to this point in Chapter 5 (§ 5.2.1), when I discuss how I will interpret the data from my experiments.

Before discussing the findings of 'test 2a', it is worth pointing out one further issue with the way in which Abelin analyses the findings of this particular experiment. Abelin does not indicate at any point whether she had considered the possibility that different onsets might have had different chances of being chosen by speakers when constructing their coined words. One cannot necessarily assume that every phonotactically-legal onset in Swedish would have stood an equal chance of appearing in speakers' coined words. For example, some onsets might have stood a higher likelihood of being chosen because they occur more frequently than others in Swedish; or because speakers tended to encounter these particular prosodies more frequently than any others in the particular linguistic contexts requested by the definitions (e.g. adjectival words, nominal words or verbal words). This calls into further question whether the results of this experiment provide reliable evidence about the productivity of these patterns. For instance, it might have been that four speakers used onset /fn/ in response to the <<dryness>> definition because this definition asked for a nominal word (such a word to name an object that is typically associated with being dry); and Swedish

speakers happen to encounter this onset more frequently than any other onset in nouns throughout day-to-day language use. If this was the case, then it would seem questionable whether the four speakers using /fn/ had, in fact, perceived the /fn/-<<dryness>> association in the vocabulary of Swedish; or whether this finding simply reflected the extent to which speakers encountered this onset in nouns. In combination with the problems arising from the small respondent cohort, the fact that Abelin does not indicate that she had considered this possibility means that it is by no means clear whether the findings of this test are meaningful at all. As a result of this, a further methodological cue can be taken for the present study at this point: In any questions which ask speakers to coin words in response to particular definitions or stimuli, some consideration will need to be given to whether each phonotactically-legal onset had an equal chance of being used (at least in English), or whether some onsets (and if so, which ones) might have stood a higher chance of being chosen than others; perhaps because they occur more frequently in the language at-large, or because speakers tend to encounter these particular prosodies most frequently in the particular linguistic contexts requested by the questions. The discussion returns to this point in Chapter 5.2.3 when I discuss my third test type, which is loosely based on this particular experiment.

Table 11, below, summarises every connotation that was attributed more than once to the onsets under test in Abelin's test 2a:

Table 11: Summary of the recurrent meanings in Abelin's (1999) Test '2a'

| **Onset Phonaestheme** | **Meaning(s) theoretically associated with this onset in stage one** | **Meaning(s) ascribed to this onset in Test 2a (ratio of respondents to ascribe this meaning)** |
|---|---|---|
| /fn/ | <<pejoration>><br><<dryness>> | <<pejoration>> (5/15)<br><<dryness>> (2/15) |
| /vr/ | <<bad mood>> | <<bad mood>> (3/15) |
| /kr/ | << crookedness>><br><<roughness>> | <<pejoration>> (undisclosed) |
| /pj/ | <<pejoration>> | <<pejoration>> (10/15) |
| /skr/ | <<destruction>><br><<roughness>><br><<speech/singing>> | <<destruction>> (7/15) |

Like test 1a, this experiment also seems to suggest that a number of the phonaesthemes may have been perceived to some extent, indicating that a similar style of experiment could work well in the present study. However, in this experiment, the numbers and proportions of speakers using the patterns productively appear markedly higher than in test 1a, suggesting that these results are potentially less likely to have arisen by chance and are more likely to reflect some degree of perception of the phonaesthemes. For instance, ten of the fifteen speakers ascribed a connotation of <<pejoration>> to the /pj/ word, reflecting the /pj/-<<pejoration>> recurrence in the vocabulary of Swedish. Similarly, seven speakers ascribed a connotation of <<destruction>> to the /skr/ word, and five speakers ascribed a connotation of <<pejoration>> to the /fn/ word. The only onset under test not associated with at least one of its recurring meanings in the vocabulary was /kr/. Abelin reported that this onset was productively assigned a connotation of <<pejoration>>, but did not disclose the number of speakers who made this association. Though not an issue with the design of the experiment, the fact that Abelin omitted this particular finding calls into question the reliability and credibility of her interpretations of the data. This omission highlights an important presentational cue for the present study: it reinforces the need to ensure that every finding is reported clearly, irrespective of whatever trend(s) it may show.

#### 3.2.2.3) Tests '1b and 2b'

Like tests '1a' and '2a', tests '1b' and '2b' exhibit both strengths and weaknesses. In terms of their strengths, both experiments once again appear to show some degree of productivity for several phonaesthemes. For instance, in test '1b', Abelin found that nineteen of the phonaesthemes under test were used productively by at least two thirds of the respondent cohort. Of particular note was onset /fj/, which was associated with <<pejoration>> in preference to any other onset by every single respondent sampled. In addition, a further five onsets were productively associated with their respective meanings by thirteen of the fourteen respondents. These included:

- /str/-<<length and thinness>>;
- /pj/-<<pejoration>>;
- /fn/-<<dryness>>;
- /vr/-<<bad mood>>;

- /sp/-<<wetness>>

A number of these trends were mirrored in the findings of test ‘2b’. In this test, four onsets were associated with their recurrent meanings in preference to any other meaning by every single respondent sampled. These included:

- Onset /fn/ and <<dryness>>;
- Onset /pj/ and <<pejoration>>;
- Onset /gr/ and <<hollowness>>;
- Onset /sn/ and <<speech/singing>>.

In addition, all but one respondent associated onset /sk/ with <<movement>>, and all but one respondent associated /spr/ with <<separation>>.

The findings of these experiments suggest that it will be appropriate to adapt this style of closed-ended test in the present study. Put simply, tests ‘1b’ and ‘2b’ generated the highest productivity of phonaesthemes seen anywhere in Abelin’s study: Test ‘1b’ generated 100% productivity of the /fj/-<<pejoration>> phonaestheme; while test 2b generated 100% productivity of the /fn/, /pj/, /gr/ and /sn/ phonaesthemes. These results suggest that constraining the choice of phonological forms and semantic domains available to respondents – thus limiting the number of possible phonological-semantic associations in each question – creates an effective context for the perception and productive use of phonaesthemes. However, there is a methodological weakness with the way in which Abelin administered these experiments which will need to be avoided in the present study. This concerns the way in which coined words were presented to respondents, and is a shortcoming which also applies to tests ‘1a’ and ‘2a’, as is discussed below.

#### 3.2.2.4) General Methodological Shortcoming of Tests ‘1a’, ‘1b’, ‘2a’ and ‘2b’: The Presentation of Coined Words

Although tests ‘1a’, ‘1b’, ‘2a’ and ‘2b’ appear to show a degree of productive use of a number of phonaesthemes, all four experiments encounter a further methodological issue regarding the way in which they provided or requested coined words. As shown in Table 9,

above, all four experiments relied on the use of coined words; and in all four experiments, these coined words were given or requested in graphological form:

- In each question of test ‘1a’, Abelin asked respondents to coin and then write down a word for a given definition. She then interpreted the phonological forms of the words coined by respondents from these written forms.
- In each question of test ‘2a’, Abelin asked respondents to assign a meaning to a given coined word based on its phonological form. Abelin presented the coined word in written form.
- In each question of test ‘1b’, Abelin asked respondents to assign one of three coined words to a given meaning based on their phonological forms. She presented all three words in written form.
- In each question of test ‘2b’, Abelin asked respondents to assign one of three meanings to a given coined word based on its phonological form. She presented the coined word in written form.

However, Abelin did not appear to have considered that providing and requesting coined words in graphological form may have been problematic. For instance, in test ‘1a’, Abelin did not discuss whether – and if so, how – she could be certain that she was correctly interpreting the [onset] phonological forms of speakers’ coined words from their graphological responses. It may be that there is a more regular grapheme-phoneme correspondence in Swedish than in English; or that each of the phonaesthemes Abelin was studying can only be realised by one graphological sequence in Swedish. But Abelin gave no indication that she had considered whether this was the case; or whether it could have been possible to misinterpret any of the phonaesthemes from the graphemes given by respondents. Of course, if such misinterpretation could have occurred, it would cast doubt on the validity of the results, since there would be no guarantee that Abelin would have interpreted the same [onset] phonological forms from the coined words as those intended by her respondents. A similar scenario is seen Tests ‘1b’, ‘2a’ and ‘2b’. These experiments are problematic for the opposite reason: in each of these tests, Abelin relied on respondents being able to interpret a particular phonological form from the graphological sequences she provided. However, Abelin did not appear to have considered the possibility that any or all of the respondents might have interpreted different phonological forms from the graphemes than those she had

intended. And if such misinterpretation could have occurred, it would again cast doubt on the validity of Abelin's results; since there would be no guarantee that every informant was responding to the same phonetic stimulus, or the stimulus intended by Abelin.

In view of these potential shortcomings, two further methodological cues can be taken for the present study: Firstly, in any experiments where respondents are asked to ascribe connotations to coined words, it will be most practical to present the coined words in spoken form. This will circumvent any potential issues of respondents misinterpreting the phonological forms of these coined words (particularly their onset consonant clusters); and will ensure that every single informant in all three language cohorts is responding to identical phonetic stimuli. Secondly, in experiments where respondents are asked to coin words of their own, where it is clearly not feasible to expect a response in audio form, consideration will need to be given to the possible graphological sequences that may be used to realise the twelve phonaesthemes in English, French and Polish. This is particularly important in English and French, since both languages feature 'very inconsistent [relationships]' between graphemes and phonemes (Meyer, 2005:133; Akamatsu, 2006:492). Establishing the graphemes that could be used to represent the phonaesthemes in each language will increase the confidence with which the use (or non-use) of the phonaesthemes can be interpreted in respondents' coined words.

### 3.2.3) Summary

Abelin's study constitutes an important contribution to both the field of enquiry and to the present thesis. In terms of the former, Abelin's work provides the first and only evidence thus far to suggest that speakers might be aware – at some level – of the phonaesthemic patterns in the vocabulary of their language. In terms of the latter, Abelin's work provides the only methodological precedent for the present study. In this respect, it offers both strengths which can be emulated and weaknesses which can be improved upon. Following the discussion above, the specific methodological cues taken from Abelin's work in the present study can be summarized as follows:

1. Tests '1a', '1b', '2a' and '2b' all appear to offer effective ways of testing the productivity (and thus perception) of phonaesthemes. This appears to be particularly true of the closed-ended tests ('1b' and '2b'). This suggests that it would be

appropriate to use adaptations of these question types in the present study, subject to points 2-9, below (cf. further discussion in Chapter 5).

2. The number of respondents sampled must be sufficiently large as to produce valid, representative findings, so as to minimize the chance of making a 'type 1' error.
3. Equal numbers of speakers from each language should be sampled, and the same cohort of speakers from each language should complete each experiment. This will ensure that the productivity of phonaesthemes can be accurately compared across experiments and between language cohorts.
4. In the analysis of data, no result can be claimed to 'prove' that a phonaestheme is perceived (or not). In an exploratory study such as this, the analysis of data will need to focus on numerical tendencies and trends, making appropriately tentative remarks as to what these findings appear to indicate.
5. Coined words should be used in every experiment(s) which asks respondents to ascribe connotations to words as a result of their phonological forms. This will circumvent the likely connotative effects of using pre-existing words with established denotations.
6. In any experiment(s) which asks respondents to coin words in response to a given definition or meaning, consideration will need to be given to whether each phonotactically-legal onset had an equal chance of being used (at least in English); or whether some onsets (and if so, which ones) might have stood a higher chance of being chosen than others. Only in light of this can one undertake an informed analysis of the data, and make observations as to whether a given result suggests perception of a phonaestheme or not.
7. It is important to remember to report all findings clearly, giving reasons for any interpretations made; irrespective of whatever trends the findings may show.
8. In any experiment(s) which asks respondents to ascribe connotations to coined words, it will be most practical to present the coined words in spoken form. This will circumvent any potential issues of respondents misinterpreting the phonological forms of these coined words; and will ensure that every informant in all three language cohorts is responding to identical phonetic stimuli.
9. In any experiment(s) which asks respondents to coin words in response to a given meaning or definition, consideration will need to be given to the possible graphological sequences that may be used to realise the twelve phonaesthemes in English, French and Polish. This will increase the confidence with which the use

(or non-use) of the phonaesthemes can be interpreted from the graphology of respondents' coined words.

The next and final subsection in the present chapter discusses my motivations and rationale for studying the presence and perception of the phonaesthemes in French and Polish alongside English. In so doing, it weighs the motivations for studying three languages against the need to sample a sufficient number of speakers in order to generate representative, comparable samples.

### 3.3) Motivating the Languages Sampled

In addition to exploring the presence and possible productivity of the phonaesthemes in English, I also chose to study the twelve patterns in French and Polish. My choice to study the patterns in other languages alongside English was motivated by a number of factors. The first factor was one of the gaps in the existing literature noted in Chapter 1: Although these patterns have been identified and discussed to a limited extent in the vocabulary of English, there is currently no evidence to suggest whether they may also extend to the vocabularies of other languages. Indeed, any or all of these phonaesthemes could be universal tendencies – patterns present in a large number of languages worldwide – but there is currently no evidence to suggest whether this may be the case. On the basis that phonaesthemes are recurring patterns in the phonetic forms used to express particular meanings, I would predict that these twelve associations would recur more frequently in English than in any other language, and that they would be perceived more frequently by English speakers than by non-English speakers. This is because different languages tend to express the same meanings using different phonetic forms (cf. discussion of arbitrariness, Chapter 2.5). As such, it seems likely that different phonaesthemic patterns would have emerged in different languages; and that there is no motivation to predict that these specific phonaesthemes would extend beyond English. This would seem to be consistent with the fact that the phonaesthemes identified by Abelin (1999) in the vocabulary Swedish (see above) are generally very different to those that have been identified in the vocabulary of English. However, this possibility currently remains unexplored. I therefore reasoned that it could be worth exploring, very tentatively, whether the phonaesthemes appeared to extend to the vocabularies of languages other than English, and whether they were patterns perceived by speakers of these languages.

The second factor motivating my choice to study the phonaesthemes in other languages was interpretability of the English data. As there is no motivation to predict that these patterns would extend to languages other than English, I reasoned that studying the phonaesthemes in other languages would provide a useful basis against which to compare and interpret the English data. This proved particularly important to interpreting the findings of the productive experiments. Consider the following scenario. Imagine that, in a given experiment, I had found the /sw/ phonaestheme was productively associated with <<fast or strong movement>> by twenty native English speakers in a sample of thirty. Without any other data against which to compare this finding, it would have been difficult to determine or substantiate the degree to which this indicated perception of the pattern by English speakers. However, with another point of reference to compare against this finding, it would be possible to reach a more informed interpretation of the data. For example, imagine that in the same experiment, five Polish speakers productively associated /sw/ with <<fast or strong movement>>. When considered alongside this finding, the English result is placed into context by comparison: The results would suggest that the phonaestheme was perceived to a somewhat greater extent by English speakers than by Polish speakers. In turn, this would seem to agree with the prediction that different phonaesthemic patterns would emerge in different languages. This type of comparison is a direct manifestation of the cue taken from Abelin's study that one cannot claim any given result 'proves' that a phonaestheme is perceived (or not); and that in an exploratory study such as this, the analysis of data needs to focus on numerical tendencies and trends, making appropriately tentative remarks as to whether the patterns appeared to have been more or less productive in each case (cf. §3.2.2.2; § 5.2.1).

While studying the phonaesthemes in other languages alongside English offered the key advantages discussed above, it is worth noting that this decision also introduced a number of limitations to my study; particularly with regards the number of informants sampled and the validity and representativity of my findings. By studying the presence and perception of these patterns in two additional languages, I was not able to sample as many respondents as if I had simply focused on studying the patterns in English. This was largely due to constraints of time. I first had to arrange for the productive experiments (discussed in Chapter 5) to be professionally translated into French and Polish; a process which took several weeks. I then had to write a separate online survey for each language cohort. Due to the complexity of the questions involved and the stimuli required in each case, this amounted

to several further weeks spent writing .html code. In combination, these processes left me with less time to recruit and sample respondents (without overrunning the deadlines of the project) than if I had restricted my enquiry to English alone. Of course, sampling fewer respondents in each language group reduced the extent to which I could claim my findings were representative of each language at-large, and the certainty with which I could claim my findings did not come about by chance (cf. discussion of Abelin's 'Test 1a', § 3.2.2.2). However, had I only studied the patterns in English, it would have been much more difficult to reach any meaningful interpretation of the data, since I would have had no baseline against which to compare the findings (see discussion above). Moreover, if I had restricted my enquiry to English alone, I could not have addressed the gap in the literature vis-à-vis whether there was any evidence of the patterns in languages other than English. For these reasons, I determined that the benefits of exploring the patterns in other languages outweighed the drawbacks of sampling smaller respondent groups. Nevertheless, given the shortcomings of Abelin's (1999) 'Test 1a' discussed above, I was mindful of the need to ensure my sample sizes were as large as possible in order to maximise the validity of my findings, and as diverse as possible in order to provide the most representative cross-sections of their various populations. For this reason, I sampled thirty native speakers of each language, on the basis that this number was feasible within the time constraints of the project, yet still represented a sufficient sample size for an exploratory study of this kind, and was twice the size of even the largest of Abelin's (1999) respondent groups. In addition, I tried to ensure where possible that each online survey was distributed to a diverse range of informants; both male and female, with a variety of different ages, education levels and socioeconomic backgrounds.

Having decided to explore the phonaesthemes in other languages alongside English, my decision to use French and Polish specifically was motivated by two factors. Firstly, the phonotactics of these languages allow most of the twelve prosodies in word-initial position; the only exceptions being that /skw/ is not used in onset position in French, and /tw/ is not used in onset position in Polish[6]. In addition to phonotactic reasons, these particular

[6]This does not, however, imply that the realisations of these prosodies are phonetically identical across the three languages. For example, the /gɹ/ phonaestheme would most likely be realised as [gɹ] in English, [gʁ] in French and [gr] in Polish. This is because since /ɹ/ tends to be realised as an alveolar approximant in English, a uvular fricative in most dialects of French, and an alveolar trill in Polish (Brooks, 1975:9; Fougeron and Smith, 1999:78-81; Ladefoged, 1999:41-44; Salter and Bousfield, 2002:746). However, none of the variations in the way the twelve prosodies are realised create contrasts of meaning in the languages studied. For example, the most likely "English" and "Polish" realisations of the /gɹ/ phonaestheme (i.e. [gɹ] and [gr])

languages presented the fewest methodological difficulties in terms of translating materials and recruiting respondents. One of my personal friends was employed as a professional French translator at the time of writing, and one of my academic colleagues was a native Polish speaker. This meant that when the time came to design and implement the productive study (cf. Chapter 5), I had regular access to individuals who were able to translate the experiment questions; as well as increased access to native speakers of both languages that I could recruit as respondents. This increased access to native speakers of French and Polish was particularly helpful. As well as enabling me to distribute the survey to informants of different ages, genders and backgrounds, it enabled me to sample informants (wherever possible) who identified themselves as non-fluent English speakers, and who had not studied English at secondary school level. In turn, this meant I could minimise any effect that knowledge of English vocabulary may have had on the responses of French and Polish speakers.

The next two chapters present the methodologies and findings of the experiments used to address the research questions posed in § 1.3. Chapter 4 discusses the vocabulary-based experiments used to explore the presence of the phonaesthemes in English, French and Polish, while Chapter 5 discusses the productive experiments used to investigate the possible perception of these patterns.

would both represent the same onset prosody to a French speaker; as would the realisations [ɡʁ] and [ɡʀ] (Fougeron and Smith, 1999:80). This meant that I could expose all three speaker groups to the same audio recordings of the prosodies (in the main study, these were produced by a native English speaker), knowing that the same sequences would be interpreted – and responded-to – by all three language cohorts.

# CHAPTER 4: VOCABULARY-BASED EXPERIMENTS

This chapter presents the methodologies and findings of the two vocabulary-based experiments I employed to address Research Questions 1, 2, and in part, Research Question 5 (cf. §1.5). The chapter comprises three subsections. The first subsection following this introduction details the methodology and findings of my first vocabulary-based experiment, designed to address Research Questions 1 and 2. To address these research questions, I analysed dictionaries of English, French and Polish. I first recorded the phonaesthemes' type frequencies in each language (the number of different words featuring each onset that also conveyed the meaning shown in §1.1.). I then wrote-down the three meanings most frequently expressed by words bearing each phonaestheme, to establish whether the onsets recurred more frequently with the meanings identified in Chapter 1 than with any other meanings; and if so, how much more frequently. Following this, the second subsection of the chapter details the methodology and findings of my second vocabulary-based experiment, designed to study the phonaesthemes' token frequencies in English. The token frequencies provide a representative measure of the extent to which speakers encounter the twelve phonaesthemes in everyday language use (see further discussion in § 4.3). In combination with the productive and type frequency data, the token frequencies of the phonaesthemes are used to address Research Question 5 at the close of Chapter 5. I undertook this stage of the investigation in English only because there was no motivation to predict that the phonaesthemes would be present in the vocabularies of French or Polish, or perceived by French and Polish speakers (cf. Chapter 3.3). To estimate the token frequencies of the twelve phonaesthemes, I analysed the British National Corpus of written and spoken English. I recorded the number of times each phonaesthemic relationship occurred in a random sample of 1,000 random words featuring its associated onset. Although the findings of this experiment are of no use in addressing Research Questions 1 and 2, it is most appropriate to outline its methodology and findings in this chapter, because the experiment involved text-based analysis in a similar way to the study of dictionary headwords. The fourth and final subsection of the chapter summarizes how the findings of the first vocabulary-based experiment address Research Questions 1 and 2.

## 4.1) Vocabulary-Based Experiment for Research Questions 1 & 2: Analysis of Dictionary Headwords

### 4.1.1) Methodology

My first vocabulary-based experiment involved analysing the vocabularies of English, French and Polish, with the aim of calculating the type frequencies of the twelve phonaesthemes in each language. Type frequency refers to the number of different words in which an onset recurs with its associated meaning identified in Chapter 1. For this experiment I used an English dictionary and bilingual French-English and Polish-English dictionaries. It was necessary to use bilingual French and Polish dictionaries, because I needed to be able to understand the definitions of the words listed (see further discussion below). For reasons of space, bilingual dictionaries usually contain fewer headwords than monolingual dictionaries. This is because bilingual dictionaries usually list vocabulary alphabetically in the first language and translate it into the second; then alphabetically in the second language and translate it into the first. However, in order to ensure comparability of the phonaesthemes' type frequencies in English, French and Polish, I needed to sample a relatively equal number of headwords in each language. Therefore, I chose to study the Concise Oxford English Dictionary, the bilingual Collins-Robert English-French dictionary, and the Oxford PWN English-Polish dictionary. All three dictionaries contained similar numbers of headwords, at between 300,000-400,000 apiece. This allowed me to sample a similarly-sized corpus of vocabulary in each language. In addition, all three volumes were of similar publication date, which allowed me to analyse contemporary vocabulary in each language.

The first phase in this experiment involved listing every head lexeme beginning with each of the twelve onsets, along with their definitions. I repeated this process in English, French and Polish[7]. For the purposes of the present study, 'head lexemes' included only root forms of head lexemes. I did not include inflections of head lexemes, nor any instances where words were clearly derived from a head lexeme and retained (but did not extend) its semantic content; instances of so-called 'syntactic derivation' (Dixon, 1982:12). For example, I included the word 'glory' in the English list of /gl/ head lexemes, but excluded 'glorious',

[7] To ensure that I had correctly distinguished head lexemes from inflections and derivations, I asked native French and Polish speakers to cross-check the respective vocabulary lists.

because the adjective does not extend or modify the core semantic idea of <<exaltation and honour>> of the root noun from which it is derived. However, I included both 'glow' and 'glow-worm', as the latter extends the core semantic idea of <<gently emitting light>> to <<a female insect that emits a shining green light from its abdomen>> (OED). Excluding inflections and syntactic derivations reduced the likelihood of my counting different forms of the same words more than once; which could have made any of the twelve phonaesthemes appear more – or less – frequent in each language than was actually the case. I also excluded any clear instances of lexical borrowings (such as 'snob' in French, which is borrowed both formally and semantically from English). My reasoning behind this was that the inclusion of such lexical borrowings in a dictionary does not guarantee that these borrowings have become as widely integrated into either the vocabulary of a language or the discourses of its speakers as 'native' words (Fabiszak, 2005:1739). While this decision may have caused me to overlook a small number of phonaesthemic words, I considered this to be a preferable alternative to including words that speakers might not regard as belonging to their [L1] language.

The next phase of analysis involved examining the twelve lists of headwords in each language, and highlighting any instance where a head lexeme – in any of its senses – expressed the meaning associated with its onset as identified in Chapter 1. To ensure possible replicability of the study, these highlighted lists are presented as Appendices 2-4[8]. Using these [now highlighted] lists, I then counted the number of head lexemes exhibiting each phonaestheme in English, French and Polish. This data constituted the phonaesthemes' type frequencies in each language. I recorded this data in a table (Table 13, below), expressing the type frequency of each phonaestheme as a fraction of the total number of headwords bearing its respective onset. This quantified the extent to which each onset recurred with the meaning under test in each of the three languages. A number of scholars in the literature have followed similar experimental methods to measure the extent to which phonaesthemes are visible in the vocabulary of a language; including Bowles (1995) and Abelin (1999) (cf. Chapters 2 and 3). The only difference between my own experiments and those of Bowles and Abelin is that I began with a predetermined list of phonaesthemes, and compared their presence in the vocabularies of the three languages; whereas Bowles and Abelin analysed the vocabularies of English and Swedish from scratch in order to find phonaesthemic relationships.

[8] See attached CD-Rom.

To ensure replicability of the present study, Table 12, below, shows the criteria on which I judged whether each head lexeme expressed the meaning associated with its phonaestheme. A set of criteria on which to judge the headwords was essential (cf. critique of Abelin, 1999; above). I could not simply look for certain key words or phrases in the definitions; this would have been too restrictive. For example, I could not simply look for the expressions "pejorative" or "unpleasant" in the definitions of /sl/ words in order to judge whether they exhibited the /sl/-<<pejoration>> phonaestheme. This is because there are many /sl/ words that are clearly <<pejorative>> in at least one sense (such as 'sly', 'slurry' and 'slog' in English); but whose definitions do not use the words 'pejorative', 'unpleasant' or even 'offensive' (cf. OED; see also critique of Abelin's experiments in §3.2). To varying extents, the same can be said for many of the phonaesthemes. Therefore, rather than classifying the head lexemes by searching for specific words in their definitions, I had to think more widely about the ways that the meanings associated with each phonaestheme could be expressed. To this end, I applied the following broad operationalisations of the meanings identified in Chapter 1 (see Table 12, overleaf):

Table 12: The criteria on which I judged whether head lexemes exhibited each phonaestheme

| Prosody | Operationalisation of Meaning Association |
|---|---|
| /sl/ | Any action, quality, person or thing which can cause any type of harm or bring about an effect that is culturally perceived (at the time of writing) to be disadvantageous to anyone or anything; any expression which can in any sense be used to offend, criticise, insult or degrade; any action, quality, person or thing which is in any sense culturally perceived as unpleasant at the time of writing. |
| /sm/ | |
| /sn/ | |
| /stɹ/ | Any action, quality, person or thing which is straight, or lacking in curves; long; stretched or under strain in any sense. |
| /gɹ/ | Any action, quality, person or thing which either physically or metaphorically grips, holds, pinches, clasps or fastens securely onto someone or something. |
| /gl/ | Any action, quality, person or thing which either emits or reflects light, or has the ability to see or affect the process of sight; any action pertaining to the emission or reflection of light or to the act of seeing; any quality of emitted or reflected light, or sight. |
| /kɹ/ | Any action, quality, person or thing which is curved or bent; wound-up; lacking in straightness; not running in a singularly linear direction; any action, quality, person or thing that is either literally or metaphorically (e.g. criminally) crooked. |
| /kl/ | Any action, quality, person or thing which has the quality of being physically thick (in the sense of density); any action, quality, person or thing having the property of being uncoordinated, clumsy or unsteady; any action, quality, person or thing which is culturally perceived to be stupid or unintelligent at the time of writing; the process of being thick (i.e. dense); the process of acting in an uncoordinated, clumsy or unsteady manner; the process of acting in a manner culturally perceived as stupid or unintelligent at the time of writing; or the properties of thickness (density), clumsiness, lack of co-ordination and stupidity in themselves. |
| /sw/ | Any action, quality, person or thing that either is in itself, or moves with, a swift/brisk or strong motion; any action, quality, person or thing that is culturally perceived as fast or strong (both physically and metaphorically) at the time of writing. |
| /skw/ | Any action, quality, person or thing which is timid, frail or weak; or which is culturally perceived as being timid, frail or 'weak' (either literally or metaphorically) at the time of writing. Any action, quality, person or thing which, by virtue of being timid, frail or weak, might be culturally perceived as being 'cute' at the time of writing. |
| /tw/ | Any action, quality, person or thing which is either physically or metaphorically small; or which is culturally perceived as being small at the time of writing. |
| /st/ | Any action, quality, person or thing which is, either physically or metaphorically, strong and resilient – whether in material or of character/personality; or which is culturally perceived as being strong or resilient at the time of writing. |

Having calculated the phonaesthemes' type frequencies, I then used the lists of head lexemes to establish the three most frequently-recurring meanings among words featuring each onset[9]. This involved a number of stages. The first stage was to identify and write a gloss of the main semantic field(s) of every head lexeme. A semantic field is a broad 'conceptual area' (Faber and Usón, 1999:67-68) or underlying meaning (Jackson and Ze Amvela, 2000:107) to which a series of words are related. For example, the words 'chef', 'fork', 'plate' and 'wine' are all related to the broad idea of <<food>>. Therefore, <<food>> constitutes a semantic field. Having identified and written a gloss of the main semantic field(s) of every head lexeme in the word-lists, I then counted the number of times each distinct semantic field (re-)occurred. I repeated this process in the vocabularies of English, French and Polish. The idea behind this phase of the experiment was to reach a more informed judgment about the presence of the phonaesthemes in each vocabulary, by establishing whether the onsets recurred with the meanings under test more frequently than with any other meaning(s).

### 4.1.2) Results

Table 13, overleaf, shows the type frequencies of the twelve phonaesthemes in English, French and Polish. This is the first stage in quantifying the phonaesthemes' presence in the vocabularies of the three languages; one of the main gaps in the current literature (cf. § 1.2).

[9] I chose to note the three most common meanings expressed (rather than the five or ten most common, or listing every single meaning with which the onsets occurred) for methodological efficiency.

Table 13: Type Frequencies of the Phonaesthemes in English, French and Polish (ranked by number of recurrences in English)

| Onset Prosody | Type Frequency of phonaestheme (as fraction of total head lexemes) - ENGLISH | Type Frequency of phonaestheme (as fraction of total head lexemes) - FRENCH | Type Frequency of phonaestheme (as fraction of total head lexemes) - POLISH |
|---|---|---|---|
| /sl/ | 63/148 | 2/12 | 3/12 |
| /st/ | 53/417 | 23/106 | 39/360 |
| /sn/ | 37/76 | 3/3 | 4/13 |
| /gl/ | 30/96 | 7/76 | 5/39 |
| /sm/ | 26/68 | 0/5 | 17/66 |
| /sw/ | 24/94 | 1/5 | 0/110 |
| /stɹ/ | 22/114 | 7/34 | 5/162 |
| /kɹ/ | 22/319 | 21/281 | 48/287 |
| /kl/ | 16/188 | 0/153 | 3/156 |
| /tw/ | 11/36 | 0/4 | n/a |
| /gɹ/ | 10/344 | 2/268 | 3/243 |
| /skw/ | 7/49 | n/a | 0/4 |

The type frequencies provide clear numerical evidence that all twelve phonaesthemes are observable in the present-day vocabulary of English. Every single onset recurs with its meaning association identified in Chapter 1; the basic prerequisite of a phonaesthemic pattern (cf. § 2.3). Even the phonaestheme which attains the lowest type frequency, onset /skw/, recurs seven times with its associated meaning of <<weakness and cuteness>>. Nevertheless, some phonaesthemes appear to be more pervasive associations in English than others. At the upper extreme, the /sl/-<<pejoration>> phonaestheme recurs sixty-three times in the 148 /sl/ words. Over a third of all /sl/ head lexemes sampled express a broadly <<pejorative>> meaning; the greatest frequency of recurrence of any phonaestheme under test. The fact that this association is so prevalent in the vocabulary of English could explain why it is one of the most frequently-cited English phonaesthemes in the literature (cf. Table 7). Also of note are the /sm/-<<pejoration>>, /sn/-<<pejoration>> and /gl/-<<light>> phonaesthemes. Although their type frequencies are not as high as /sl/, there is similarly clear evidence for these phonaesthemes in proportional terms. Like /sl/, over a third of all /sm/ and /sn/ head lexemes

express something <<pejorative>>; and nearly a third of all /gl/ head lexemes express meanings related to <<light>>. Overall, only two phonaesthemes recur in less than ten percent of their headwords; the /gɹ/-<<gripping>> and /kl/-<<clumsiness, thickness and stupidity>> phonaesthemes. Referring back to Table 7 (p. 68), the type frequencies of these phonaesthemes appear to be congruent with the fact that both associations are among the least frequently-cited of the phonaesthemes under test. Only Firth (1930) and Bowles (1995) discuss the association between /gɹ/ and <<gripping>>; while Firth (1935) alone suggests an association between /kl/ and <<clumsiness, thickness and stupidity>>.

The non-English type frequencies indicate that many of the phonaesthemes are also observable in the vocabularies of French and Polish. In French, seven of the twelve onsets recur with their associated meanings identified in Chapter 1; the exceptions being /sm/, /sw/, /kl/, /tw/ and /skw/ (/skw/ does not occur in onset position in French). Similarly, in Polish, nine of the twelve onsets recur with their associated meanings; the only exceptions being /sw/, /skw/ and /tw/ (/tw/ does not occur in onset position in Polish). It appears that the /st/-<<strength and stoicism>> and /kɹ/-<<crookedness>> phonaesthemes are particularly pervasive in the vocabularies of French and Polish. In both languages, these phonaesthemes are by far the most recurrent; attaining type frequencies greater than twice those of the next most frequently-recurring phonaesthemes. Indeed, the type frequency of /kɹ/ in Polish is more than double that in English; and in French, it is only a single word lower than in English. However, despite this, all of the onsets except /kɹ/ are more pervasive in the vocabulary of English than in French or Polish. Every single phonaestheme attains a higher type frequency in English than in French; and only the /kɹ/-<crookedness>> phonaestheme attains a higher type frequency in Polish than in English. Further, where all twelve phonaesthemes attain type frequencies of seven or more in English, only three of the phonaesthemes recur more than seven times in Polish (/st/, /kɹ/ and /sm/); and only two recur more than seven times in French (/st/ and /kɹ/).

To reach a more informed interpretation of the type frequency data, Table 14, overleaf, shows the three most common meanings (semantic fields) expressed by words featuring each onset in English, French and Polish. This represents the second stage in quantifying the phonaesthemes' presence in the vocabularies of the three languages.

Table 14: The three most common meanings expressed by words with each onset in English, French and Polish (number of head lexemes expressing each meaning)

| ONSET | ENGLISH | FRENCH | POLISH |
|---|---|---|---|
| **/sl/** | 1. pejoration (63)<br>2. shape (e.g. sleek, slender) (11)<br>3. sleeping/bodily functions (8) | 1. pejoration (2)<br>-<br>- | 1. pejoration (3)<br>-<br>- |
| **/sm/** | 1. pejoration (26)<br>2. smoke/smoking (7)<br>3. facial expressions (e.g. smile, smug) (4) | - | 1. pejoration (17)<br>2. tasting/culinary (13)<br>3. slimy/glutinous substances (11) |
| **/sn/** | 1. pejoration (37)<br>2. nasal/oral noises (11)<br>3. animals (6) | 1. pejoration (3)<br>-<br>- | 1. pejoration (4)<br>2. paper production (2)<br>3. spinning/revolving (2) |
| **/st/** | 1. strength/stoicism (53)<br>2. stars/celestial (12)<br>3. stones/stone constructions (10) | 1. strength/stoicism (23)<br>2. pointed objects (6)<br>3. pejoration (4) | 1. strength/stoicism (39)<br>2. old fashioned/aged (32)<br>3. loud noises (9) |
| **/stɹ/** | 1. straightness/stretching (22)<br>2. pejoration (11)<br>3. constriction/restriction (9) | 1. straightness/ stretching (7)<br>2. geology/weather (5)<br>3. light/sight (2) | 1. negative emotions (29)<br>2. guns/ammunition/use thereof (13)<br>3. straightness/stretching (5) |
| **/gl/** | 1. light/sight (30)<br>2. medical (e.g. gland, glaucoma) (14)<br>3. chemicals/chemistry (8) | 1. ice/water (7)<br>2. light (7)<br>3. sticky/glutinous substances (7) | 1. soil/clay products (8)<br>2. pejoration (7)<br>3. light (5) |
| **/gɹ/** | 1. pejoration (30)<br>2. gripping (10)<br>3. oral noise/expression (8) | 1. writing/drawing/engraving (13)<br>2. largeness (7)<br>3. incremental increases/decreases (4) | 1. flora (i.e. trees, soil, fruit etc.) (15)<br>2. writing/engraving (14)<br>3. fatness (6) |

(Table continues overleaf)

| ONSET | ENGLISH | FRENCH | POLISH |
|---|---|---|---|
| **/kl/** | 1. clumsiness/thickness/stupidity (16)<br>2. grabbing/clamping (9)<br>3. clarity/clearness (6) | 1. emission of noise /noisy objects (19)<br>2. confined/secluded/segregated (7)<br>3. valves/flaps/dividers (3) | 1. adhesion/stickiness (10)<br>2. slapping/smacking (9)<br>3. emission of noise /noisy objects (8) |
| **/kɹ/** | 1. nature (46)<br>2. food/drink/consumption (37)<br>3. pejoration (34) | 1. pejoration (46)<br>2. crookedness (21)<br>3. roughness/crustiness (18) | 1. pejoration (61)<br>2. crookedness (48)<br>3. crystalline (6) |
| **/sw/** | 1. Fast/strong movement (26)<br>2. food/drink/consumption (14)<br>3. nature (11) | - | 1. speech/hearing (22)<br>2. fatness/elephantine (9)<br>3. sweetness (of sugar etc.) (6) |
| **/tw/** | 1. smallness (11)<br>2. twistedness (5)<br>3. time (4) | 1. twistedness/twisting (2)<br>-<br>- | n/a |
| **/skw/** | 1. weakness (esp. squeamishness) (7)<br>2. pejoration (5)<br>- | n/a | - |

In many respects, the findings in Table 14 support the type frequency data. Once again, these findings show clear evidence of all twelve phonaesthemes in the vocabulary of English. For ten of the onset prosodies, the meaning under test recurs more frequently among their headwords than any other meaning(s):

- English head lexemes featuring onsets /sl/, /sm/ and /sn/ express <<pejorative>> meanings more frequently than any other;
- English head lexemes featuring onset /st/ express <<strength and stoicism>> more frequently than any other meaning;
- English head lexemes featuring onset /stɹ/ express <<straightness and stretching>> more frequently than any other meaning;
- English head lexemes featuring onset /gl/ express <<light>> more frequently than any other meaning;
- English head lexemes featuring onset/kl/ express <<clumsiness, thickness and stupidity>> more frequently than any other meaning;
- English head lexemes featuring onset /tw/ express <<smallness>> more frequently than any other meaning;
- English head lexemes featuring onset /skw/ express <<weakness and cuteness>> more frequently than any other meaning;
- English head lexemes featuring onset /sw/ express <<fast or strong movement>> more frequently than any other meaning;

The only prosodies which do not recur most frequently with the meanings under test are /gɹ/ and /kɹ/. The meaning of <<crookedness>> is actually the fifth most-common meaning among English /kɹ/ headwords; semantic fields of <<nature>>, <<food/drink/consumption>>, <<pejoration>> and <<flora>> all recur more frequently (cf. Chapter 6.2). Similarly, the meaning <<gripping>> is only the second most-common meaning among English /gɹ/ head lexemes; a <<pejorative>> semantic field recurs more frequently.

In general, the French and Polish data in Table 14 also supports the type frequency data analysed above. Once again, there is clear evidence of a number of the phonaesthemes in the vocabularies of French and Polish. For instance:

- Onsets /sl/ and /sn/ recur more frequently with <<pejoration>> than with any other meaning in the vocabularies of both French and Polish;
- Onset /st/ recurs more frequently with <<strength and stoicism>> than with any other meaning in the vocabularies of both French and Polish;
- Onset /gl/ recurs more frequently with <<light>> than with any other meaning in the vocabularies of both French and Polish;
- Onset /stɹ/ recurs more frequently with <<straightness and stretching>> than with any other meaning in the vocabulary of French; while <<straightness and stretching>> is the third most frequently-recurring meaning among /stɹ/ headwords in Polish;
- Onset /sm/ recurs more frequently with <<pejoration>> than with any other meaning in the vocabulary of Polish;
- Onset /kɹ/ recurs frequently with <<crookedness>> in both French and Polish.

However, Table 14 also reinforces the observation that not all the phonaesthemes are as clearly evident in French and Polish as they are in English. For a number of the onsets, the meanings under test are not among those most frequently-recurring among their head lexemes. For instance, <<clumsiness, thickness and stupidity>> is not one of the three most common meanings among /kl/ headwords in either French or Polish. A similar situation is seen for a further four phonaesthemes:

- Onset /gɹ/ and <<gripping>> in both languages;
- Onset /sw/ and <<fast or strong movement>> in both languages (there are actually very few /sw/ head lexemes in French, which explains the lack of recurring meanings here);
- Onset /skw/ and <<weakness; cuteness>> in Polish (no onset /skw/ in French); and
- Onset /tw and <<smallness>> in French (no onset /tw/ in Polish).

Taken at-large, the findings in Tables 13 and 14 suggest that, while a number of the phonaesthemes are clearly evident in the vocabularies of French and Polish, the twelve relationships are generally more pervasive in English. Not only are the phonaesthemes' type frequencies almost exclusively higher in English than in French or Polish (with the exception of /kɹ/ as discussed above); but for all twelve prosodies in English, the meanings outlined in

§ 1.1 are among the most frequently-recurring meanings among their headwords. Indeed, for ten of the twelve phonaesthemes, the meanings under test recur more frequently among their headwords than any other meaning. The same is clearly not true in French and Polish. Of course, these findings do not provide any guarantee that the phonaesthemes will be more productive in English than in French or Polish; or that the phonaesthemes will be productive at all. This is something that only the experiments presented in Chapter 5 can address.

The next section in this chapter presents the methodology and a brief overview of the findings of the second and final vocabulary-based experiment I implemented. As previously discussed, the purpose of this experiment was to generate a representation of the phonaesthemes' token frequencies in English; that is, an indication the frequencies with which English speakers encounter the twelve associations in day-to-day language use. This data is compared against the phonaesthemes' type frequencies in English and their overall frequencies of productivity in English in Chapter 5, to establish whether it is the type or token frequency of a phonaestheme (or both, or neither) that most appears to affect its productivity. This comparison addresses Research Question 5 (cf. § 1.3). I undertook this stage of the investigation in English only because there was no motivation to predict that the phonaesthemes would be present in the vocabularies of French or Polish, or perceived by French and Polish speakers (cf. § 3.3). Although the findings of this experiment are of no use in addressing Research Questions 1 and 2, it is most appropriate to outline its methodology and findings in this chapter, because the experiment involved text-based analysis in a similar way to the study of dictionary headwords above. However, the presentation of its findings is necessarily brief here. The main discussion of this data is given in § 5.4, after the discussion of the productive data.

## 4.2) Vocabulary-Based Experiment to Establish Token Frequencies of Phonaesthemes in English: Corpus Analysis

### 4.2.1) Methodology

To calculate a representation of each phonaestheme's token frequency in English, I used a corpus methodology. I consulted the British National Corpus ('BNC'), 'a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English' (BNC, 2009; cf. Biber, 1993 for discussion on corpus size and representativeness). Using the BNC, I measured the proportion of words with each onset that also expressed the meaning under test in a sample of actual language use. For each phonaestheme, I began by generating a random sample of 1,000 words containing the onset. I chose not to filter or restrict the search to particular word classes; the idea being to generate a sample of all the lexical contexts in which speakers tend to encounter the twelve onsets. I selected a sample of 1,000 instances because many corpus studies seem to regard this number as being sufficient to generate a representative sample (cf. Hanks, 2004: 91; Deignan, 2005: 155-157). I then asked the corpus software to display each of these 1,000 words, and their surrounding linguistic contexts, in a separate line of text (a concordance line). Each concordance line showed one of the 1,000 words in the centre, with its surrounding linguistic context to the left and the right (this format is termed 'Key Word In Context', or 'KWIC'; for a fuller discussion, see Russell-Bernard and Ryan, 2010: 192-193). The full sample of concordance lines for each onset are included as Appendices 5-16 on the attached CD-Rom.

Having generated the concordance lines for each phonaestheme, I then counted how many of the 1,000 randomly-chosen words expressed the meaning associated with their respective onset (cf. Chapter 1, pages 8-9). These are displayed as highlighted concordance lines in Appendices 5-16. I judged any word to have exhibited the associated meaning if it was one of the head lexemes (or an inflection or derivation thereof) I had identified in the dictionary entries as demonstrating the phonaestheme, using the criteria in Table 12. I then recorded these frequencies in a table. These are discussed in Chapters 5 and 6.

I calculated the token frequency for all twelve phonaesthemes following the procedure outlined above. However, for onsets /sl/, /sm/ and /sn/, one further stage in the

analysis was required. This was because these onsets are associated with the meaning of <<pejoration>>, which in some cases can be dependent on the context of utterance as well as the literal meaning of the particular words. To clarify: In analysing the dictionary entries, the criteria I used to determine whether the lexeme expressed something <<pejorative>> was "any action, quality, person or thing which causes any type of harm to anyone or anything, or which is culturally perceived as inherently 'unpleasant'". On the basis of these criteria, I judged lexemes such as 'smell' to be broadly <<pejorative>>, since this expression can often carry culturally unpleasant connotations such as <<bad odour>>; particularly when used as a noun. However, this does not predicate that every use of the word 'smell' need be associated with something unpleasant or pejorative; a reference could be made to a 'nice smell', for example. I found a number of such cases throughout the corpus samples. For instance, one of the concordance lines for /sm/ refers to 'the smell of Lagerfeld's favourite opopanax candles' (Appendix 6, p.14, line 6); while another reads '….the bread smells wonderful. He sighed' (Appendix 6, p.1, line 37). Words with the other nine onsets are not subject to the same context-dependent meaning as /sl/, /sm/ and /sn/ words. For example, every use of the word 'glow' necessarily denotes the property <<light>>, irrespective of the surrounding linguistic context; just as every use of the word 'straight' denotes the property <<straightness>>. To resolve this potential issue of context-dependence, I also examined the surrounding lexical context of words beginning /sl/, /sm/ and /sn/. I excluded from the token frequencies any instance where a supposedly <<pejorative>> word was clearly used in a context where the association was not <<pejorative>> or <<unpleasant>>, or where the association was unclear. While my own interpretation of <<pejorative>> contexts introduces a degree of subjectivity, I considered this a preferable alternative to including words in the token count that were clearly not being used in a pejorative sense – as demonstrated by the above example of 'smell'.

#### 4.2.2) Results

Table 15, overleaf, shows the token frequencies of the twelve phonaesthemes calculated from the BNC samples:

Table 15: The phonaesthemes' token frequencies in English (ranked in descending order)

| **PHONAESTHEME (Appendix no.)** | **TOKEN FREQUENCY** |
|---|---|
| /gl/-<<light>> (10) | 459/1000 |
| /sn/-<<pejoration>> (7) | 371/1000 |
| /sl/-<<pejoration>> (5) | 201/1000 |
| /sw/-<<strong/fast movement>> (15) | 197/1000 |
| /stɹ/-<<straightness and stretching>> (8) | 191/1000 |
| /st/-<<strength and stoicism>> (9) | 142/1000 |
| /kɹ/-<<crookedness>> (13) | 134/1000 |
| /sm/-<<pejoration>> (6) | 113/1000 |
| /skw/-<<weakness and cuteness>> (14) | 55/1000 |
| /gɹ/-<<gripping>> (11) | 23/1000 |
| /kl/-<<clumsiness, thickness and stupidity>> (12) | 16/2000*[10] |
| /tw/-<<smallness>> (16) | 7/1000 |

Like the type frequency data, the token frequencies suggest that some phonaesthemes are more pervasive in English than others. At the upper extreme, the /gl/-<<light>> phonaestheme attained a token frequency of 459/1000 tokens (46%). This suggests that nearly half of the /gl/ words encountered by English speakers in day-to-day language tend to express a meaning related to <<light>>. This is followed by the /sn/-<<pejoration>> phonaestheme, which attained a token frequency of 371 (37%). This suggests that around a third of /sn/ words encountered by English speakers in day-to-day language tend to express a broadly <<pejorative>> meaning. At the opposite pole, the /tw/-<<smallness>> phonaestheme attained a token frequency of only 7 recurrences in the sample of 1000 /tw/

[10] As shown in Appendix 12, I sampled 1,000 English words beginning with "cl" and 1,000 beginning with "kl" for onset /kl/, since this prosody can be spelled using either orthographic sequence. While this is also true of onset /kɹ/, the analysis of dictionary headwords revealed that there are no /kɹ/-<<crookedness>> words in English that use the graphological sequence "kr" (unlike /kl/-<<clumsiness>>, which is seen in words such as '**cl**umsy' and '**kl**utz'). As such, I only sampled orthographic "cr" words to establish the /kɹ/-<<crookedness>> token frequency.

words. That is to say, only around 1% of the /tw/ words encountered by English speakers in day-to-day language tend to express a meaning of <<smallness>>. Similarly, only around 2% of the /kl/ words in the sample expressed <<clumsiness, thickness and stupidity>>; and only 2% of /gɹ/ words expressed meanings related to <<gripping>>. If a phonaestheme's token frequency does in fact influence its perceptibility and productivity, then based on this data it would seem reasonable to suggest that the /gl/ and /sn/ phonaesthemes are more likely to be productive in English than the /gɹ/, /kl/ and /tw/ phonaesthemes.

A trend worthy of particular note in these results is that some phonaesthemes attain relatively similar type and token frequencies (in proportional terms), whereas others attain markedly different type and token frequencies. This is clearly demonstrable by comparing the type and token frequencies by percentage, as shown in Table 16, below:

Table 16: A comparison of the phonaesthemes' type and token frequencies in English (ranked by percentage type frequency)

| ONSET | TYPE FREQUENCY (%) | TOKEN FREQUENCY (%) |
|---|---|---|
| /sn/ | 49 | 37 |
| /sl/ | 43 | 20 |
| /sm/ | 38 | 11 |
| /gl/ | 31 | 46 |
| /tw/ | 31 | 1 |
| /sw/ | 26 | 20 |
| /stɹ/ | 19 | 19 |
| /skw/ | 14 | 6 |
| /st/ | 13 | 14 |
| /kl/ | 9 | 2 |
| /kɹ/ | 7 | 13 |
| /gɹ/ | 3 | 2 |

Table 16 shows that a number of the phonaesthemes attained broadly similar type and token frequencies in proportional terms. For instance, the /stɹ/-<<straightness and stretching>> phonaestheme was evident in 22 of the 114 /stɹ/ head lexemes (19%), and was exhibited by 191 tokens in the corpus sample (also 19%). Similarly:

- The /gɹ/-<<gripping>> phonaestheme attained a type frequency of three percent and a token frequency of two percent;

- the /st/-<<strength and stoicism>> phonaestheme attained a type frequency of thirteen percent and a token frequency of fourteen percent;
- the /kɪ/-<<crookedness>> association attained a type frequency of fourteen percent and a token frequency of thirteen percent;

However, the same cannot be said of all phonaesthemes. This becomes very clear if one compares the /sl/ and /gl/ phonaesthemes. The /sl/-<<pejoration>> phonaestheme attained the highest type frequency of all twelve onsets (cf. Table 13), and the second-highest type frequency in percentage terms. In contrast, the /gl/-<<light>> phonaestheme attained only the sixth-highest type frequency of the twelve onsets (cf. Table 13), and the fourth-highest in percentage terms. However, in the corpus sample, the proportion of the 1,000 /gl/ words expressing <<light>> is over double the proportion of the 1,000 /sl/ words expressing <<pejoration>>. In other words, the /sl/ phonaestheme has a higher type frequency than the /gl/ phonaestheme, yet the /gl/ phonaestheme attains a token frequency more than double that of the /sl/ phonaestheme. This suggests that, while the /sl/-<<pejoration>> phonaestheme is more pervasive in the vocabulary of English, the /gl/ words encountered by speakers are more likely to express <<light>> than any /sl/ words are to express <<pejoration>>. A similar situation is seen with the /tw/-<<smallness>> phonaestheme. In proportional terms, nearly a third of all /tw/ head lexemes expressed a meaning related to <<smallness>>. However, the proportion of /tw/-<<smallness>> words actually observed in the corpus sample is the lowest of any phonaestheme under test (as discussed above), at less than one percent. This suggests that although the phonaestheme is fairly pervasive among /tw/ head lexemes, it is very likely that speakers will tend to encounter /tw/ words which do not express meanings related to <<smallness>>.

The above analysis suggests that there does not necessarily appear to be any relationship or link between a phonaestheme's type and token frequency; at least in English. That is, the extent to which a phonaestheme is present in the vocabulary of the language does not appear to influence the degree to which speakers tend to encounter this sound-meaning association in everyday language use.

The discussion returns to the token frequency data in more detail in § 5.4. Before discussing the productive experiments, the next and final section in the present chapter

summarizes how the findings of the first vocabulary-based experiment address Research Questions 1 and 2.

### 4.3) Summary of Vocabulary-Based Findings

Research Question 1 asked: 'To what extent are the twelve phonaesthemes observable in the present-day vocabulary of English?' In response to this question, the type frequency data shows that there is some degree of evidence for all twelve phonaesthemes in the vocabulary of English: All twelve onsets *recur* with the meanings identified in § 1.1, and thus exhibit the necessary conditions to be termed 'phonaesthemes' (cf. § 2.3). However, the type frequency data indicates that some phonaesthemes are more pervasive associations in the vocabulary of English than others. The most frequently-recurring of the twelve phonaesthemes is the /sl/-<<pejoration>> association, which recurs in 63 of the 148 /sl/ head lexemes. This is followed by the /st/-<<strength and stoicism>> phonaestheme (53 recurrences) and the /sn/-<<pejoration>> phonaestheme (37 recurrences). Conversely, the least frequently-recurring phonaesthemes are the /skw/-<<weakness and cuteness>> association (7 recurrences); the /gɹ/-<<gripping>> association (10 recurrences) and the /tw/-<<smallness>> association (11 recurrences). However, despite the fact that some phonaesthemes attain much higher type frequencies than others, ten of the twelve prosodies recur more frequently with the meanings under test than with any other meanings. The only exceptions to this trend are onsets /kɹ/ and /gɹ/. In combination, then, these findings show evidence for all twelve phonaesthemes in the vocabulary of English; although they indicate that the /sl/-<<pejoration>> and /st/-<<strength and stoicism>> phonaesthemes are markedly more pervasive than the /skw/-<<weakness/cuteness>>, /gɹ/-<<gripping>>, /tw/-<<smallness>> and /kɹ/-<<crookedness>> phonaesthemes.

Research Question 2 asked: 'To what extent are the twelve phonaesthemes observable in the present-day vocabularies of languages other than English?' In response to this question, the type frequency data shows that many of the phonaesthemes are also clearly observable in the vocabularies of French and Polish, the two languages studied in addition to English. In French, seven of the twelve onsets recur with the meanings identified in § 1.1; the exceptions being /sm/, /sw/, /kl/, /tw/ and /skw/ (/skw/ does not occur in onset position in French). Similarly, in Polish, nine of the twelve onsets recur with the meanings identified in § 1.1; the only exceptions being /sw/, /skw/ and /tw/ (/tw/ does not occur in onset position in Polish).

As such, it is only the /sw/, /skw/ and /tw/ phonaesthemes that are not evident to some extent in either language. Notably, the /st/-<<strength and stoicism>> and /kɹ/-<<crookedness>> associations are particularly pervasive in the vocabularies of both languages. Indeed, the type frequency of /kɹ/ in Polish is more than double its type frequency in English. To a large extent, these findings are supported by the semantic field analysis. This reveals that, in both languages, four of the onsets (/sl/, /sn/, /st/, /gl/) recur more frequently with the meanings identified in §1.1 than with any other meanings. For a further two onsets (/stɹ/ and /kɹ/), the meanings identified in § 1.1 are among the most frequently-recurrent in the onsets' headwords. However, despite this evidence for these phonaesthemes in the vocabularies of French and Polish, eleven of the twelve associations are more pervasive (frequent) in the vocabulary of English. Every single phonaestheme attains a higher type frequency in English than in French; and only the /kɹ/-<crookedness>> phonaestheme attains a higher type frequency in Polish than in English. In addition, there are a number of phonaesthemes whose meaning associations are not among the most common semantic fields of their onsets in French and Polish. This is true of /kl/, /gɹ/ and /sw/ in both languages; as well as /skw/ in French and /tw/ in Polish. This contrasts with English, where, for every phonaestheme except /kɹ/, the meanings under test are within the two most recurrent semantic fields in their onsets' headwords. On balance, then, while there is evidence of a number of the phonaesthemes in the vocabularies of French and Polish, there is clearer evidence of all but one phonaestheme in English.

The next chapter presents the methodologies and findings of the productive experiments designed to address Research Questions 3-5.

# CHAPTER 5: PRODUCTIVE EXPERIMENTS

This chapter presents the methodologies and findings of the productive experiments I implemented. As outlined in Chapter 1, the purpose of implementing these experiments was to address Research Questions 3 and 4, by investigating whether the twelve phonaesthemes are patterns that are actually perceived by speakers of English and other languages (in this case, French and Polish). Research into speakers' perception of phonaesthemes constitutes one of the main gaps in the current literature, and is the primary aim of the present study (cf. Chapter 1). The present chapter consists of five subsections. Following this introduction, the first subsection discusses my pilot study. Taking cues from Abelin's (1999) experiment (cf. § 3.2.3), I constructed five different productive experiments. I trialled these experiments on a small number of respondents, in order to test the reliability of the methodologies and data collection methods, and to check whether respondents appeared to understand what was being asked of them. Upon analysing the results of the pilot study, and with the benefit of participant feedback, I made a number of changes for the main study. These included removing one experiment, simplifying another experiment, and clarifying several parts of the study that participants reported as being difficult to understand.

Following the discussion of the pilot study, the second subsection of the chapter presents the methodologies and findings of each of the four productive experiments I implemented as the main study. This is presented on an experiment-by-experiment basis. Having discussed the methodologies and findings of the experiments individually, the third subsection of the chapter then calculates the overall productivity of each phonaestheme in each language. This provides a cumulative picture of the most and least productive phonaesthemes in each language; something which is difficult to judge when analysing the findings of the individual experiments. It also enables a broad comparison of the productivity of the phonaesthemes across the three language cohorts. This three-way comparison is supplemented by statistical testing. For two of my four experiments, I statistically compared the results of English, French and Polish speakers. This increased the certainty with which I could conclude that certain phonaesthemes appeared to have been more productive in some languages than in others. Thus, the third subsection of the chapter concludes by outlining the statistical tests I performed, and discussing their findings.

In its fourth subsection, the chapter investigates whether it is the type or token frequency that most appears to affect the productivity (and thus perception) of the phonaesthemes. This is in response to Research Question 5, which asked about what sorts of factors might affect the extent to which the twelve phonaesthemes are perceived by speakers. (As discussed in § 4.3.1, this part of the investigation is only undertaken in English.) In this section, the phonaesthemes' type and token frequencies are compared to their frequencies of overall productivity. The aim of this is to reveal whether it is the type or token frequencies that are most similar to the phonaesthemes' productivity. The findings of this comparison suggest that neither the type nor token frequencies of the phonaesthemes are wholly responsible for their productivity (and thus perception). As such, two other possible factors - labelled "multiple semantic recurrence" and "multiple phonetic recurrence" – are considered in Chapter 6. "Multiple semantic recurrence" is the label given to the idea that one [onset] prosody could recur with the expression of many different meanings; while "multiple phonetic recurrence" is the label given to the idea that one meaning could recur with multiple [onset] prosodies.

The fifth and final subsection of the chapter summarises how the findings of the productive experiments, both individually and cumulatively, address Research Questions 3-5.

### 5.1) The Pilot Study

The first phase in my design and implementation of the productive experiments was to trial a number of different test types in a small-scale pilot study. The main motivation for conducting a pilot study was the lack of existing literature to have investigated the perceptibility of phonaesthemes; and the fact that there were no established ways of testing this phenomenon that I could have applied or adopted. As a result, the methodologies I employed were almost entirely novel – albeit partially motivated by my critique of Abelin's (1999) experiments (cf. § 3.2) – and were very much exploratory. As such, I decided that it was most appropriate to conduct a pilot study before using any of my proposed tests to collect data for analysis, in order to see which of these tests appeared to be most effective in generating evidence about the productivity (perceptibility) of phonaesthemes.

In constructing the pilot study, the first methodological decision I had to make was to determine the most appropriate type of investigation for my research questions. Abbuhl, Gass

and Mackey (2013: 116) note that in making this choice, careful consideration must be given 'to the goals of the study…as well as to the advantages and disadvantages of the various data collection techniques that are available'. Fortunately, there is a larger body of literature concerning data collection techniques in linguistics than on the perception of phonaesthemes. Detailed, critical accounts of the most common approaches used in various fields of linguistics are provided by Wray and Bloomer (2006), Meyer and Nelson (2006), Dörnyei (2007), Litosseliti (2010) and Posdeva and Sharma (2013) among other scholars; and there are many more publications offering overviews of different data collection techniques in the social sciences and humanities at large (e.g. Blalock (1974), Weathington, Cunningham and Pittenger (2010), and van Peer, Hakemulder and Zyngier (2012)).

After consulting the available literature, I chose to collect my productive data using a survey. There are a number of other data collection methodologies I could have used – most notably open-ended or semi-structured interviews. However, I chose to use a survey because it offered the greatest number of advantages (relative to the goals of my study) of the various techniques discussed in the literature, which included intuition/introspection, survey questions and experiments, ethnography/fieldwork, structured or unstructured interviews and psycholinguistic experiments such as neuroimaging (cf. summary by Abbuhl, Gass and Mackey, 2013: 116-134). The aims of my study meant that it would not have been appropriate to use my own intuition or introspection as a means of collecting data. This would not have revealed anything about the perception of the phonaesthemes in any language other than my own opinions, which would inevitably have been biased towards recognising the twelve patterns after engaging with the literature in Chapter 2. Equally, it would have been inappropriate to conduct an ethnographic study into the perception of the phonaesthemes. The logistics involved would have made this impossible. In the time I would have been able to allocate for collecting data, I would have had no guarantee of observing situations in which everyday speakers of the three languages would have encountered the twelve patterns; less still situations where they would have had the opportunity to show any explicit awareness that they recognised these patterns. Moreover, as neither a fluent French or Polish speaker, it is unlikely I would have been able to conduct an ethnographic study in either of these languages anyway. And while psycholinguistic experiments such as neuroimaging might well be useful for studying phonaesthemes in the future – for instance, one could measure neural activity when exposing speakers to the patterns – this approach would not have enabled me to address the goals of the present study: to explore the extent to

which English, French and Polish speakers perceived some sort of [connotative] affinity between the twelve meaning associations and prosodies that was not present with other prosodies, and whether they perceived some sort of [connotative] affinity between the twelve prosodies and the meaning associations that was not present with other meanings. With the benefit of hindsight, it might have been useful to conduct some form of open-ended, face-to-face data collection methodologies (such as interviews or focus groups) alongside the survey, as these could have allowed me to explore qualitatively, with participants, the reasons *why* they may have given particular answers, and *why* they may have associated particular prosodic forms with particular meanings. I return to this point in more detail in Chapter 6.5 when I discuss the limitations of the study, and in Chapter 7.2 when I outline possible directions for future research in this field. However, while face-to-face interviews would have offered this key affordance, they would also have created similar logistical difficulties as an ethnographic study. The time constraints of the project would have made it very difficult to conduct face-to-face interviews with ninety people, then transcribe the data and analyse the discourse for themes and patterns. Moreover, as a non-fluent French and Polish speaker, it would have been impractical for me to have personally conducted interviews with French and Polish respondents. Instead, I would have had to recruit an individual to conduct these interviews on my behalf, which would have added substantially to the time and financial costs of the project and would have introduced a further source of subjectivity to the data collection process.

Choosing to collect my productive data by way of a survey offered a number of key advantages over other data-collection methodologies. Unlike introspection, a survey allowed me to explore linguistic features and patterns (in my case, the perception of phonaesthemes) among multiple, real-life speakers; rather than relying solely on my own intuition (Meyer and Nelson, 2006:99). Using a survey also presented fewer logistical difficulties than ethnographic or interview-based approaches. It allowed me to collect data from afar (Schilling, 2013: Ch.6), eliminating the need for face-to-face or one-to-one contact with the individuals I sampled. In turn, this meant I did not need to travel to French and Polish speech communities, or acquire linguistic competence in French or Polish within an impractically short time. Instead, I was able to arrange to have the survey professionally translated before its distribution to the non-English speaker groups (see discussion in § 3.3). In combination, these factors substantially reduced the temporal and financial pressures on the project. Finally, unlike psychological or psycholinguistic methodologies, a survey allowed me to

conduct experiments that were appropriate to my research questions; namely, experiments to measure the extent to which speakers associated particular prosodies with particular meanings (and vice versa); and the numbers of speakers to do so in each case.

Having determined that a survey was the most appropriate type of methodology for my particular study, I then had to decide whether I would use an electronic (online) or paper-based survey. Dörnyei and Taguchi (2010: 63-72) discuss the advantages and disadvantages of different types of survey administration, and argue that the electronic approach offers a number of key advantages, irrespective of one's research aims. These advantages include 'easy access to populations who would otherwise be difficult to reach' (ibid: 69), as well as savings in time and costs at both the distribution and data-entry stages. These savings arise principally from the absence of postal delays or charges, the fact that there is no need to manually input data into a computer for analysis (or pay someone to do the same), since responses are automatically collected and stored; and the ability to instantly track the number of completions and responses to the survey (rather than having to wait for a period of time for surveys to be returned in order to see how many people have responded). Together, these advantages recommended the use of an online survey in my study, since I was to be sampling informants from a number of areas across the United Kingdom, France and Poland. By using an online survey, I could distribute the experiments instantly and cheaply to these areas, track the number of completions instantly and accurately (distributing further surveys electronically if necessary), with the knowledge that all completed surveys would be ready for analysis immediately with no further data-entry required, which could have risked researcher error.

In addition to the advantages discussed by Dörnyei and Taguchi, above, an online survey offered a further key advantage with respect to the goals of my study:

- The key affordance of using an electronic survey, and the factor which ultimately motivated my decision to adopt this approach, was that it allowed me to incorporate both recordings and sound files into the questions, by means of embedded hyperlinks. As the purpose of my study was to test the connotations ascribed to coined words on the basis of their **phonetic structures**, it was crucial to provide these words in spoken form rather than written, so as to ensure that every informant was responding to identical phonetic stimuli (cf. critique of Abelin's (1999) tests '1a', '1b', '2a' and '2b' in Chapter 3). Notably, there would have been no way to feasibly incorporate spoken

words in a paper-based survey. This would have been possible had I conducted an interview-based questionnaire, as I could personally have uttered the coined words I was asking respondents to judge. However, face-to-face interviews would have been logistically difficult – particularly with French and Polish respondents – and far more time consuming, as outlined above. Moreover, this could have introduced a further source of subjectivity and variability into the study, as there would have been no guarantee that I would have produced each coined word phonetically identically (with the same intonation, pitch and so on) for each informant. It is possible that such variations could have affected the connotations ascribed to these phonetic forms by respondents. On balance, then, an online survey offered the only reasonable means of presenting all ninety informants with identical **phonetic** stimuli in each question – an absolute necessity – within the constraints of the project.

Of course, while there were many advantages to using an online survey in my particular study, this does not entail that this method is completely without its disadvantages. Notably, Dörnyei and Taguchi (2010: 71), in common with Bryman (2008) and Shih and Fan (2008), suggest that response rates tend to be lower in online surveys by comparison to paper-based surveys; and that online surveys tend to attract a higher response rate among younger individuals – particularly those in their late teens and twenties. This second observation is cited in much of the literature pertaining to survey design. For instance, Kays, Keith and Broughal (2013:163) stress that when implementing an online survey, it is very important 'to consider the…computer familiarity of those that will be completing [it]', since one can tend to expect 'higher rates of Web-survey responses by younger participants, and…higher rates of pencil-and-paper survey responses by older participants' (ibid; see also Börkan, 2010). However, while these factors could have presented disadvantages to some studies, they were less problematic for my investigation. The fact that I only sought thirty participants in each respondent group, and could track the number of completions instantly and accurately (distributing further surveys if required), meant that the potential for a lower response rate was less of a concern than it may have been if I had conducted a larger-scale study in a single language. Moreover, while it may be that online surveys tend to favour a younger respondent demographic[11], this did not pose a problem for my study since age was not a variable under

[11] Nevertheless, Crystal (2001) points out that many more people are making use of the Internet and computer-mediated technology with every passing year, and many people of older generations are becoming

investigation in this project. This is because there is currently no motivation in the literature to predict that speakers' age would have affected their perception of phonaesthemic patterns (assuming all respondents were adults with the level of L1 proficiency normally expected in adulthood). For example, there is no motivation to suggest that a 20-year old English L1 speaker would be any less likely to perceive the /sl/-<<pejoration>> phonaestheme than a 70-year old English L1 speaker (however, see possible hypotheses of age and education level affecting phonaestheme perception in Chapter 7.2).

Since the possible disadvantages of online surveys identified in the literature did not present serious drawbacks to my study, and did not detract from the key affordances discussed above, I decided to use an online (web-based) survey to administer my productive experiments in both the pilot and main studies. Having taken this decision, I then began designing the questions to be trialled in the pilot study (discussed in 5.1.1-5.1.6, below), and determining the number and native language(s) of the respondents I would sample for this exercise. For the pilot study, I chose to survey only English- and Polish-speaking informants. It was sufficient to test only two languages in the pilot, since the aim here was not to generate findings about the phonaesthemes themselves. Rather, the aims of the pilot study were:

- Principally, to test whether my proposed experiments appeared to be reliable, given that they were almost entirely novel with no existing precedents in the literature;
- To test whether my proposed experiments could actually be implemented with the electronic survey software;
- To test that respondents appeared to understand what was being asked of them;
- To ensure that the electronic survey software operated as intended.

Indeed, my only reasons for including Polish were to verify that the survey software was capable of displaying graphemes other than those in the standard Latin alphabet (such as the Polish "ł"); and to verify that the software was capable of collecting responses from Internet users in different countries.

From each language, I sought five participants, taking care to make a note of the

---

increasingly computer-literate.

people I had sampled in order to avoid recruiting these respondents again for the main study. All English respondents were my own acquaintances; all Polish respondents were known to my doctoral colleague who translated the study into Polish. Surveying only five participants meant that I was able to analyse and reflect on the pilot study relatively quickly. This allowed me more time to design and implement any modifications before distributing the main study. The survey was totally anonymous; I did not collect any personally-identifiable information from participants. This was a decision I also took in the main study. My choice to make the survey anonymous was motivated by Dörnyei and Taguchi (2010:72), who suggest that assurances of anonymity tend to increase response rates in online surveys. As such, I informed participants that I would have no way of identifying them personally from the responses they provided. In line with the ethical guidelines of my institution, I also informed participants that their consent was entirely voluntary and that they were free to withdraw from the survey at any point.

I chose not to inform respondents about the research aims until after the pilot study data had been gathered. This was a process I later repeated in the main study. The main motivation behind this decision was to try to minimise any possible acquiescence bias in the data. 'Acquiescence bias' refers to the idea of participants providing the answers that they think will be most beneficial to the researcher (Wagner, 2010:35). Therefore, in the introduction to the survey, I simply stated that I was investigating the processes of naming and describing things, which could have potential implications in a number of domains; including advertising and the study of literary texts. As a precaution against participants attempting to identify patterns, I set up the survey software to block the use of the 'back' button in Internet browser windows. This meant that participants could not return through the survey and change their answers. I informed participants in the introduction to the survey that any answer, once submitted, could not be changed.

The next six subsections provide an overview of each experiment type trialled in the pilot study. For each question, I discuss the factors which motivated me to use this particular form, and provide a critical review of the advantages and disadvantages of this particular method of investigation. The pilot findings and participant feedback are then critically reviewed, and a discussion given of the modifications I implemented to the experiments for the main study.

### 5.1.1) Pilot Question Type 1: Matching Coined Words to Images

The first test type I trialled in the pilot study used a closed-ended, multiple choice format. In this test, I asked respondents to choose one word from a series of options to name or describe a given meaning. As there are no established methods for testing the perceptibility of phonaesthemes, my primary motivation for choosing this style of experiment was that it seemed to provide an effective context for the perception of phonaesthemes in Abelin's (1999) Swedish study (cf. Abelin's 'Experiment 1b', Chapter 3.2.2.3). In each question of this type, I provided respondents with a single, large image[12], depicting, as unambiguously as possible, the meaning associated with the given onset under test (cf. Chapter 1; see discussion of possible image ambiguity in §§ 4.3.1.7 and 6.2.3). In this test, as in Question Type 2 (see below), I chose to portray the meanings associated with the phonaesthemes using images rather than linguistic contexts such as definitions or descriptions. There were several factors motivating this decision. The first was so that I could be absolutely certain I had not primed respondents to think about any of the phonaesthemes by unintentionally or unavoidably using the prosodies when presenting the twelve meanings. For example, had I tested the /stɪ/-<<straightness>> phonaestheme by asking respondents to choose a word for something "straight", I could not have known whether my informants were making analogies to the /stɪ/-<<straightness>> pattern in English (i.e. recognising the phonaesthemic pattern); or had simply been primed into choosing an /stɪ/ word by the word "straight" in the definition.

My decision to present the meanings through images was also motivated by my engagement with the literature on survey design. Many scholars (e.g. Ben-Nun, 2008; Dörnyei and Taguchi, 2010:9; Stoop, Billiet, Kock and Fitzgerald, 2010:24; Rovai, Baker and Ponton, 2014:502) warn that lengthy surveys, or those where respondents are required to process a large amount of material, can be susceptible to respondent fatigue; a 'phenomenon that occurs when survey participants become tired of the survey…and the data they provide begins to deteriorate' (Ben-Nun, 2008). Bethlehem and Biffignandi (2012:223-225) warn that the consequences of survey fatigue can include 'satisficing' (cf. Simon, 1957; Krosnick, 1991); a situation whereby respondents may begin choosing answers at random, or devoting markedly less care and comprehension to the answers they provide, simply to complete the

[12] All images used throughout both the pilot and main studies were linked from the photo-hosting site Flickr, and were either the intellectual property of the author or were legally used under a Creative Commons License.

survey as quickly and easily as possible. Issues of possible respondent fatigue and its consequences were particularly relevant to my study – both in its pilot and main phases – since my aim was to test speakers' perception of the phonaesthemes using a variety of experiments. As such, I was aware from the outset that the study would be lengthy, and that respondents would be required to process a large amount of material in both survey(s). Indeed, even if I had used just two or three different question types, testing all twelve phonaesthemes would have amounted to twenty-four or thirty-six questions in all. To this end, I recognised the need to keep the survey questions as varied and stimulating as possible for respondents, in order to retain their attention and willingness to cooperate for as long as possible (cf. Couper and Bosnjack, 2010:543). Presenting the phonaesthemes' meanings through images, rather than definitions or descriptions, provided such stimulation in this experiment, through the visual mode.

While choosing to present the phonaesthemes' meanings through images offered a potential reduction to any survey fatigue, it is notable that this decision was not without its disadvantages. Chiefly, these centred around issues of ambiguity. The advantage with using linguistic contexts (such as definitions or descriptions) to present the phonaesthemes' meanings would have been their specificity. I could have deliberately constructed these linguistic contexts so as to be confident that respondents were all reacting to the same semantic fields that I had intended. However, using images did not offer this guarantee. Even though I strove to use images in which my intended meanings were as large, central, bright, saturated and prominent as possible, so as to focus respondents' gazes and maximise the semantics I intended to communicate (Machin, 2007:66, 70-75, 132-138, 147-149); I had no guarantee that respondents **would** look at or react to this particular part of the image, or interpret its meaning in the same way as I had, even if they did. After all, any image has the potential to be interpreted in different ways by different people (cf. Mansfield, 2004:1087-1088; Gee and Hayes, 2011:113-114; ). There are any number of reasons why such differences in interpretation can occur; including cultural and personal experiences, and, in the context of a survey, any other images that participants have encountered during the survey itself (Tourangeau, Conrad and Couper, 2013:7). In practice, this potential disadvantage of using images became clear when I analysed the pilot study findings, and reflected on the feedback provided by the participants (cf. § 5.1.7, below). One participant actually noted that they "[weren't] sure about what some of the…images were showing". As a result of the feedback I received on this and my other image-based experiment (test type 2,

discussed below), I made a number of changes to this style of question for the main study. Chief among these was to add an instruction in each question for participants to study and respond **to a specified part of the image**. This resolved the issue of whether participants would look at and react to the same part of the image I had intended (i.e. the part that I felt portrayed the meaning associated with the phonaestheme under test). However, even with this adaptation, I could not be certain whether participants would interpret this part of the image in the same way I had intended. This meant that I had to be more cautious about claiming in my analysis that any result suggested the perception of a phonaestheme (or not) in this any my other image-based experiment (test 2). I discuss this point further in § 5.1.7, below. Nevertheless, I decided to retain image-based questions in the main study in spite of their potential drawbacks. This was primarily because of the advantages they offered vis-à-vis stimulation and minimising respondent fatigue, as discussed above. I reasoned that because my study did not rely solely on image-based questions, but also used experiments which provided the twelve meaning associations by way of linguistic contexts (see below), any inconsistencies among the image-based tests would likely be identified, and any uncertainties resolved, when their findings were viewed alongside those of the other experiments I had conducted. Indeed, as I discuss further in § 5.3, below, it is equally revealing to consider the findings of the experiments cumulatively as it is individually. In practice, the two image-based tests used in the main study did not appear to generate substantially different findings from the other experiments, which presented the phonaesthemes' meaning associations through linguistic contexts (i.e. definitions and descriptions) (see further discussion in § 5.3).

As well as the single large image, in each question of this experiment, I also provided respondents with audio recordings of three coined words. These words were constructed and delivered as follows:

- All three coined words were monosyllabic. This ensured that each coined word only presented respondents with one syllable-onset prosody.
- All three words were phonologically identical except for their onsets (i.e. they featured identical syllable rimes). Only one of the words contained the onset phonaestheme associated with the meaning portrayed in the image; the other two featured different, randomly-chosen onsets.

- All three words in every question used phonotactically-permissible onsets in English and Polish, with the exception of the question testing onset /tw/ which, as previously noted, does not occur in word-initial position in Polish.
- English respondents heard a native English speaker (the present author) pronouncing the coined words, while Polish respondents heard a native Polish speaker (the translator of the study) pronouncing the words. This was with the aim of ensuring, as far as possible, that the speaker accent did not cause difficulty for the respondents in interpreting the phonological forms of the words (however, I ultimately changed this method in the main study, as is discussed in § 5.1.8).

I invited participants to listen to the audio recordings of each coined word as many times as they wished; then asked them to tick a radio button next to whichever word they felt best named or described the content in the image (i.e. the meaning associated with the phonaestheme). Here, as in question types 2 and 4 (as well as questions 1, 2 and 3 in the main study), some questions asked respondents for a noun-like word – one to 'name' the content in the image – while others asked for an adjective-like word – one to 'describe' the content in the image. On reflection, it might have been more practical to have asked respondents to choose words to 'describe' the image content in every question; or words to 'name' the image content in every question. This is because there is no guarantee that asking for a noun-like word would have generated the same frequency or strength of phonaesthemic associations among speakers as asking for an adjective-like word. However, I varied the questions between asking for words to 'name' and 'describe' to try to minimise the chance of respondents inferring the associations I was testing, as this could have generated answers biased towards acquiescence. I reasoned that since connotations can attach to any content-bearing lexical item (cf. § 2.1), the grammatical class of word requested would hopefully not affect the strength or frequency of onset-meaning associations perceived.

For the purpose of clarification, a screenshot is included of this question type, taken from the English pilot study (Figure 7, overleaf). Screenshots of every question in the pilot study are included as Appendix 17.

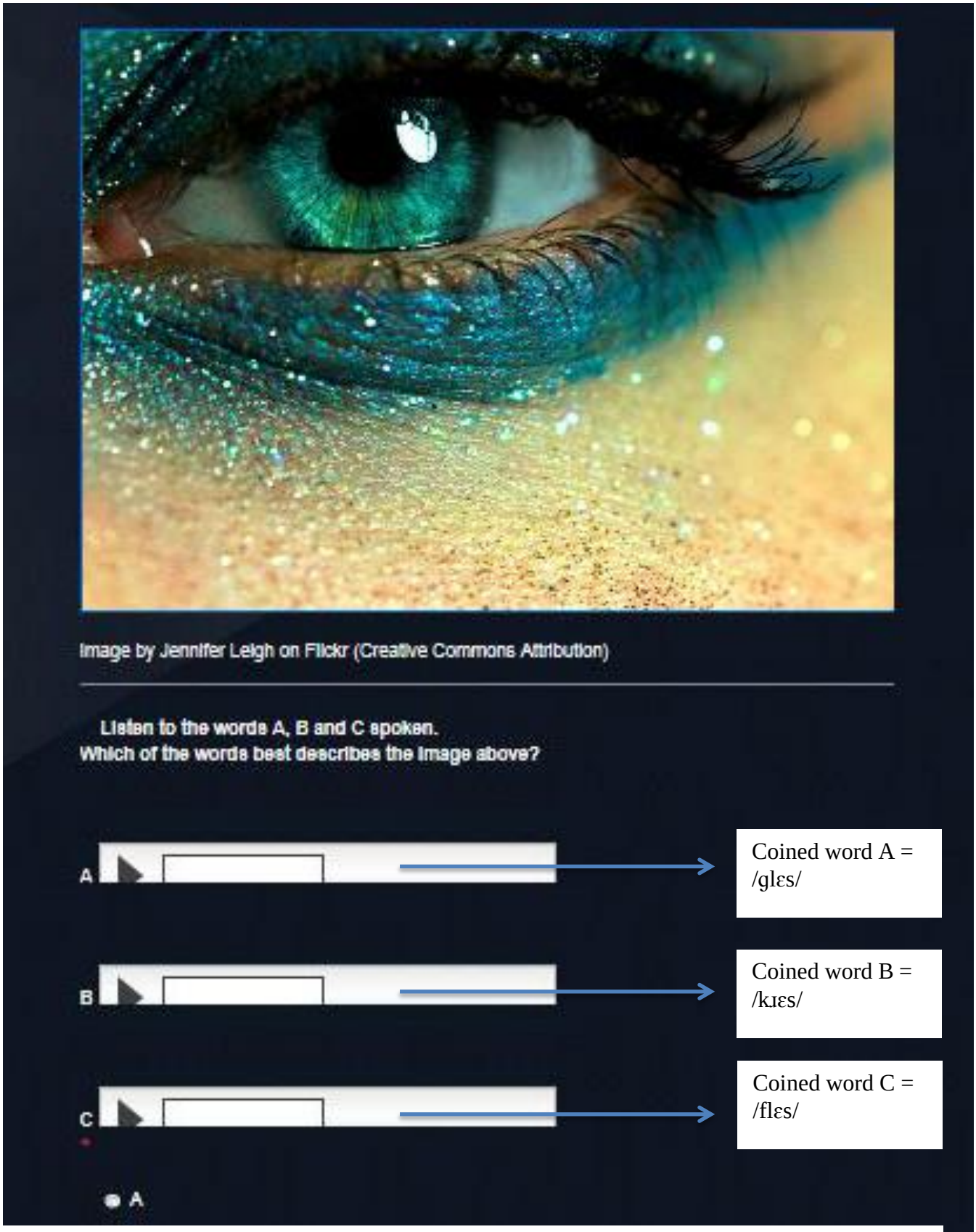



Figure 7: Screenshot of the pilot study question type 1 testing the /gl/-<<light>> phonaestheme. (If phonaestheme were used productively, word A would be chosen.)

### 5.1.2) Pilot Question Type 2: Matching Images to Coined Words

In this experiment, I used a similar closed-ended, multiple choice format as in test type 1. However, whereas test type 1 had asked participants to choose a word for a given meaning, this test operated in reverse, asking participants to choose a meaning for a given word. There is no motivation in the literature to suggest that phonaesthemic patterns would be perceived any more or less frequently depending on the "directionality" of the connotation (i.e. whether speakers experienced a sound and associated it with a particular meaning, or whether they experienced a particular meaning and associated with a particular sound). This is not to say that this phenomenon might not be worth exploring in the future; but it was not the reason for my piloting this experiment, or for taking it forward into the main study. My main reason for reversing the order of stimuli from test 1 was in order to provide another potential opportunity for participants to use the phonaesthemes productively, given that this type of experiment appeared to provide an effective context for the perception of phonaesthemes in Abelin's (1999) study (cf. Chapter 3.2.2.3).

There were a number of factors which motivated the specific form taken by this experiment. As noted above, I decided to use a multiple choice test for the same reason as in the first experiment; because this format seemed to have provided an effective context for the perception of phonaesthemes in Abelin's (1999) Swedish study (cf. Abelin's 'Experiment 2b', Chapter 3.2.2.3). However, this decision was also influenced by my engagement with the survey literature. Many scholars (e.g. Kothari, 2004:103; Dörnyei and Taguchi, 2010:33) suggest that multiple choice tests are advantageous – particularly in longer surveys – because they present a relatively low cognitive load for participants, they are comparatively quick to answer, and because they use a method with which many respondents are likely to be familiar. As such, I reasoned that re-using this methodology could also help to reduce the possibility of any respondent fatigue (see above). It was for this reason that I also retained the use of images in this question. While the same potential drawbacks to using images apply as in test type 1 (see above), I reasoned that the potential disadvantages of respondent fatigue could have been equally problematic; and therefore decided to retain the use of images for visual stimulation. However, as in experiment 1, I strove to use images in which my intended meaning (i.e. that associated with the phonaestheme under test) was as large, central, bright, saturated and prominent as possible in the frame, so as to try and maximise its salience (cf.

Machin, 2007:66, 70-75, 132-138, 147-149).

In each question of this test type, I provided respondents with an audio recording of a single coined word, which featured the phonaestheme under test, and a selection of three images. Only one of the images depicted the meaning associated with the phonaestheme of the coined word (cf. Chapter 1). In this experiment, I took great care to choose images whose colours were as bright and saturated as possible. This is because some scholars (e.g. Charoenruk and Stange, 2014) suggest that in web-based surveys where multiple images are used – particularly simultaneously – participants may tend to study the first image in greater detail than the remaining images. Choosing bright, highly saturated images offered some precaution against this possibility, since these qualities are known to attract viewers' attention (e.g. Machin, 2007:135).

In each question, I asked participants to listen to the audio recording of the [single] coined word as many times as they wished, then select which of the three images they felt was best named or described by the word. As in test type 1, I tested all twelve phonaesthemes using this method. Notably, in both this experiment and test type 1 (above), I did not offer respondents a 'none of the above' option among the multiple choice options; just a forced choice between the three coined words (or three images in test type 1). Several scholars (e.g. Converse, 1964; Ehrlich, 1964; Osterlind, 1998:151; Dörnyei and Taguchi, 2010:33) advocate the careful use of a 'none of the above' option in multiple-choice tests, on the basis that it can be helpful to participants in the event that they do not know or cannot choose a suitable answer from the options provided. However, many scholars in the literature (e.g. McClendon and Alwin, 1993; Krosnick, 1999a) argue that a noncommittal 'none of the above' option is in fact surplus to requirements, and that omitting this option, and forcing respondents to choose from the options provided, does not actually diminish the reliability or validity of the data gathered in any way. Indeed, Krosnick (1999a:44) notes, 'when people are pushed to offer an opinion instead of saying "don't know", the reliability and validity of data collected is no lower than when a "no opinion" option is offered and people are encouraged to select it […] That is, people who would have selected this option if offered nonetheless report meaningful opinions when it is not offered'.

As before, a screenshot is included of a question from this test type, taken from the English pilot study (Figure 8, overleaf):

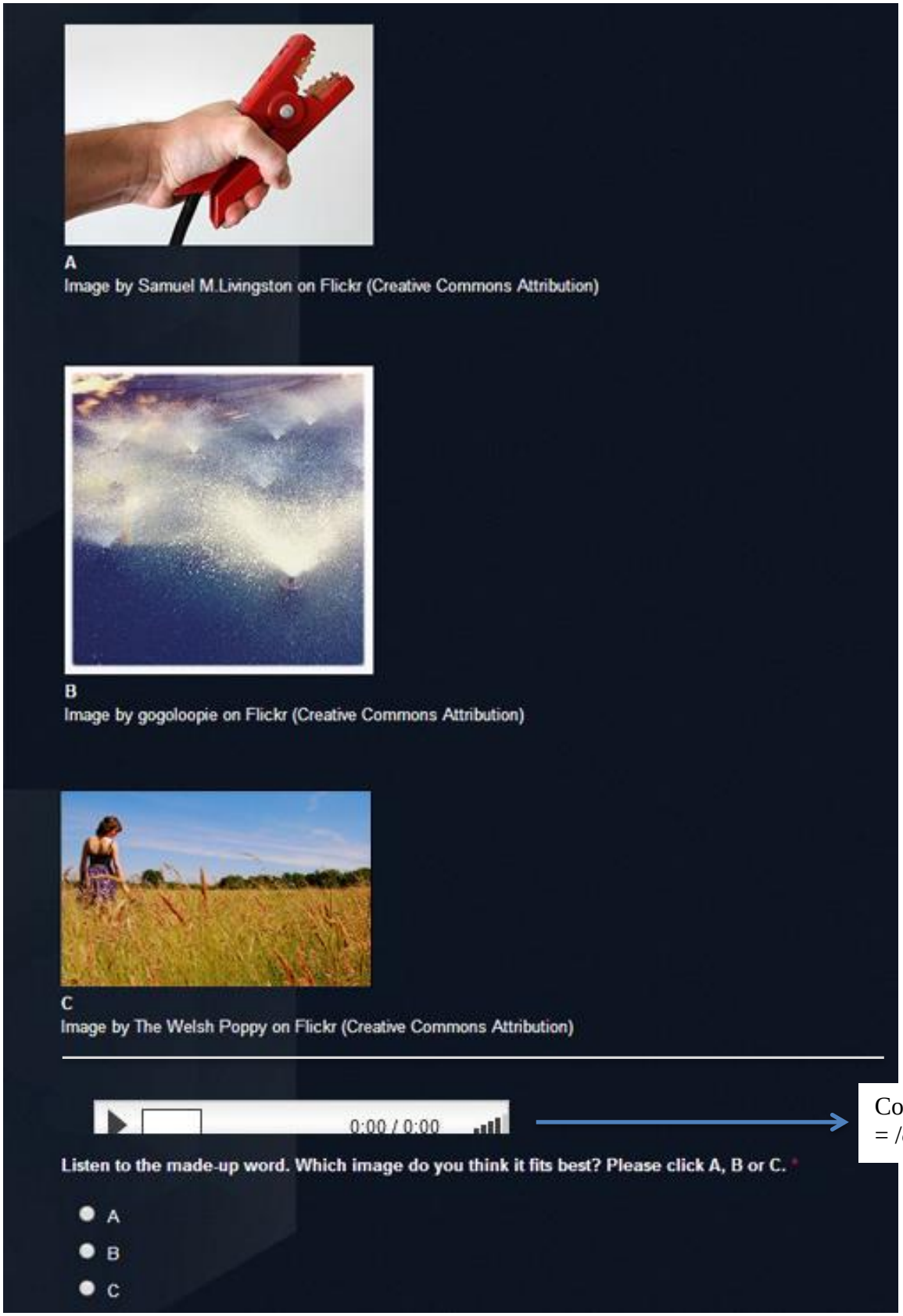



Figure 8: Screenshot of the pilot study question type 2 testing the /gɹ/-<<gripping>> phonaestheme. (Image A represents <<gripping>>. Thus, if phonaestheme were used productively, image A would be chosen.)

### 5.1.3) Pilot Question Type 3: Word-Building

My third style of experiment trialled in the pilot study used a completely novel approach to studying phonaesthemes. This method was entirely of my own design. In this question type, I asked respondents to 'build' words using supplied prosodies, and then attribute these coined words to a series of images. I implemented four questions of this type; each testing three of the twelve onset phonaesthemes. My motivation for piloting this experiment was to test whether, when given a **simultaneous** choice of several onsets **and** several meanings, participants would productively identify the twelve onset-meaning phonaesthemes under test. In test types 1 and 2, I had always completely controlled one of the variables; the "meaning" (image) in test 1 and the onset (coined word) in experiment 2. However, in this third experiment, I offered respondents choices in each variable simultaneously; a choice of onsets and a choice of meanings. I reasoned that this methodology offered a key advantage over tests 1 and 2, as it provided a more challenging context for the perception of the phonaesthemes (i.e. one which gave greater opportunities for respondents to associate the twelve onsets with different meanings and the twelve meanings with different onsets; and one in which it was less likely that the patterns under test would be identified by chance). As such, I reasoned that this test could have provided more robust evidence for the perception of the phonaesthemes, had this been identified; increasing the certainty with which I could make claims in my analysis.

The other key advantage that I saw to this methodology was that it required respondents to interact with the survey to a greater extent than tests 1 and 2. This interaction would necessarily take place on two levels: First, participants would have to construct words for themselves using the syllable prosodies I had provided. Second, having constructed these words, participants would have to decide which of the words was most appropriate for each image. I reasoned that this level of interaction would be a useful exercise to keep participants stimulated; and that the question could therefore serve as a useful "break" if placed between the multiple choice questions, so as to vary the style of questioning and help reduce the likelihood of respondents experiencing any fatigue (see further discussion of question ordering in § 5.1.7).

As noted above, in each of the four questions of this type, I provided respondents with three onsets. Each onset was one of the twelve phonaesthemes under study. I also provided

three randomly-chosen syllable rimes. All syllable onsets and rimes were phonotactically legal in English and Polish (with the exception of onset /tw/ in Polish as discussed above). I asked respondents to combine the onset and rime prosodies in any way they chose, such that they 'built' three words. My only stipulation was that they should try to avoid using each prosody more than once. Here, for the first and only time during the study, I provided the syllable prosodies in graphological rather than spoken form. This was because I considered that it would be easier for participants to see the prosodies in written form, rather than trying to imagine how individual audio clips of phonotactic sequences would sound in combination. In addition, it participants did not have to try to spell-out phonetic structures using graphemes, which could easily have led to misspellings or misinterpretations of responses by the researcher. However, despite these clear motivations, this was arguably not the best decision to have taken in hindsight. Having been well aware of the potential issues with representing the prosodies graphologically (rather than phonetically) from my critique of Abelin's (1999) study (cf. § 3.2.2.4), it would probably have been better if I had continued using auditory stimuli, and redesigned the experiment. The reason for this is that I could not be sure when I was analysing the data that all respondents had interpreted the same phonotactic sequences from the graphological stimuli, or the phonotactic sequences I had intended. As such, even though this test generated a comparatively high level of inter-test congruences with the results of other experiments (cf. analysis of pilot study findings, § 5.1.7, below), I could not ultimately know whether this was due to respondents having perceived the phonaesthemic patterns, or may simply have come about by chance. As such, I could not accurately determine whether this experiment had provided a context which facilitated the perception of phonaesthemes or not. It was for this reason, as well as for reasons of web-browser timeouts (see § 5.1.7), that I did not retain this experiment for the main study; although it is notable that this test type may be suitable for exploring phonaesthemes further in the future, if the method of presenting the prosodies were to be revised.

In addition to the six prosodies provided in each question (three onset phonaesthemes and three randomly-chosen rimes), I provided respondents with three images. Each image depicted a meaning associated with one of the three onset phonaesthemes under test, insofar as this was possible (cf. critique of images in § 5.1.1, above). Having 'built' their three words using the onset and rime prosodies supplied, I asked respondents to decide which of the words they felt was most appropriate for each image, and label each image accordingly.

For the purpose of clarification, a screenshot is included below of one of the questions of this type, taken from the English pilot study:

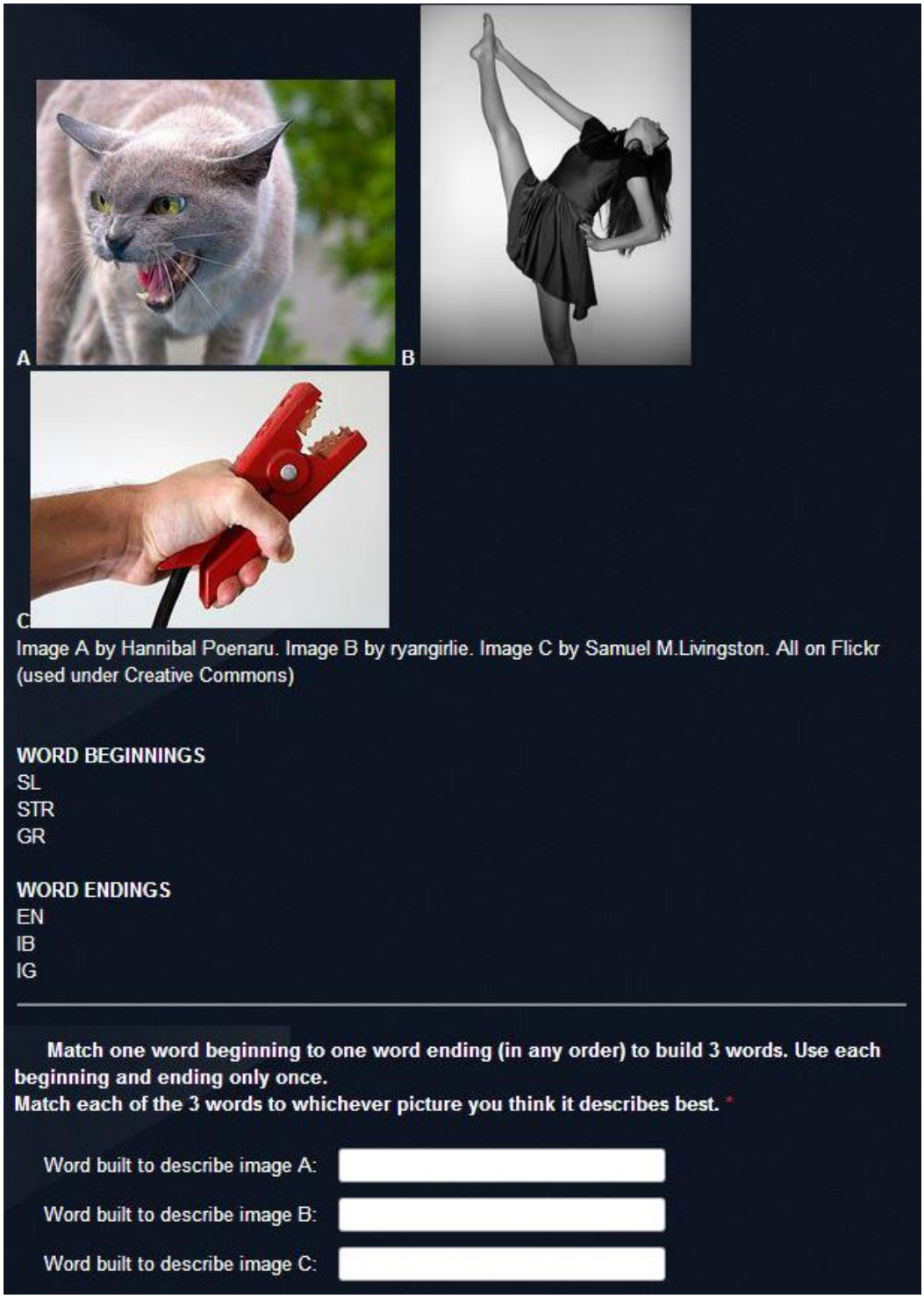



Figure 9: Screenshot of the first pilot study word-building question, testing the productivity of the /sl/ <<pejoration>>, /stɹ/ <<straightness/stretching>> and /gɹ/ <<gripping>> phonaesthemes. (Image (A) represents <<unpleasantness; pejoration>>; image (B) represents <<straightness>>; image (C) represents <<gripping>>.)

### 5.1.4) Pilot Question Type 4: Open-Ended Definitions

My fourth question type used a fully open-ended methodology. My idea in this experiment was to generate evidence about the productivity (perception) of the phonaesthemes when respondents were totally unprompted in any way. In each question of this test type, I asked respondents to invent any word of their choosing to name or describe a given definition. Each definition related to one of the meanings associated with the twelve onsets under test. For the /sl/, /sm/, /st/, /stɹ/, /kl/, /kɹ/, gl/ and /gɹ/ phonaesthemes, I asked for a word to describe the content in the definition. For the /sn/, /sw/, /skw/ and /tw/ phonaesthemes, I asked for a word to name the content in the definition (cf. consideration of 'name' and 'describe' in § 5.1.1, above). I imposed only one constraint in this test: I asked respondents to try to avoid knowingly using any existing word in any language they spoke.

There were several factors which motivated the form taken by this experiment, and several advantages to using this particular form. First, this open-ended style of experiment seemed to have provided an effective context for the perception of phonaesthemes in Abelin's (1999) Swedish study (cf. experiments '1a' and '2a', Chapter 3.2.2.2). However, I also reasoned that it was advantageous to include at least one experiment in my study which relied on linguistic contexts (such as definitions and descriptions) to test the productivity of the phonaesthemes, in addition to my image-based tests (cf. § 5.1.1, above). This is because this combination of methods gave greater rigour to my study, by offering multiple [different] contexts for the possible perception of these patterns. Naturally, each of these methods presented their own affordances and constraints. While my use of images in tests 1 and 2 had offered the advantages of visual stimulation and possible reduction of respondent fatigue, the major drawback to this decision was the potential for participants to misinterpret the meanings I had intended to convey (cf. critique in §§ 5.1.1 and 5.2.2, above). By contrast, the main advantage of using linguistic contexts, such as definitions and descriptions, was their degree of specificity. I was able to carefully construct these linguistic contexts so as to be confident that respondents were all reacting to the same semantic fields that I had intended (see screenshots in Appendix 17 for a full list of these definitions); something I could not guarantee in tests 1 and 2. Moreover, by combining the use of linguistic contexts with the use of fully open-ended questions, the resulting test offered substantial advantages in terms of certainty. Since open-ended questions allow participants total freedom of expression (Morrow, Jackson, Disch and Mood, 2011:182), and the ability to provide answers 'using

their own frame of reference, without undue influence from prespecified alternatives' (Edwards, Thomas, Rosenfeld and Booth-Kewley, 1997:26), this method provided the most robust test for the phonaesthemes of all the experiments in my study. With no pre-defined options from which to choose, respondents could have created any phonotactic sequence from any [phonotactically-legal] prosodies any language(s) they spoke; with potentially infinitely many different combinations. Even in English, there are at least eighty-one phonotactically-legal onset sequences from which participants could have chosen (Kreidler, 2004: Ch.6; Cruttenden, 2008: 254-255). By comparison, respondents were asked to select from just three different possibilities in tests 1 and 2, and from nine different possibilities in test 3. As such, this experiment offered the most challenging context for the perception of the phonaesthemes in any of my tests, and the greatest possible opportunity for respondents to associate the twelve meanings with onsets other than those under test. Consequently, it offered the greatest possible certainty that any pattern(s) in the findings might reflect the perception (or non-perception) of the phonaesthemes. For instance, if a particular onset was used by fifty percent of participants in response to a given definition, I could more confidently suggest in my analysis that this reflected some level of perception of a phonaestheme, and had not simply come about through chance or because of the limited number of response options in the question.

Although the design of this experiment offered the key advantages discussed above, it is notable that this test type was not without its disadvantages. One of the main drawbacks with using this methodology was the degree of caution I had to exercise in constructing each linguistic context, to ensure that I did not use the prosody under test anywhere in the definition/description or question. Indeed, this was one of the primary motivations behind my use of images in tests 1 and 2 (cf. § 5.1.1). The very rationale behind this test was to explore the productivity of the phonaesthemes **without** lexically priming respondents in any way. However, the fact that I could not use the prosodies in conveying their meaning associations often made it very difficult to construct the linguistic contexts in the first place. For instance, in question testing the /stɹ/-<<straightness>> phonaestheme, it was very difficult to convey the idea of <<straightness>> without using an /stɹ/ word[13]. The resulting contrivance required in constructing each definition meant that these questions took extensive time to design before I was satisfied that I had conveyed the meanings sufficiently precisely without using

[13] In practice, the definitions I chose to test this phonaestheme used the adjectives 'long' and 'thin' in the pilot study, and 'long' and 'level' in the main study respectively.

the prosodies under test. Another potential drawback with this methodology was the possibility of participants experiencing respondent fatigue. Despite their advantages in specificity, using linguistic contexts to convey the meanings associated with the prosodies provided less visual stimulation for participants than images. It is possible that the level of interaction required by this question – the fact that participants were asked to coin words completely unaided – may have alleviated the likelihood of fatigue somewhat. Nevertheless, I decided that it was probably best to minimise the use of any other definition or description-based questions in the study, in order to reduce where possible the likelihood of respondents becoming bored or fed-up.

### 5.1.5) Pilot Question Type 5: Word-Ranking

Like the third experiment, my final test in the pilot study used a completely novel methodology. My aim in this test was to measure the 'strength' with which speakers associated the twelve prosodies and meanings. Specifically, I wanted to explore whether speakers perceived coined words featuring the onsets as being 'better suited' to expressing their associated meanings than existing words which did not feature the onsets. For example, would speakers deem a coined word featuring onset /gl/ a more appropriate word to represent the meaning <<light>> than an existing word like 'sparkle' or 'shine'?

To generate data that could investigate this point, I used another closed-ended method. In each question of this test, I provided respondents with audio recordings of five words, labelled A-E. In the majority of questions, one of these words was a coined word, and featured one of the twelve prosodies under test. The remaining four words were existing English words that were, to varying extents, synonymous with the meaning association of the prosody in the coined word. For example, in the question testing /st/ and <<strength; stoicism>>, I used the coined word /stɔːt/ alongside the English words "resolute", "anchored", "braced" and "moored". (I varied this pattern on three occasions, however, as is discussed below). I invited respondents to listen to the recordings of the five words as many times as they wished, then rank the words based on the extent to which they sounded like they would express 'X' (where X was the meaning associated with the onset(s) under test). For example, returning to the question testing the /st/-<<strength; stoicism>> phonaestheme, I asked respondents to rank the words "stort" (/stɔːt/), "resolute", "anchored", "braced" and "moored" in terms of how much they thought each one sounded like it would express

something ‘firm or immovable’. The ‘firmest’ or ‘most immovable’ sounding word would be placed at the top of the scale, while the ‘least firm’ sounding word would take lowest place. I reasoned that if I were to take this experiment forward into the main study, I would then calculate and compare the most common (mode) rank position attained by each of the words. This would reveal the ‘strength’ of the associations perceived between each of the words and the meaning in question. For example, in the question testing the /st/-<<strength and stoicism>> phonaestheme, the words “resolute” and “stort” (the coined word) might have attained a mode rank of 1, while the remaining words might have attained a mode rank of 4. This would suggest that speakers perceived the words “resolute” and “stort” to be most appropriate for expressing something <<strong and stoic>>; this despite the fact that “stort” was a coined word with no pre-existing definition, and that respondents had only the sound structure of this word on which to make their judgments about its possible meaning. In the event that such a pattern had been found, it may have suggested the possibility that phonaesthemic associations could, in some contexts, be an equally strong or perhaps even stronger form of meaning than denotation; since “stort” (/stɔːt/) would have been more strongly associated with <<strength and stoicism>> than three existing words with established denotations related to this meaning (namely anchored, braced and moored).

As noted above, there were several questions in this experiment type where I varied the ratio of one coined word to four existing words. These variations were as follows:

- I tested the /sl/, /sm/ and /sn/-<<pejoration>> phonaesthemes together in a single question, and therefore used three coined words and two existing <<pejorative>> English words;
- I tested the (theoretically-opposing) /kɹ/-<<crookedness>> and /stɹ/-<<straightness and stretching>> phonaesthemes together in a single question, and therefore used two coined words; combining these with existing English words expressing <<straightness>>[14]
- In the final question of this type, I included three coined words, although only one featured the phonaestheme under test (/gɹ/).

[14] Theoretically, were I to have been analysing the results of these experiments, I would have expected the /stɹ/ word to be more highly associated with <<straightness>> overall than the /kɹ/ word.

Each of these variations was motivated. The motivation for using three coined words in the question testing /gɹ/ was as a control; to investigate whether respondents really were making associations between the prosody of the coined word and the meaning under test, or whether they were simply choosing the coined words because they were not existing English words. The motivation behind the other variations was to test whether I could compare the 'strength' of several phonaesthemes simultaneously using this type of experiment, were I to take it forward into the main study.

There were a number of factors which motivated my decision to pilot this particular experiment. First, like test types 3 and 4, the questions in this test required a greater level of engagement from participants than simply clicking a box or a radio button. Having listened to the audio recordings of the five words, participants had to drag the options A-E into designated rows, depending on how much or how little they thought each word sounded like it would express the given meaning. In the uppermost row, I asked participants to place the word they thought sounded most like it would express the given meaning; in the lowermost row, the word they thought sounded least like it would express the given meaning. By introducing this level of interactivity into the test, my aim was to provide ongoing stimulation for participants, thereby reducing the effects of any possible fatigue. I reasoned that this interactivity was particularly advantageous in this experiment, since this would be the final test respondents would encounter in the pilot study. By the time participants reached this experiment, they would already have been exposed to forty-eight prior questions (four test types, each testing the twelve phonaesthemes). This was because there was only one piece of survey software available to me (at the time of data collection) which allowed for the integration of audio recordings into the particular types of questions I had designed. Unfortunately, this software did not allow for the randomisation of questions within an experiment type; only total randomisation of questions throughout the study. Given that each of my test types was unique, and warranted its own set of instructions to respondents, I was not able to take advantage of this facility. This meant that each participant encountered the same questions in the same order; and that this test was always encountered last. (However, when I designed the survey, I deliberately varied the order in which the twelve phonaesthemes were tested in each question type, as a precaution against participants trying to identify patterns in the questions and their stimuli, and providing answers biased by a desire to acquiesce (cf. § 5.1)). Consequently, I decided that any action I could take to reduce possible fatigue at this point would be beneficial. (I return to the issue of question ordering

and the amount of material respondents were required to process in § 5.1.7 when I discuss the modifications I made to the pilot methodology for the main study.)

In addition to its level of interactivity, my decision to pilot this experiment was also motivated by engagement with the survey literature. Ranking scales are cited as being particularly common data collection methods used in a wide variety of social research domains, including business and marketing (Bajpai, 2011:52), education (Burke Johnson and Christensen, 2014:161-162) and community development (Hutson and Kolbe, 2010:137). As such, I decided that a ranking scale would form a useful basis for this experiment, as it was likely to be a familiar method to many respondents; one which did not necessitate a high cognitive load and would encourage participants to remain attentive even if they were becoming tired (Glastonbury and MacKean, 1991:238; Lum, 2013:204). Moreover, Krosnick (1999b) suggests that simultaneous ranking scales such as those used in this test offer a number of advantages over other forms of rating scales, such as those where respondents rate stimuli individually on a particular dimension (cf. Taylor and Kinnear, 1971; McIntyre and Ryans, 1977; Reynolds and Jolly, 1980). Krosnick (1999b:556) suggests that participants 'are more likely to make mistakes' when responding to individual rating scales; including 'failing to [respond to] an item more often than when ranking' (cf. Neidell, 1972; Brady, 1990). As such, Krosnick (1999b:556) argues that ranking scales tend to be more reliable; a point cited by a number of scholars throughout the literature, including Elig and Frieze (1979), Munson and McIntyre (1979), Rankin and Grube (1980), Reynolds and Jolly (1980) and Meithe (1985).

Although the form of this experiment offered a number of advantages as discussed above, it is notable that there was one main drawback to this methodology. This concerned the difficulty in communicating its instructions clearly and unambiguously to participants. While the principle of a ranking scale offered a relatively straightforward process with a relatively low cognitive load, the way in which I asked participants to use this methodology was rather more complex. The idea of ranking each word "on the basis of how much it sounds like it would represent [X]" was a difficult idea to convey in simple and clear terms; particularly as 'X' (the meaning under test) varied in each question. It also made this question difficult and time consuming to translate into Polish. Indeed, even though I had presented the English instructions in what I felt was the clearest format (with short sentences and simple clauses where possible), my Polish translator requested clarification of this question type in

person. As a result, I decided that it would be useful to include an example of a completed [hypothetical] question of this type in the instructions, to provide further clarification for participants. This example was purely hypothetical, and did not use any of the phonaesthemes or meanings under test, so as to avoid priming respondents in any way. Rather, it showed a possible completed ranking of three random coined words on the dimension of sounding 'like they would describe a sea creature'. However, the effect of including this example with the already fairly complex directions was that the instructions for this test became the longest of all the experiments in my study. In hindsight, this was probably not the best decision to have made, given that this was the last test respondents were to encounter in the study, at a point when the effects of fatigue may have been setting-in (cf. Glastonbury and MacKean, 1991:238; Cole, 2014:288). In practice, it appears that my concerns about both the complexity of the test type and the length and detail of its instructions were very much founded: One pilot participant explicitly drew attention to this test as being 'confusing' in their feedback, as is discussed in § 5.1.7, below. As a result of these practical difficulties, I made a number of modifications to simplify this test for the main study (see further discussion in §§ 5.1.8 and 5.2.4 below).

### 5.1.6) Participant Feedback

After the final question, and prior to the submission of their completed surveys, I invited respondents to provide optional feedback. The main purpose of this was to try to establish whether respondents had understood what was being asked of them in each question, and to establish whether respondents had experienced any difficulties – either with the survey at-large or with any of the individual test types. To this end, I provided an expandable text box, so that participants could type as much or little as they chose. As is discussed below, participant feedback played a crucial role in identifying a number of shortcomings in the pilot study methodology, and in shaping the form taken by the main study.

### 5.1.7) Critical Evaluation of Pilot Study Findings and Participant Feedback

This section provides a critical evaluation of the main findings and participant feedback from the pilot study. Here I use the term 'findings' to mean 'what the pilot study reveals about the methodologies of the productive experiments', rather than to imply or refer to any data

pertaining to the phonaesthemes. As discussed above (§ 4.3.1), my purpose for implementing the pilot study was to test the different types of productive experiment and resolve any issues before conducting the main study. My main priorities were:

- To check that the methodologies of these experiments could be implemented within the constraints of the survey software;
- To check that the survey tool operated as it should have (in terms of displaying non-English graphemes and collecting data from other countries);
- To test whether my proposed experiments appeared to be reliable;
- To establish whether respondents understood what was being asked of them in each experiment; and to establish any inconsistencies or difficulties respondents may have encountered, either with the study at-large or with any of the specific test types.

With regards to the first of these points, the software offered me a 'trial run' of the survey before I circulated its weblink to informants and began collecting their responses. From this trial run, I was able to verify that the survey tool was successful in displaying the embedded images and playing the embedded audio clips; and that the five test types were all correctly formatted with the appropriate images, audio clips and response options in each question. The trial run also revealed that non-English graphemes were successfully displayed. Once I began collecting participants' responses, the survey software also demonstrated that it was successful in gathering data from other countries: All five Polish respondents successfully submitted their answers.

Once I had fully gathered the pilot study data, I needed to establish whether the experiments appeared to be reliable, insofar as I could establish. To accomplish this, I investigated the extent to which the five experiments tended to generate similar (and dissimilar) results. I reasoned that the test types which most consistently generated findings in-line with the dominant trends would be the most reliable; while any that tended to generate dissimilar findings would be less reliable. The format and findings of this analysis are discussed below. For the purpose of this exercise, I used only the English data. This is because the type frequency data (cf. § 4.2) indicated that a number of the phonaesthemes were not evident in the vocabulary of Polish. Given that the type frequency of a phonaestheme could potentially determine its perceptibility (cf. §§ 1.2, 5.4), I reasoned that

these associations may not be perceptible patterns in Polish. If this was the case, then in a number of the experiments, Polish respondents may have chosen answers at random. This could well have resulted in dissimilarities between the results of these experiments. These dissimilarities would not necessarily indicate anything about the (un-)reliability of the experiments; only about the perceptibility of the phonaesthemes in Polish. For this reason, I focused exclusively on the English data. As all twelve phonaesthemes are evident to some extent in the vocabulary of English (cf. § 4.2), I reasoned that they were potentially more likely to be perceptible patterns in English. By extension, this would mean that any similarities or dissimilarities between the English results were more likely to indicate the reliability or unreliability of the experiments, rather than simply speakers' perception of the phonaesthemes.

To establish the extent to which the five experiments appeared to generate similar (and dissimilar) results, I devised the following method. First, I took each phonaestheme in turn, examined its data and noted:

- The details of every experiment in which the phonaestheme was more productive than any other sound-meaning association;
- The details of every experiment in which the phonaestheme was not more productive than another sound-meaning association (i.e. where it was equally productive as another sound-meaning association or where a different sound-meaning association was more productive than the phonaestheme).

For test type 1, this involved counting and comparing the number of times each coined word was chosen. For test type 2, this involved counting and comparing the number of times each image was attributed to the coined word. For test type 3, this involved counting and comparing the number of times each word onset was attributed to each image. For test type 4, this involved counting and comparing the number of times each distinct word onset was used. Finally, for test type 5, this involved calculating the mean rank position attained by the five words, and comparing the mean rank of the phonaestheme(s) to those of the non-phonaesthemic words.

Having noted the above, I then determined the main trend in the results for each

phonaestheme. This meant establishing whether the majority of the five tests showed the phonaestheme to be more productive than any other sound-meaning association; or whether the majority of the tests showed it to be no more productive than another sound-meaning association. I then produced a table summarising the main trend of each phonaestheme and the experiments in which this trend was shown. This data is displayed in Table 17, overleaf. This provided an overall picture of the tests that were congruent in showing the main trend for each phonaestheme. It also revealed which tests generated incongruent (i.e. dissimilar) results in each case.

Table 17: The pilot study questions producing congruent results for each phonaestheme

| **Phonaestheme** | **Description of main trend in results** | **Tests whose results were congruent in showing main trend** |
|---|---|---|
| /sl/ | Phonaestheme is no more productive than another association | 3, 4, 5 |
| /sm/ | Phonaestheme is no more productive than another association | 1, 3, 4, 5 |
| /sn/ | Phonaestheme is more productive than any other association | 2, 3, 5 |
| /stɹ/ | Phonaestheme is no more productive than another association | 1, 2, 4, 5 |
| /kɹ/ | Phonaestheme is no more productive than another association | 3, 4, 5 |
| /gl/ | Phonaestheme is more productive than any other association | 1, 2, 3, 5 |
| /sw/ | Phonaestheme is more productive than any other association | 1, 2, 3, 4 |
| /gɹ/ | Phonaestheme is no more productive than another association | 3, 4, 5 |
| /tw/ | Phonaestheme is more productive than any other association | 1, 2, 3, 5 |
| /skw/ | Phonaestheme is more productive than any other association | 1, 2, 3, 5 |
| /kl/ | Phonaestheme is no more productive than another association | 1, 4, 5 |
| /st/ | Phonaestheme is no more productive than another association | 1, 2, 5 |

Key to highlighting:

Phonaestheme is equally or less productive than another association

Phonaestheme is more productive than any other association

The table shows that all five tests seemed to generate a greater number of consistent findings than inconsistent findings. At least seven of the twelve results from each test type were consistent with the main trends identified. This suggested that the five test types appeared to be relatively reliable; and were potentially suitable for retention in the final study. Of the five experiments, test type 5 generated the greatest number of congruent findings; eleven of its twelve results were consistent with the main trends of the phonaesthemes. This was followed by test type 3 (nine congruent results); test type 1 (eight congruent results), and test types 2 and 4 (seven congruent results apiece).

However, while the experiments appeared to be relatively reliable, the feedback provided by respondents indicated that a number of individuals had experienced difficulty in understanding some questions. I received four feedback comments from participants overall (three in English and one in Polish); of which two cited issues of understanding:

*"q5 a bit confusing but i sorta [sic.] got what i had to do" (English Respondent 2);*

*"wasn't sure about what some of the early images were showing - like there was one of a road, did you mean the road itself or the stuff around it?"(English Respondent 4).*

In the first of the comments above, English respondent 2 indicated that they found question type 5 confusing. So although the methodology of this test type seemed to be fairly reliable, there appeared to have been something about this experiment that respondents found unclear. It may have been that the confusion experienced by this respondent was caused by the nature of the activity I had asked participants to complete; or the way in which the instructions for this activity were presented. Unfortunately, the respondent did not indicate the precise cause of their confusion. Nevertheless, this feedback provided a direct manifestation of my concerns about the methodology and instructions of this experiment (cf. § 5.1.5, above); and suggested that the design of this test would need to be modified to avoid confusion if it were to be implemented in the main study. As a result of this feedback, I simplified this particular experiment, changing the simultaneous ranking scale design to a Semantic Differential scale (cf. §§ 5.1.8 and 5.2.4, below).

In addition to the potential difficulty with test type five, the feedback provided by

English respondent 4 suggested that a number of the image-based questions may have been problematic; in that participants might have been unsure about which parts or qualities of the images they were being asked to name or describe. To exemplify, this participant cited the image of the straight Indian road I had used to test the /stɹ/-<<straightness and stretching>> phonaestheme in test 1 (cf. p.4 and p.16 of Appendix 17). They reported that they had been unsure about whether I was asking respondents to choose a word to describe the road itself, or a word to describe some other part of the image. Again, this feedback provided a direct manifestation of my concerns about the methodology of this experiment; and of the disadvantages of using images in tests 1 and 2, rather than linguistic contexts, to convey the meanings associated with the phonaesthemes (cf. § 5.1.1). Although I had tried to make the relevant feature of each image as salient as possible through size, centrality, brightness, saturation and framing (Machin, 2007:66, 70-75, 132-138, 147-149), it is clear from this feedback that greater specificity was needed in the instructions of these image-based questions, to ensure that all participants studied and responded to the same part of the image I had intended. In response to this, I ensured that every 'type 1' question in the main study (i.e. every question which asked respondents to choose one of three coined words to name or describe a given image) indicated precisely which part of the image respondents were being asked to name or describe (see further discussion below). In theory, it would also have been beneficial to adopt this practice in the 'type 2' questions; those which asked respondents to choose one of three images to attribute to a given coined word. However, in practice, this might have been less successful. These questions would have had to direct respondents to look at specified parts of **three** images, and to attribute the coined word to one of these specified parts. This would have made the instructions for this test type more difficult to explain; and I had already determined from test type 5 that the complexity of tests and their explanations might have been a source of confusion for participants. For this reason, I retained the original instructions used in the pilot study for the 'type 2' questions; asking respondents to choose which of the images they thought the word '[fitted] best'. However, I reviewed each of the images used in this experiment to try to ensure that the meanings I intended to convey were the most salient features in their respective images; and were as large, central, bright and saturated as possible (ibid.)

Of course, it is notable that, even after making these changes to the 'type 1' experiments, and directing participants to study and respond to specified parts of each image, there was still no guarantee that respondents would interpret the images in the same way, or

indeed the way I had intended (cf. § 5.1.1, above). This was particularly true for the 'type 2' experiments – those where participants had to choose one of three images for the single coined word – since I could not even direct respondents to specific parts of these images (see above). For this reason, I still had to exercise caution in making any claims about the phonaesthemes when analysing the findings of my image-based experiments; I knew I would not be able to claim that any given result offered 'proof' about the (non-)productivity of any phonaestheme. However, as I previously noted in my critique of Abelin's (1999) study (cf. § 3.2), and as I discuss further in § 5.2.1, it would have been inappropriate to make such claims in the present study anyway; not least because of the size of my respondent groups and the absence of existing studies in this area. In an exploratory study such as this, I knew that the analysis of data would need to focus on numerical tendencies and trends, and that any remarks about what these patterns appeared to show would need to be made appropriately tentatively. However, in hindsight, it might have been helpful to conduct one or more focus groups at this stage, to establish the suitability of the images I had chosen to represent each meaning association. During these focus groups, I could have asked participants to discuss how each meaning association would be best represented through an image or images, and why; that is, whether there were any features they felt the various images would need to include. For example, would the idea of <<pejoration>> be most clearly represented by an image depicting an act of violence, or by an image depicting an object capable of causing physical harm, such as a gun? Or some other content? Would the idea of <<straightness>> be most clearly represented through an image depicting some kind of straight object, or an image depicting some kind of action? Or something else? I could then have shown participants the images I intended to use, and asked them to discuss whether they felt these images were suitable to represent the intended meanings, or whether it would be advisable to choose alternative images. Focus groups would have been an ideal method for this stage of the investigation, for a number of reasons. They would have offered a relatively quick means of sampling multiple people; they would have allowed me to clarify any necessary points with respondents quickly and simply; they would have generated a large amount of rich and useful qualitative data vis-à-vis the suitability of images (Stewart and Shamdasani, 2015:45-46), and, most importantly, they would have allowed respondents the discursive freedom 'to react to and build on the responses of other group members' (ibid.), which could have generated ideas and suggestions that would not otherwise have been revealed by one-to-one sampling. Naturally, the logistics of this exercise would mean that it would only have been feasible to conduct focus groups with English speakers. Moreover, these focus groups would

necessarily have had to sample different participants from those I would later survey in the main study. Nevertheless, this approach would have provided a useful insight for my image-based tests, and would have afforded me greater certainty that participants were likely to have interpreted the meanings I had intended to convey in any images I used. If I were to conduct similar experimental research into phonaesthemes in the future, this is an approach I would ideally employ in my study. I return to this point in my discussion of the limitations of this study in Chapter 6, and in my discussion of the directions for future research in Chapter 7.

Whereas the first two feedback comments from the pilot study had raised issues of understanding specific questions, the remaining comments both raised a different issue; specifically, the time taken to complete the survey:

*"Survey timed out twice before I could get to the end! This was my third attempt. Guess you have to be quite quick with your answers" (English Respondent 1)*

*"questionnaire slightly long but still interesting" (trans.) (Polish Respondent 4)*

The fact that both responses raised time-related issues with the survey is important. Firstly, as shown above, Polish respondent 4 reported that they found the survey 'slightly long'. As discussed in the methodology of my experiments (cf. sections 5.1.1-5.1.5, above) I was well aware of the potential issues concerning 'respondent fatigue' when designing my survey. I knew that it would have been counterproductive to implement a study with so many experiments that respondents became fed-up or bored whilst completing it (Ben-Nun, 2008; Tavakoli, 2012:551). This is because feelings of boredom and tiredness could have led to respondents choosing answers at random, or devoting markedly less care and comprehension to the answers they provided, in order to finish the survey as quickly as possible (cf. discussion of 'satisficing' by Simon, 1957; Krosnick, 1991, Bethlehem and Biffignandi, 2012:223-225 and in § 5.1.1). This, of course, would have defeated the whole point of the experiments. Therefore, the fact that this participant specifically commented on the length of the survey suggested it was worth investigating whether the number of questions could be reduced in the main study, so as to minimise the effects of any possible fatigue. In practice, one way in which I tried to achieve this was by removing one of the tests (see discussion below).

It was not just the number of questions in the survey that caused time-related problems for respondents. As shown above, English respondent 1 reported issues with browser 'time outs'. 'Timing out' is the termination of an Internet connection or process, which usually occurs when a web server fails to download a webpage or does not receive a command from a user's browser within a certain length of time (OED: Online). It is possible that time-outs reported by English respondent 1 were caused by the quantity of embedded images and audio files in the survey. With every question featuring either three embedded images or at least three audio clips, the web browser and bandwidth available to many participants would have been placed under a heavy load throughout the survey. This could have increased the time taken for each question to download, and could have resulted in participants having less time to read each question and submit their answers before their web browser terminated the process. While clearly a source of frustration for this respondent, there could potentially have been more serious implications of browser time-outs to the main study. The first is that not all respondents might have restarted the survey if their browser timed-out, particularly if the length of the survey had already been an issue for participants (as suggested by Polish respondent 4). This could have led to many respondents starting but not finishing the survey, resulting in an incomplete data set. However, even if respondents were to have restarted the survey following a time-out, there is a risk that they could have begun to recognise the co-occurrences of onsets and meanings in the questions, or begun to speculate about the linguistic patterns being studied. This could then have skewed their answers in some way. For example, if the respondents had believed they had identified the patterns under investigation, they might have provided answers biased by a desire to acquiesce (cf. 'acquiescence bias' in § 5.1, above). The most obvious solutions to the problem of server time-outs were to reduce the data load for participants' web browsers and to shorten the survey; the latter already a priority in view of the feedback from Polish respondent 4. This is given further discussion below.

### 5.1.8) Methodological Changes Implemented for Main Study

In light of the participant feedback, I made a number of modifications for the main study; both to individual experiments and to the study at large. These are now briefly discussed. Following this, the next section of the chapter presents the [revised] methodologies and findings of the main study.

The first modifications I made were designed to address the time-related issues discussed above. I began by slightly shortening the length of the survey, in response to the potential problem of respondents becoming bored or fed-up with the task. This I achieved by removing question type 3 (the 'word-building' exercise). I chose to remove this particular experiment for a number of reasons. The first was because of the methodological disadvantage I had identified vis-à-vis representing the prosodies graphologically rather than phonetically (cf. § 5.1.3). I decided to represent the prosodies graphologically in this experiment because I considered that it would be easier for participants to 'build' their words using written sequences rather than audio clips; and because it meant that participants did not have to try to spell-out phonetic structures using graphemes, which could easily have led to misspellings or misinterpretations of responses by the researcher. However, the consequence of this decision was that I could not be sure, when analysing the data, that all respondents had interpreted the same phonotactic sequences from the graphological stimuli, or even the phonotactic sequences I had intended (for further discussion of interpreting phonetic forms from graphological structures, see §§ 3.2.2.4 and 5.2.3.). Therefore, even though this experiment seemed to generate a number of congruent findings with the other tests in the pilot, I could not ultimately know whether this was due to respondents having perceived the phonaesthemic patterns, or whether this may simply have come about by chance. As such, I could not fully determine whether this experiment had provided a context which facilitated the perception of phonaesthemes or not; and whether the test type was suitable for retention in the main study or not. Consequently, I decided this test was probably the most sensible choice of experiment to remove, on balance. This decision was also aided by the fact that test type 3 had placed some of the heaviest data-loads on respondents' web-browsers; requiring three images to be retrieved and displayed simultaneously. I reasoned that by removing this question, I would not only shorten the survey and reduce any possible respondent fatigue, but also reduce the likelihood of browser time-outs, and maximise the chances of respondents successfully completing the study on their first attempt.

Having removed test type 3, I then re-ordered the remaining experiments in the study. This was to ensure that respondents were not exposed to more than twelve consecutive questions of the same format; another measure aimed at reducing respondent fatigue. Discussions of this phenomenon (e.g. Smith and Fletcher, 2004:168; Ben-Nun, 2008; Stoop, Billiet, Kock and Fitzgerald, 2010:24; Bethlehem and Biffignandi, 2012:223-225; Tavakoli, 2012:551) suggest that respondent fatigue is not just influenced by the methodology of the

individual questions or the length of the study itself; but also by factors such as the degree of repetition, and the number of similar questions asked consecutively without any variation or break. As Smith and Fletcher (2004:168) note, 'respondent 'fatigue' can set in when the same type of question is asked over and over again, with each question having exactly the same type of response options […] Respondents will cease to discriminate between [the different options or statements provided], and opt for the answer that will most speedily get them to the end of this chore' (cf. discussion of 'satisficing', § 5.1.1). To minimise the chances of this situation arising, I moved question type 2 (multiple choice of images for single coined word) from its original position following experiment 1 to a point nearer the end of the survey. This meant that participants would not encounter two consecutive sets of twelve questions which all involved the same kind of material (i.e. multiple choice questions with three possible answers). The resulting format for the main study was as follows:

1. Twelve multiple choice 'type 1' questions, where respondents would choose one of three coined words for a single image (one question per phonaestheme);
2. Twelve Semantic Differential scale questions (see below), where respondents would place a coined word on a scale to indicate the meaning they thought it would most likely express;
3. Twelve multiple choice 'type 2' questions, where respondents would choose one of three images for a single coined word;
4. Twelve open-ended questions, where respondents would invent any word of their choosing in response to a given definition.

As noted above (§ 5.1.5), the constraints of the survey software meant that I was not able to randomise the ordering of questions within each set. Rather, I would only have been able to randomise the order of all questions throughout the study. This would not have been an appropriate decision to have made, since each type of question necessitated its own set of instructions. It would have been impractical and inappropriate to repeat these instructions for each question of each test type; particularly in view of the additional time and quantity of reading this would demand from participants. Instead, it was most practical to run all twelve questions of each test type consecutively. This meant that I also had to decide on the order in which the phonaesthemes would be tested **within** each test type, and ensure that this order was kept constant across each language group to ensure accurate comparability. In so doing, I deliberately varied the order in which the phonaesthemes were tested within the four sets of questions. For instance, I ensured that the /sl/-<<pejoration>> phonaestheme was not always the first pattern to be tested; and that the /gl/-<<light>> phonaestheme was not always the last

pattern to be test. I reasoned that this variation within test types would hopefully reduce (where possible) the likelihood of participants noticing the patterns under test, and giving answers biased by a desire to acquiesce. In practice, the only potential disadvantage to my new question order was that it left the open-ended definition questions at the end of the study. This may have been problematic, since these questions placed what was arguably the heaviest cognitive load on participants of all the tests in my study. Here, respondents were required to coin twelve words from scratch, without any prompts or primes other than the stimulus definitions. Cole (2014:288) advises against placing such cognitively demanding questions at the end of a survey, since this is the point at which any effects of fatigue are most likely to affect the quality of responses. Fortunately, it does not appear that this disadvantage was manifested in practice. As I note when I discuss findings of this experiment (§ 5.2.3), the questions in this test seem to have provided some of the most convincing evidence for the perception of phonaesthemes in the entire study. Given the very low likelihood that 'satisficing' or randomly choosing answers would generate use of the phonaesthemes in this experiment (§§ 5.1.4; 5.2.3), it appears that participants did not experience substantial levels of fatigue at this point; and that the results of this test were still very much valid.

As well as removing question type 3, I implemented a further change to reduce the likelihood of Internet browser time-outs; specifically, I changed the size (resolution) of the embedded images. In the pilot study, I had linked the images directly from the hosting site Flickr in their largest file sizes for the purpose of quality. These often amounted to several megabytes of data per image. However, the Flickr server allows users to view most images at a number of different resolution settings. For the main study, I therefore embedded links to slightly smaller versions of the images; limiting their width to 500 pixels along their longest dimension. This gave a good balance between a smaller file size (which ensured faster download speeds and less browser/bandwidth load) and a clear resolution on a standard computer monitor.

Having shortened the survey and taken steps to reduce the likelihood of browser time-outs, I then implemented a number of modifications to address the potential ambiguities of image-based questions. The first of these modifications was to ensure that every 'type 1' question indicated precisely which part of the image respondents were being asked to name or describe. In addition, I replaced a number of images in question types 1 and 2 to ensure I was maximising the prominence of all twelve meanings under test. The first image to be

replaced was the Indian road, given that this image had been explicitly singled-out as ambiguous in the participant feedback. For its replacement in depicting <<straightness>>, I used a photograph of three horizontal telegraph wires bisecting the centre of the frame, perpendicular to a straight road running across the bottom quarter of the image. Out of necessity, I also had to find replacements for several other images I had used, because these had been removed from Flickr since I had collected the pilot study results. All replacement images are shown in Appendix 18, where the English text version of the main study is provided in the form of screenshots.

Further to the above, I also implemented a number of other modifications. The first of these was to simplify test type 5 (the 'word-ranking' test). As discussed above, the participant feedback suggested that the instructions and/or methodology of this question may have been difficult for respondents to understand. In response to this, I redesigned this question type in order to make use of a simpler and even more widely-used methodology; a Semantic Differential scale (Cook, Hepworth, Wall and Warr, 1981; Dumont, 2008). The revised form of this experiment is presented in § 5.2.4, below. I also added a further corpus-based analytical stage to the open-ended definitions experiment, to increase the confidence with which I could make claims about the productivity of phonaesthemes from its results. This change is more appropriately discussed with the presentation of the revised experiments used in the main study (see § 5.2.3 for further details). Finally, I re-recorded every single coined word used in the study, using a speaker with a Received Pronunciation (RP) accent. The main reason for this was to maximize interpretability of the sounds in the coined words for respondents; RP being, in principle, one of the more widely intelligible accents of English (Brown, 1991:33). I gave these re-recorded coined words to all three respondent cohorts in the main study, rather than using a native speaker of each language to pronounce the coined words as in the pilot. Although this could have been disadvantageous to a study assessing the attitudes of respondents towards the speaker (cf. Giles and Powesland, 1975; Giles and Bourhis, 1976; Cruttenden, 2008:77), the cultural values ascribed to an RP accent are of less importance to the present study. During this process of re-recording the coined words, I became aware that two of the forms used in the pilot study, /kɹɛs/ and /klɛf/, were both existing words in English (/klɛf/ is also an existing word in French). Therefore, I took the opportunity to change the these words for the main study; combining the /kɹ/ and /kl/ phonaesthemes with different syllable rimes in their respective questions, and taking care to ensure that the resulting combinations did not realise any other existing words in English,

French or Polish.

In addition to making the modifications outlined above, I decided to provide an incentive for respondents to take part in the main study. The respondents of the pilot study were either my personal friends or friends of my Polish academic colleague. While I was able to rely on the goodwill of these individuals to complete the pilot, it would have been impractical to expect ninety people to complete the main study simply out of kindness. Indeed, Dörnyei and Taguchi (2010:72) suggest that offering an incentive is one of the primary ways of maximising participation in online surveys. As such, I offered respondents the chance to be entered into a prize draw for an Apple iPod Nano in return for completing the main study. Entry to the prize draw was optional, which I informed participants at the start of the study alongside their right to withdraw from the survey at any point and the guarantee of anonymity. To enter the prize draw, respondents simply had to enter their email address on the final screen. When all responses were collected, I printed out the list of email addresses, and asked a member of my extended family (who had not taken part in the study) to select a winning email address at random. In order to preserve the anonymity I had promised respondents, I kept all email addresses separate from the survey data; downloading the two sets of information as separate files.

## 5.2) Methodologies and Findings of Main Productive Study

This section presents the methodologies and findings of the four productive experiments I implemented as the main study[15].

### 5.2.1) Experiment 1: Matching Coined Words to Single Image

#### 5.2.1.1) Methodology

For my first experiment, I retained question type 1 from the pilot study in its entirety, save for the modifications outlined above. In each experiment of this test type, I presented respondents with audio recordings of three monosyllabic coined words. The coined words were identical in phonetic form apart from their onsets. Only one of the words featured the

[15] My sincere thanks to Dr. Angela Williams and Dr Andreas Buerki (Cardiff University) for their invaluable advice and assistance in interpreting the statistical data from these experiments.

onset phonaestheme under test. In addition to the three coined words, I presented respondents with a single large image, depicting the meaning associated with the particular phonaestheme under test. I invited respondents to listen to each word as many times as they wished, and then select the word which they felt best described or named a specified part of the image. As before, a screenshot of this question type, taken from the English survey, is included below for the purpose of clarification (Figure 10, overleaf):

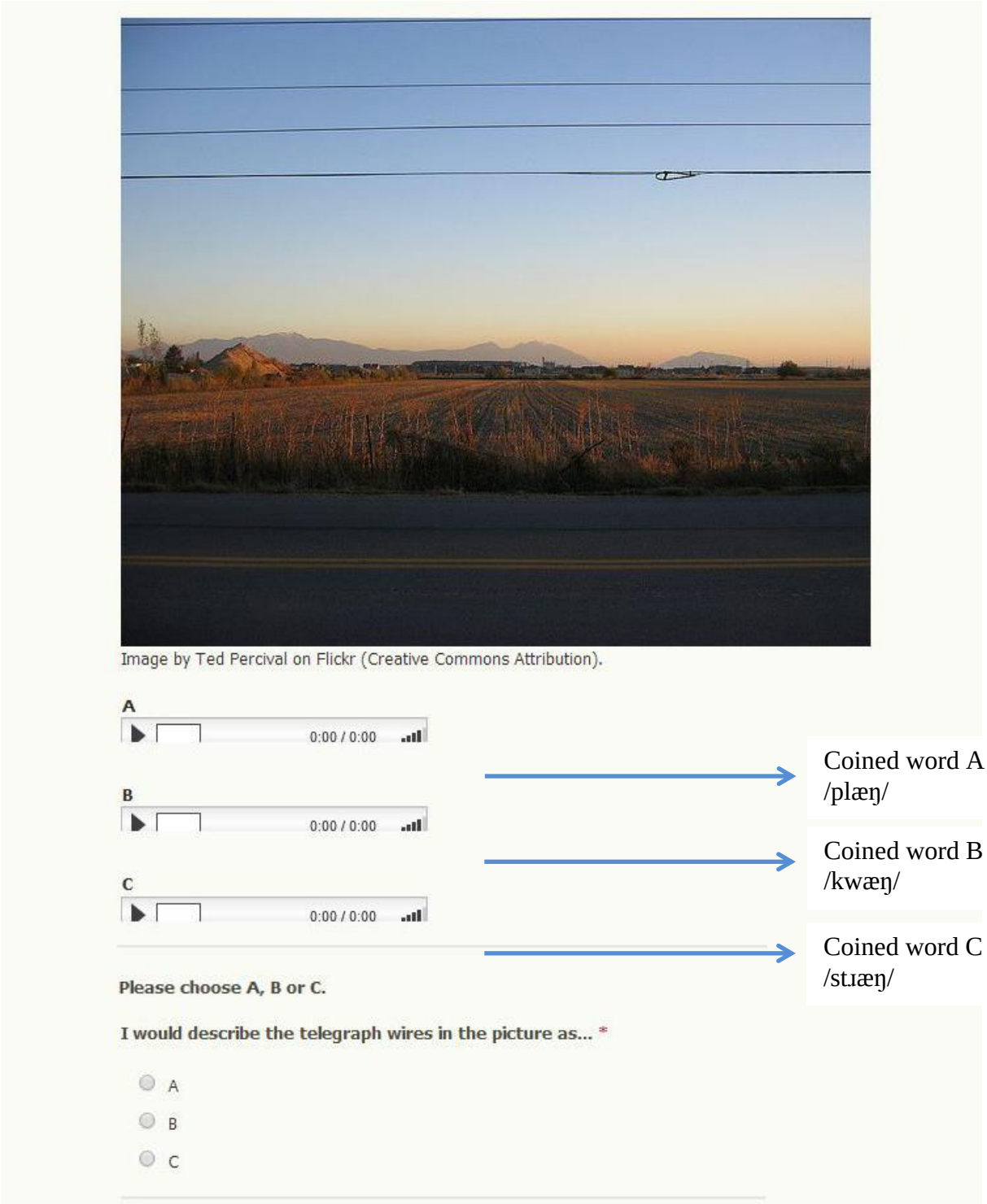



Figure 10: Screenshot of question type 1 testing the /stɹ/-<<straightness and stretching>> phonaestheme (main study). (Word C would be chosen if the phonaestheme was productive.)

My analysis of the findings from this experiment proceeded as follows. I began by counting the number of respondents in each language who chose the phonaestheme in each question. I recorded this data in a contingency table (cf. Table 18, below). I then studied the data, beginning with the responses of the English speakers. I identified the most and least productive phonaesthemes, as well as the main trends in the data; and discussed what the findings (both individually and at-large) suggested about the perception of the phonaesthemes in English. For example, if the majority of respondents used a given phonaestheme productively, it would suggest that this association was a perceptible pattern in English (cf. definition of 'perceive' in § 1.2). Conversely, if the majority of respondents avoided using a phonaestheme productively, it would suggest that the pattern might not be perceptible. I then repeated this process for the French and Polish data, and compared the productivity of the phonaesthemes in French and Polish with their productivity in English, with the aim of revealing which of the phonaesthemes were most productive in each language.

The findings of Experiment 1 are discussed below. Before presenting the data, it is worth noting an important caveat to this analysis, which also applies to the findings of Experiment 2. In multiple choice experiments such as these, there is a higher probability that a proportion of the results may have come about through chance; particularly in comparison to more open-ended tests such as Experiment 3. This is simply due to the fact that the range of potential answers is more constrained: In Experiment 3, participants could choose from a potentially infinite number of different answers (phonotactic sequences). By contrast, in Experiments 1 and 2, participants had to choose between one of three possible answers. The consequence of this constraint is that there are fewer options from which respondents could have selected by chance; i.e. if they did not perceive any kind of relationship between the image(s) and the coined word(s) supplied. Indeed, it is possible that as many as a third of participants could have chosen the phonaestheme by chance in each question from Experiments 1 and 2: If none of the participants perceived a sound-meaning relationship between the coined word(s) and image(s), simple probability would suggest that a roughly equal proportion of speakers would select each available option. This is akin to flipping a coin: If one flips a normal coin which has not been modified in any way (i.e. where there is nothing particular about the coin to dictate which way it would land), one would expect the coin to land heads-up and tails-up a roughly equal number of times (Meester, 2008:38; Taylor, 2015:4-5). The implication of this for my analysis is that, since a third of respondents could potentially have 'used' each phonaestheme by chance, the number of respondents

identifying the phonaesthemes will ideally need to be as far in excess of ten as possible (in each language) in order to suggest that the patterns in question may have been perceived. The greater the number of respondents [over one third] to have used each pattern productively, the greater the extent to which the findings will suggest that the phonaesthemes were in fact perceived. Conversely, the fewer the number of respondents [under one third] to have used each pattern productively, the greater the extent to which no perception of the pattern can be inferred from the data.

#### 5.2.1.2) Results

Table 18, below, shows the number of speakers who chose the coined words featuring the phonaesthemes in response to each image in Experiment 1:

Table 18: The number of English, French and Polish speakers using each phonaestheme productively in Experiment 1 (ranked in descending order in English)

| **Phonaestheme** | **No. English speakers choosing phonaesthemic word for image** | **No. French speakers choosing phonaesthemic word for image** | **No. Polish speakers choosing phonaesthemic word for image** |
|---|---|---|---|
| /gɹ/-<<gripping>> | 27 | 17 | 18 |
| /gl/-<<light>> | 25 | 12 | 14 |
| /st/-<<strength and stoicism>> | 23 | 21 | 20 |
| /tw/-<<smallness>> | 23 | 17 | 20 |
| /skw/-<<weakness; cuteness>> | 20 | 16 | 13 |
| /sw/-<<fast/strong movement>> | 19 | 16 | 20 |
| /stɹ/-<<straightness/stretching>> | 18 | 9 | 16 |
| /kl/-<<clumsiness, thickness, stupidity>> | 17 | 11 | 8 |
| /sl/-<<pejoration>> | 11 | 14 | 13 |
| /sm/-<<pejoration>> | 11 | 11 | 12 |
| /sn/-<<pejoration>> | 11 | 12 | 8 |
| /kɹ/-<<crookedness>> | 7 | 8 | 8 |

Considering first the English data; eight of the twelve phonaesthemes were used productively by more than half of all English speakers in this experiment. Onsets /gɹ/, /gl/, /st/, /tw/, /skw/, /sw/, /stɹ/ and /kl/ were each associated with the meanings identified in Chapter 1 by more than fifteen speakers. Not only do these frequencies exceed the roughly ten speakers who might have been expected to choose the phonaesthemes by chance (see above), but they also entail that the eight onsets in question appear to have been preferred over any others available for their respective images[16]. This suggests that these eight phonaesthemes may be perceptible patterns in English; and that at some level, speakers may have recognised associations between these onsets and the meanings under test. Among these, the most productive phonaestheme in this test was /gɹ/, which was attributed to the image depicting <<gripping>> by twenty-seven English respondents. This was followed by onset /gl/, which was associated with <<light>> by twenty-five respondents; and onsets /st/ and /tw/, which were associated with <<strength; stoicism>> and <<smallness>> by twenty-three respondents apiece. In proportional terms, more than three-quarters of English respondents associated these four onsets with the meanings under test.

Not all of the phonaesthemes attained similar levels of productivity, however. At the lower extreme, just seven respondents associated onset /kɹ/ with <<crookedness>>, while eleven respondents associated the /sl/, /sm/ and /sn/ words with <<pejoration>>. The fact that less than a third of the cohort associated /kɹ/ with <<crookedness>> suggests that speakers may actually have chosen to **avoid** using this pattern; at least in this particular test. It is potentially more difficult to interpret the /sl/, /sm/ and /sn/ findings; since eleven speakers represent roughly one third of the respondent cohort, it is possible that the frequencies attained by these phonaesthemes may have come about by chance (see above). However, the findings indicate that /kɹ/, /sl/, /sm/ and /sn/ were the four least productive phonaesthemes in this experiment. What is more, the raw data showed that English speakers preferred one or more non-phonaesthemic prosodies over each of these onsets in response to their respective images. In combination, these findings suggest that the twelve phonaesthemes might vary in the extent to which they are perceptible patterns in English. It is particularly notable that the /sl/-<<pejoration>> phonaestheme was the second least-productive phonaestheme in this experiment. As discussed in Chapter 4 (§ 4.2.2), this phonaestheme attained the highest type frequency in the vocabulary of English, and the third-highest token frequency. The fact that

[16] I double-checked the raw frequency data to confirm that this was not also true of the remaining phonaesthemes.

this phonaestheme was one of the least productive therefore suggests that the productivity (and by extension, the perception) of a phonaestheme might not be dictated by either its type or token frequency. This suggestion appears to agree with the fact that onset /gɹ/ attained the highest level of productivity of all twelve phonaesthemes in this experiment, yet attained the second lowest type frequency and third lowest token frequency (cf. § 4.2.2). This point is discussed further when the overall productivity of the phonaesthemes [in English] is compared to their type and token frequencies (cf. § 5.4).

In both French and Polish, five of the twelve phonaesthemes were used productively by more than half the respondents. These included /gɹ/, /st/, /sw/ and /tw/ in both languages; as well as /skw/ in French and /stɹ/ in Polish. As in English, these frequencies exceed the roughly ten speakers who might have been expected to choose the phonaesthemes by chance. They also entail that these onsets must have been preferred over any others available for their respective images. In combination, these findings suggest that the patterns might not have been perceived by English speakers alone in this experiment; but might also have been recognised to some extent by French and Polish speakers. It is particularly notable that /skw/ was attributed to <<weakness and cuteness>> by over half the French speakers sampled, since /skw/ does not occur in onset position in French. As such, the use of this prosody in onset position should not – in theory – evoke any connotations in French. However this does not appear to have been the case. The same is true of /tw/ in Polish. As the phonotactics of Polish do not permit /tw/ in onset position, the use of this prosody in onset position should not – in theory – evoke any connotations among Polish speakers. But again, this does not appear to have been the case. Over half of all Polish speakers sampled associated /tw/ with <<smallness>>.

In French and Polish, as in English, there were differences in the levels of productivity attained by the twelve phonaesthemes. In French, the most productive phonaestheme was onset /st/. Twenty-one speakers associated the /st/ word with the image depicting <<strength and stoicism>>; a number almost equalling the twenty-three English speakers who made the association. By contrast, the least productive phonaestheme was onset /kɹ/. Eight French speakers associated the /kɹ/ word with the image depicting <<crookedness>>; a number exceeding the seven English speakers who made the same association. This finding could reflect the fact that <<crookedness>> is not one of the most frequently-recurring meanings of /kɹ/ headwords in English (cf. Table 14, § 4.2.2). Perhaps

English speakers might have avoided associating /kɹ/ with <<crookedness>> because they associated this onset with another [different] meaning (see further discussion in Chapter 6). In Polish, the most productive phonaesthemes were onsets /st/, /tw/ and /sw/, which were associated with their respective images by twenty respondents apiece. For onset /sw/, this frequency was greater than the number of English respondents who made the association (nineteen speakers). The least productive phonaesthemes were onsets /sn/, /kɹ/ and /kl/, which were associated with their respective images by eight respondents apiece. As in French, this meant that a greater number of Polish speakers associated the /kɹ/ word with <<crookedness>> than the number of English speakers who made the association. Notably, a greater number of both French and Polish speakers also associated the /sl/ word with <<pejoration>> than the number of English speakers who made the association, even though the type frequency of /sl/ in English was over twenty times higher than its type frequency in either French or Polish (cf. § 4.2.2).

Despite the fact that a number of the phonaesthemes appear to have been more productive in French and/or Polish than in English, it is clear that a number of the phonaesthemes appear to have been more productive in English than in either other language. Onset /gɹ/ was associated with <<gripping>> by twenty-seven English speakers, in comparison to eighteen Polish and seventeen French speakers. Similarly, onset /gl/ was associated with <<light>> by twenty-five English speakers in this experiment, in comparison to fourteen Polish and twelve French speakers. Both onsets were over thirty percent more productive in English than in either French or Polish, and attained the two highest productive frequencies in the entire experiment. In addition, onsets /kl/, /st/, /tw/, /skw/ and /stɹ/ were all associated with their respective meanings by a greater number of speakers in English than in French or Polish. This suggests that these seven phonaesthemes – but particularly /gɹ/ and /gl/ – could be more widely-perceived patterns in English than in French or Polish.

### 5.2.2) Experiment 2: Matching Images to a Single Coined Word

#### 5.2.2.1) Methodology

I retained question type 2 from the pilot study in its entirety, save for the modifications outlined in § 5.1.8 (above). In each question of this type, I presented respondents with an audio recording of a single coined word, featuring one of the twelve phonaesthemes. I also

presented respondents with three images. Only one of the images depicted the meaning associated with the phonaestheme in the coined word. I invited respondents to listen to the coined word as many times as they wished, and then select the image they felt was best described or named by the coined word.

When constructing the coined words for this experiment, I made sure that each phonaestheme was combined with a different syllable rime to the one used in Experiment 1. The reason for this was as a control measure. Although not discussed anywhere in the literature, it is possible that the connotations ascribed to the phonaesthemes could have been affected by the syllable rimes with which these onsets were combined. By varying the syllable rimes across experiments, I sought to minimise the chance of any phonaestheme attaining unrepresentative findings in multiple tests if a particular rime had affected its connotations. For this reason, I also varied the rimes in the coined words used in Experiment 4. (The specific rimes I used in each question are shown in Appendix 18, where the main study is provided in screenshots, with phonemic transcriptions of the coined words used in each question.) When I analysed the findings of the experiments, it transpired that the productivity of the phonaesthemes appeared to be largely congruent across the four experiments (cf. summary in § 5.5). This suggests that the rime prosodies did not substantially affect the productivity of the onsets. However, at the time of conducting the study, this was something I could not have presupposed.

A screenshot of this question type is included overleaf for the purpose of clarification:

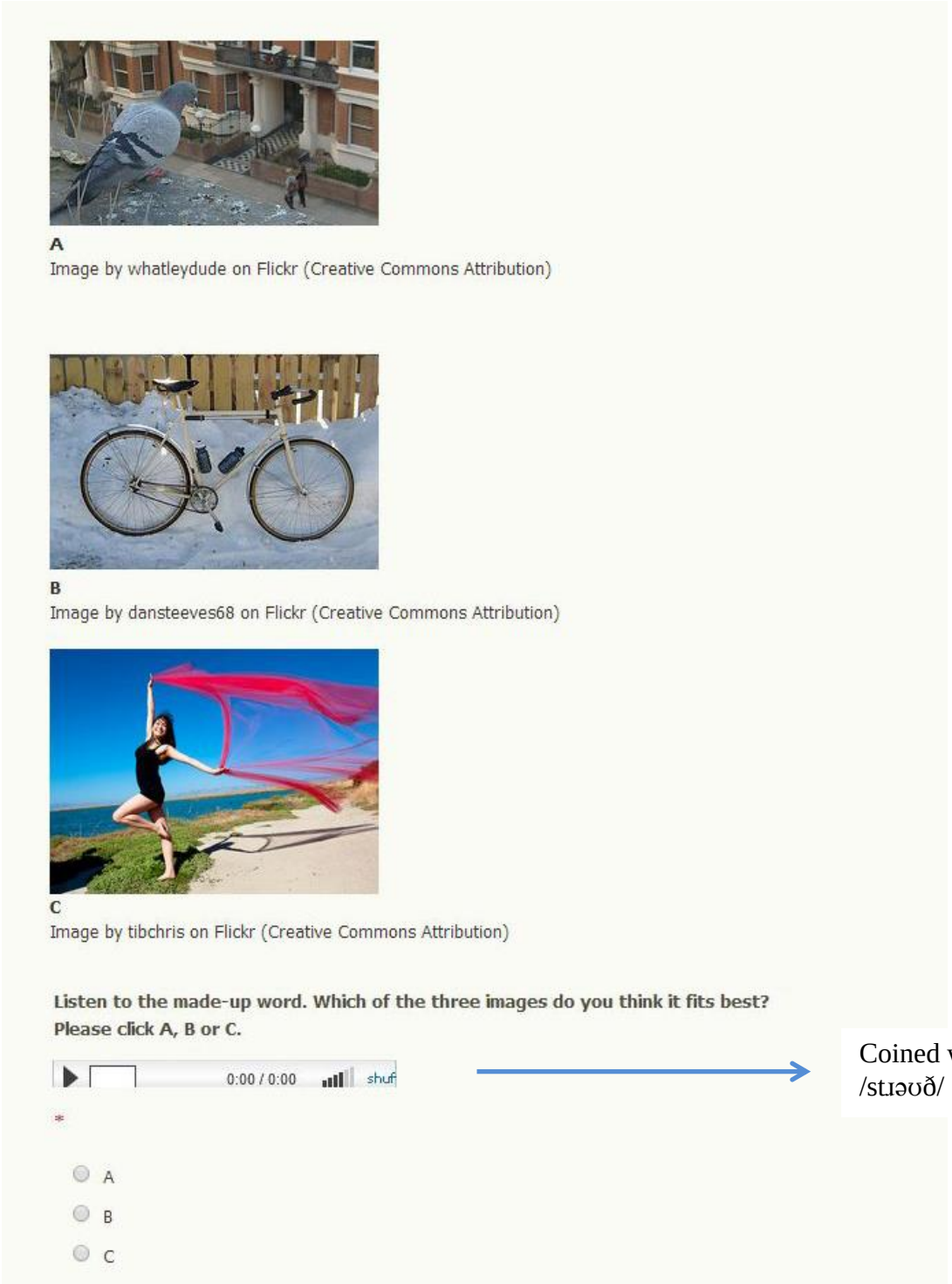



Figure 11: Screenshot of question type 2 testing the /stɹ/-<<straightness and stretching>> phonaestheme (main study). (Image C represents <<straightness and stretching>>.)

My analysis of findings from this test involved similar stages to those used for Experiment 1. For each phonaestheme, I counted the number of respondents in each language

cohort who attributed the intended image to each coined word. I recorded this data in a contingency table (Table 19, below). I then studied the data, beginning with the responses of the English speakers. I identified the most and least productive phonaesthemes, as well as the main trends in the data; and discussed what the findings (both individually and at-large) suggested about the perception of the phonaesthemes in English. In so doing, I briefly considered whether the findings were largely congruent or incongruent with those of test 1. I then repeated this process for the French and Polish data; during which I compared the productivity of the phonaesthemes in French and Polish with their productivity in English, with the aim of revealing which of the phonaesthemes were most productive in each language.

#### 5.2.2.2) Results

Table 19, overleaf, shows the number of speakers who chose each phonaestheme's associated image in response to the coined words in Experiment 2.

Table 19: The number of English, French and Polish speakers using each phonaestheme productively in Experiment 2 (ranked in descending order in English)

| **Phonaestheme** | **No. English speakers choosing phonaesthemic image for word** | **No. French speakers choosing phonaesthemic image for word** | **No. Polish speakers choosing phonaesthemic image for word** |
|---|---|---|---|
| /gl/-<<light>> | 27 | 18 | 20 |
| /sw/-<<fast/strong movement>> | 23 | 11 | 12 |
| /sl/-<<pejoration>> | 22 | 16 | 13 |
| /sn/-<<pejoration>> | 22 | 14 | 13 |
| /st/-<<strength and stoicism>> | 22 | 11 | 16 |
| /kl/-<<clumsiness, thickness, stupidity>> | 21 | 16 | 14 |
| /gɹ/-<<gripping>> | 19 | 21 | 12 |
| /sm/-<<pejoration>> | 19 | 13 | 18 |
| /kɹ/-<<crookedness>> | 18 | 15 | 15 |
| /skw/-<<weakness; cuteness>> | 18 | 13 | 15 |
| /stɹ/-<<straightness/stretching>> | 17 | 7 | 12 |
| /tw/-<<smallness>> | 15 | 12 | 18 |

Considering first the English data; in this experiment, all twelve phonaesthemes were identified by fifteen speakers or more. These frequencies all appear to exceed the roughly ten speakers who might have been expected to choose the phonaesthemes by chance. This appears particularly true of the /gl/, /sw/, /sl/, /sn/, /st/ and /kl/ phonaesthemes, which were used productively by more than twenty speakers each (two thirds of the respondent cohort). By definition, these findings entail that eleven of the twelve phonaesthemes (all except /tw/) were used productively by more than half the respondent cohort. For these onsets, then, the images depicting the meanings under test were the ones most frequently-chosen in response to the coined words. Since the /tw/ phonaestheme had been used by exactly half the respondents, I looked back at the raw frequency data. In-keeping with the other eleven phonaesthemes, I found that the image depicting <<smallness>> had also been attributed to

the /tw/ coined word by a greater number of speakers than either other image (the non-phonaesthemic images were used ten and five times apiece). Overall, these findings suggest that English speakers may have perceived all twelve phonaesthemic patterns in this experiment: Every single onset was associated with the meaning shown in § 1.1 to a greater extent than with any other meanings. In general, these findings appear to be congruent with those of Experiment 1; the only exceptions being onsets /sl/, /sm/, /sn/ and /kɹ/, which did not appear to have been so productive in Experiment 1.

As in Experiment 1, however, not all of the phonaesthemes attained similar levels of productivity. At the upper extreme, the /gl/-<<light>> phonaestheme was the most productive association in this experiment; a finding that seems congruent with its comparatively high level of productive use in Experiment 1. In this second test, twenty-seven speakers attributed the image depicting <<light>> to the coined word. This was followed by the /sw/-<<strong/fast movement>> phonaestheme, where twenty-three speakers attributed the image depicting <<fast movement>> to the coined word. The level of productivity attained by /gɹ/ in Experiment 1 fell by eight responses in this experiment. Here, nineteen speakers attributed the image of <<gripping>> to the /gɹ/ word. Notably, all three <<pejorative>> phonaesthemes appear to have been more productive in this experiment than in Experiment 1. However, the /sl/-<<pejoration>> phonaestheme was still less productive than the /gl/ and /sw/ phonaesthemes. This seems to strengthen to the earlier interpretation that a phonaestheme's type frequency might not wholly dictate its productivity. However, in this experiment, the phonaestheme with the highest *token* frequency (/gl/) attained the highest level of productivity; so it is possible that token frequency may influence the productivity of a phonaestheme to some extent. The discussion returns to this point in § 5.4. At the lower extreme, the least-productive phonaestheme in this test was the /tw/-<<smallness>> association, as noted above.

In both French and Polish, four phonaesthemes were used productively by more than half the respondent cohort. In French, more than half the respondents chose the relevant images in response to the /gl/, /sl/, /kl/ and /gɹ/ coined words[17]; while in Polish, more than half the respondents chose the relevant images in response to the /gl/, /st/, /sm/ and /tw/

[17] i.e. those depicting <<light>>, <<pejoration>>, <<clumsiness; thickness; stupidity>> and <<gripping>>, respectively

coined words[18]. As in English, these frequencies exceed the roughly ten speakers who might have been expected to choose the phonaesthemes by chance. Moreover, these frequencies also entail that the images depicting these meanings were preferred over any others in response to the coined words. As in Experiment 1, this suggests that the phonaesthemic associations might not have been perceived by English speakers exclusively in this experiment. Rather, a number of the phonaesthemes may also have been perceived by French and Polish speakers.

Notwithstanding this, however, the findings suggest that the majority of the phonaesthemes appear to have been more productive in English than in French and Polish. Ten of the twelve phonaesthemes (/gl/, /sw/, /sl/, /sn/, /st/, /kl/, /sm/, /kɹ/, /skw/ and /stɹ/) attained higher levels of productivity in English than in either other language. Of the two remaining onsets, /gɹ/ was more productive in French than English, but more productive in English than Polish; and /tw/ was more productive in Polish than English, but more productive in English than French. The fact that the majority of phonaesthemes appear to have been more productive in English than in either other language is another finding congruent with Experiment 1. In Experiment 1, seven of the twelve phonaesthemes (/gl/, /gɹ/, /kl/, /st/, /tw/, /skw/ and /stɹ/) were more productive in English than in French or Polish. The above lists show that, over the two experiments, four phonaesthemes appeared to have been consistently more productive in English than in either other language. These include /gl/, /st/, /skw/ and /stɹ/.

Despite the numerous congruences between these findings and those of Experiment 1, there are a number of notable contrasts between the French and Polish data in this test and the French and Polish data in Experiment 1. Whereas the /st/ phonaestheme was the most productive in French in Experiment 1, the /gɹ/ phonaestheme was the most productive in French in this experiment. Similarly, whereas the /tw/, /st/ and /sw/ phonaesthemes were the most productive in Polish in Experiment 1, the /gl/ phonaestheme was the most productive in Polish in this experiment. In both languages, a number of phonaesthemes rose in productivity from experiment 1 (these include /gɹ/, /gl/, /kl/, /kɹ/, /sn/ and /sm/); while a number of other phonaesthemes fell in productivity from Experiment 1 (these include /st/, /tw/, /sw/ and /stɹ/).

[18] i.e. those depicting <<light>>, <<strength; stoicism>>, <<pejoration>> and <<smallness>>, respectively.

Interestingly, there does not appear to be any clear trend vis-à-vis the phonaesthemes that rose and fell in productivity in either language.

### 5.2.3) Experiment 3: Open-Ended Definitions

#### 5.2.3.1) Methodology

For my third question type in the main study, I retained question type 4 from the pilot. In each question of this test, I provided respondents with a definition related to one of the twelve phonaesthemes under test. I then asked respondents to 'invent' any new word of their choosing to name or describe the definition (cf. discussion of 'name' and 'describe', § 5.1.1). Words could be of any length or syllable structure, since I was only interested in the word onsets used by respondents; i.e. those in the first syllable. My only stipulation was that I asked respondents to avoid knowingly using an existing word in any language they spoke. The definitions I provided were as specific as possible without actually using a word containing the phonaestheme under test, which could have primed respondents (see discussion in § 5.1.4). For example, in the question testing the /stɹ/-<<straightness and stretching>> phonaestheme, I asked respondents; "please invent a name for a long, linear metal bar used to keep things level". (All definitions used are shown in Appendix 18.)

I provided one definition to test each phonaestheme, with the exception of the three <<pejorative>> onsets (/sl/, /sm/ and /sn/). For these phonaesthemes, I provided three different <<pejorative>> definitions:

- "a large, vicious animal which silently stalks its prey and attacks when least expected";
- "a dodgy nightclub frequented by people of low moral standards";
- "the satisfied look worn by someone who feels they are better than everyone else"

Since any of the three <<pejorative>> onsets could plausibly have been used in response to each definition, I reasoned that I would treat each of these questions as a potential test for /sl/, /sm/ and /sn/. However, my motivation for using three different <<pejorative>> definitions was to explore whether speakers might associate onsets /sl/, /sm/ and /sn/ with different sorts of <<pejorative>> meanings. This is because Firth (1930:185) claimed that the pejorative

connotations of onsets /sl/, /sm/ and /sn/ were all 'subtly different', but did not suggest how. Thus, the first definition was intended to evoke connotations of animalistic violence; the second to evoke connotations of promiscuity and the third to evoke connotations of gloating and condescension. Using the same criteria on which I had judged the /sl/, /sm/ and /sn/ head lexemes in Chapter 4[19], I reasoned that each of these associations would likely be deemed <<pejorative>> in the speech communities under investigation. Of course, there are many other meanings which would also cohere with these criteria, and which would also be deemed <<pejorative>> in the speech communities surveyed. However, I chose these three associations because of their congruence with the words provided by Firth (and other scholars) to exemplify the /sl/, /sm/ and /sn/ phonaesthemes. To clarify: The <<pejorative animalistic>> associations of the first definition appeared to be congruent with many of Firth's (1930; 1935) exemplar words for the /sn/ phonaestheme, including 'sneak', 'snap' and 'snatch'. Similarly, the <<pejorative promiscuous>> connotations of the second definition appeared to concur with many of Firth (1930; 1935) and Crystal's (2002) exemplar words for the /sl/ phonaestheme, such as 'slut', 'slag', 'sleazy' and 'slattern'. Finally, the associations of <<pejorative condescension>> evoked in the third definition appeared to be congruent with many of Firth's (1930; 1935) exemplar words for the /sm/ phonaestheme, including 'smirch', smirk' and 'smug'. As such, I reasoned that speakers might tend to coin words featuring onset /sn/ in response to definition one, by analogy to the various /sn/-<<pejorative animalistic>> words in English. I reasoned that speakers might tend to coin words featuring onset /sl/ in response to definition two, by analogy to the various /sl/-<<pejorative promiscuous>> words in English. Finally, I reasoned that speakers might tend to coin words featuring onset /sm/ in response to definition three, by analogy to the various /sm/-<<pejorative condescending>> words in English.

Although my decision to test /sl/, /sm/ and /sn/ with three different <<pejorative>> definitions offered a way of exploring possible differences in connotation between the onsets, there is a possible criticism which could be directed at this method. It could be argued that analogous modelling on existing word forms, in the manner shown above, would not necessarily demonstrate evidence that speakers had perceived the /sl/, /sm/ and /sn/ phonaesthemes. For instance, if a speaker coined an onset /sl/ word in response to the second

[19] Namely, 'any action, quality, person or thing which can cause any type of harm or bring about an effect that is culturally perceived (at the time of writing) to be disadvantageous to anyone or anything; any expression which can in any sense be used to offend, criticise, insult or degrade; any action, quality, person or thing which is in any sense culturally perceived as unpleasant at the time of writing' (cf. Table 12).

definition by analogy to the existing English words ‘slut’ and ‘sleazy’, one could argue that this did not necessarily indicate speakers’ perception of an experiential link between word-initial /sl/ and a meaning of <<pejoration>> (or a particular type of pejoration); just a productive analogy to the sound structures of two existing English words. Similarly, if a speaker coined an onset /sm/ word in response to the third definition by analogy to the existing English words ‘smug’ and ‘smirk’, one could argue that this did not necessarily indicate speakers’ perception of an experiential link between word-initial /sm/ and a meaning of <<pejoration>> (or a particular type of pejoration); just another productive analogy to the sound structures of two existing English words. In reality, this is a potential criticism which could be levelled at every question of this test type. Unfortunately, the possibility that such analogies could be mistaken for the perception of phonaesthemes is a point not discussed anywhere in the literature; perhaps unsurprisingly, given the general lack of discussion about the perceptibility of phonaesthemic patterns and how this might come about (cf. Chapter 1.) However, I determined that this scenario would not actually pose any issues for my interpretation of the data from this experiment, and could not form the basis of a counterargument. There are several reasons for this. First, if speakers were to coin words with onsets /sl/, /sm/ or /sn/ in response to any of the <<pejorative>> definitions, it would not necessarily matter what sorts of cognitive processes had led to this outcome[20]. The end result would still represent a productive use of one or more of the patterns under investigation – in this case, a productive association between onsets /sl/, /sm/ and/or /sn/ and a meaning of <<pejoration>> – which is ultimately what my survey experiments were designed to explore. The same would also apply for the questions testing the other nine phonaesthemes. For instance, if speakers coined words featuring onset /gl/ in response to the definition expressing <<light>>, this would still constitute a productive use of the /gl/-<<light>> phonaestheme; howsoever this productive use came about. Moreover, even if speakers **were** to make analogies to existing words in the manner described above, this would not necessarily entail that speakers had not perceived the phonaesthemes; or that I would be mistaken to suggest this interpretation in my analysis of the data. Rather, the end result would suggest that speakers **had** perceived the phonaesthemic patterns to some extent. This is because speakers would ultimately have made a decision to use these particular onsets, even though there were other existing words – which did not feature the phonaesthemes – to which they could also have made analogies. For example, if a speaker coined an onset /sn/ word in response to

[20] And in any case, it would be very difficult if not impossible to determine the precise cognitive processes at work (cf. discussion in § 1.4.2).

definition one, the speaker would have chosen to use /sn/ in preference to any other onset, even though there are many other English words which do not feature onset /sn/ which they could also have used analogously in this context (such as ‘growl’, ‘roar’, ‘predator’, ‘stealthy’, ‘nasty’, and so on). Similarly, if a speaker coined an onset /sl/ word in response to definition two, the speaker would have chosen to use /sl/ in preference to any other onset, even though there are many other English words which do not feature /sl/ which they could also have used analogously in this context (such as ‘dirty’, ‘skeazy’, ‘grotty’, and so on). Thus, if /sl/, /sm/ and/or /sn/ were used more frequently than any other onsets in response to the three <<pejorative>> definitions, it would suggest that the speaker group(s) in question had still perceived a level of association between /sl/, /sm/ and/or /sn/ and <<pejoration>>: Speakers would have chosen to use these particular onsets over any others in their coined words, despite the fact that there were other phonetic forms which could also have been used analogously.

To analyse the data generated by this experiment, I counted and compared the number of speakers in each language group who did and did not use each of the phonaesthemes productively in their coined words. I then identified the most and least productive phonaesthemes in this test, as well as the main trends in the data; and discussed what the findings (both individually and at-large) suggested about the perception of the phonaesthemes in each language group. In so doing, I briefly considered whether the findings were largely congruent or incongruent with those of experiments 1 and 2.

In order to establish from respondents’ written (graphological) responses whether or not they had used the phonaesthemes in their coined words, I first had to research the possible graphological sequences that could be used to represent the twelve onsets in English, French and Polish (cf. § 3.2.3). These are shown in Table 20, overleaf.

Table 20: Possible graphological representations of the twelve onsets in English, French and Polish (Wijk, 1966; Tranel, 1987; Swan, 2002; Ager, 2014).

| **Language** | **Graphological sequence(s) used to realise each onset** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | /sl/ | /sm/ | /sn/ | /st/ | /stɹ/ | /kl/ | /kɹ/ | /gl/ | /gɹ/ | /sw/ | /skw/ | /tw/ |
| English | “sl” | “sm” | “sn” | “st” | “str” | “cl”; “kl” | “cr”; “kr” | “gl” | “gr” | “sw” | “squ” “skw” | “tw” |
| French | “sl” | “sm” | “sn” | “st” | “str” | “cl”; “kl” | “cr”; “kr” | “gl” | “gr” | “sw” | - | “tw” |
| Polish | “sl” | “sm” | “sn” | “st” | “str” | “kl” | “kr” | “gl” | “gr” | “sł” | “skł” | - |

Although this open-ended question type appeared to have been fairly reliable in the pilot study (cf. § 5.1.7), there was a small complication to interpreting the data it generated. If I had claimed anything about the productivity of phonaesthemes from this data without any further testing, I could have implied that every possible [phonotactically-permissible] onset in each language had an equal likelihood of being chosen; and that it was only the connotations ascribed to the phonaestheme that caused this onset to be chosen in preference to any other. However, as noted in my discussion of Abelin (1999) in Chapter 3, this is something that I could not necessarily have presupposed. There may have been another factor besides the phonaestheme affecting speakers’ choice of onsets; specifically, the frequency with which speakers encountered particular onsets in their day-to-day experiences of language. It is possible that the onsets encountered most frequently by speakers in everyday linguistic experiences could have stood a greater chance of becoming productive (cf. discussion of Usage-Based and Probabilistic models, § 3.1). For example, imagine that ten English speakers coined a word featuring onset /sl/ to describe the <<pejorative>> definition, and this frequency was greater than that attained by any other onset in this question. At face value, this result could have suggested fairly convincing evidence about the productivity of the /sl/-<<pejoration>> phonaestheme. However, it may be that English speakers tend to encounter /sl/ more frequently than any other onset in words describing things (i.e. adjectives). In this case, the result would have appeared less convincing. Without supplementing the findings

with knowledge about which onsets speakers tended to encounter most and least frequently in everyday language use, it would have been more difficult to reach an informed judgment of the findings.

To help overcome this issue in interpreting the frequency data, I added a final corpus-based analysis to this experiment. Since there was no motivation to predict that the phonaesthemes would be productive in French and Polish (cf. § 3.3), I undertook this analysis in English only. Using the BNC, I sampled 1,000 random nouns (including mass, singular and plural common nouns and proper nouns), and 1,000 random adjectives. My reasoning for sampling both nouns and adjectives was that some of the open-ended questions asked respondents for a 'word to name...' (i.e. a noun), whereas others asked for 'a word to describe…' (i.e. an adjective) (cf. discussion of 'name' and 'describe' in § 5.1.1). I displayed each of the 1,000 nouns and adjectives as a separate concordance line in KWIC format (the 1,000 concordance lines of nouns and adjectives are displayed as Appendices 19 and 20 respectively). I then counted the number of times each discrete onset appeared in these samples. I excluded from this count any instances of zero onset, any abbreviations (such as 'kg' or 'Mrs.'), any noun beginning with a numeral or symbol (such as '1000s' or '£30'), and any multi-word expressions (such as 'out-of-the-way' or 'close-cut'). I then compiled two tables; one showing the frequency with which each discrete onset was found in the noun corpus, and the other showing the frequency with which each discrete onset was found in the adjective corpus (Tables 22 and 23, below). These gave a representation of the onsets that English speakers tend to encounter most and least frequently in nouns and adjectives in everyday language use. I then used these tables to give a more informed consideration of the findings from each open-ended experiment: Whenever a phonaestheme was used in one or more speakers' coined words, I consulted the tables to see how frequently this onset was encountered in the type of lexical context required by its question (whether nominal or adjectival). The lower the frequency with which the phonaestheme was encountered in that particular lexical context, the greater the confidence it would offer that the result reflected some level of productivity (and thus perception) of the pattern; and was not simply caused by the frequency with which speakers encountered the onset.

### 5.2.3.2) Results

Table 21, below, shows the number of speakers in each language who coined (and did not coin) words featuring the phonaestheme under test in each definition. In this table, the leftmost column provides a synopsis of the definition and indicates the phonaestheme (onset) under test. From left to right, the next three columns show the number of English, French and Polish speakers who did and did not use this onset in their coined words. A tick indicates the number of speakers who used the onset under test; a cross indicates the number of speakers who did not:

Table 21: The number of English, French and Polish speakers who coined and did not coin words featuring each phonaestheme (Experiment 3)

| Synopsis of Definition **(phonaestheme under test)** | No. English Speakers using (✓) and not using (x) phonaestheme | No. French Speakers using (✓) and not using (x) phonaestheme | No. Polish Speakers using (✓) and not using (x) phonaestheme |
|---|---|---|---|
| Light<br>**- /gl/** | ✓12<br>x 18 | ✓1<br>x 29 | ✓3<br>x 27 |
| Pejoration<br>[definition 3]<br>**- /sl/, /sm/, /sn/** | ✓10 (/sm/)<br>x 20 | ✓3 (/sn/)<br>x 27 | ✓1 (/sl/)<br>x 29 |
| Clumsiness, thickness, stupidity<br>**- /kl/** | ✓9<br>x 21 | ✓1<br>x 29 | ✓1<br>x 29 |
| Fast/strong movement<br>**- /sw/** | ✓9<br>x 21 | ✓1<br>x 29 | ✓1<br>x 29 |
| Strength and stoicism<br>**- /st/** | ✓8<br>x 22 | ✓2<br>x 28 | ✓6<br>x 24 |
| Smallness<br>**- /tw/** | ✓6<br>x 24 | ✓1<br>x 29 | ✓0<br>x 30 |

Table Continues Overleaf

| Synopsis of Definition **(phonaestheme under test)** | No. English Speakers using (✓) and not using (x) phonaestheme | No. French Speakers using (✓) and not using (x) phonaestheme | No. Polish Speakers using (✓) and not using (x) phonaestheme |
|---|---|---|---|
| Gripping **- /gɹ/** | ✓4<br>x 26 | ✓2<br>x 28 | ✓2<br>x 28 |
| Straightness and stretching **- /stɹ/** | ✓4<br>x 26 | ✓2<br>x 28 | ✓0<br>x 30 |
| Weakness; cuteness **- /skw/** | ✓4<br>x 26 | ✓1<br>x 29 | ✓0<br>x 30 |
| Pejoration [definition 1] **- /sl/, /sm/, /sn/** | ✓8 (/sl/ = 5, /sn/ =3)<br>x 22 | ✓6 (/sn/ =5, /sm/ =1)<br>x 24 | ✓1 (/sl/)<br>x 29 |
| Pejoration [definition 2] **- /sl/, /sm/, /sn/** | ✓2 (/sl/)<br>x 28 | ✓2 (/sl/)<br>x 28 | ✓3 (/sl/)<br>x 27 |

In English, every single onset was used by two or more speakers in response to its associated definition. Given that respondents could have selected from eighty-one phonotactically permissible word onsets in English (Kreidler, 2004: Ch.6; Cruttenden, 2008: 254-255), the fact that all twelve phonaesthemes attained two or more uses in thirty responses suggests that all twelve phonaesthemic associations may have been perceived by English speakers to some extent. (This is discussed further below, in light of the onsets found most frequently in the sample of 1,000 random nouns and adjectives from the BNC.) This interpretation appears to agree with the interpretations drawn from both previous tests; particularly Experiment 2. However, even though all twelve phonaesthemes appear to have been perceived in English, it is once again clear that some attained greater productivity than others. In common with Experiment 2, the /gl/-<<light>> association was the most productive phonaestheme in this experiment. Twelve speakers – some forty percent of English respondents – coined a word featuring onset /gl/ in response to the definition ‘a word to describe a distant star sparkling in the heavens’. This was the greatest level of productivity attained by any phonaestheme across all three language groups in this experiment. The next

most-productive phonaestheme among English speakers in this experiment was the /sm/-<<pejoration>> association. Onset /sm/ appeared in ten of the thirty words coined for the third <<pejorative>> definition ('the satisfied look worn by someone who feels they are better than everyone else'). The presence of /sm/ as the second most-productive phonaestheme is notable. In Chapter 4, I found that /sm/ had only the seventh highest type frequency and eighth highest token frequency of the phonaesthemes in English (cf. Tables 13 and 15). Moreover, in Experiments 1 and 2, /sm/ appeared to be a relatively less productive phonaestheme in comparison to many of the other patterns under test; particularly in Experiment 1, where it was the joint second least productive of the twelve phonaesthemes. However, in this experiment, /sm/ was more productive than every single phonaestheme aside from /gl/. What is more, /sm/ attained this level of productivity in response to the particular definition that I had predicted would elicit /sm/ coined words (see further discussion below). As such, it appears that English speakers may, to some extent, have constructed the forms of their coined words by analogy to existing words in the language. Other notable results in this test include onsets /kl/, /sw/ and /st/, which, in common with /gl/ and /sm/, were all used by more than a quarter of English speakers in their coined words. Once again, these findings appear to be largely congruent with those of Experiments 1 and 2. In Experiment 2, /kl/, /sw/ and /st/ were all among the most productive phonaesthemes; attaining frequencies of use of 21, 23 and 22 respectively. Similarly, in Experiment 1, the /sw/ and /st/ phonaesthemes were again among the most productive patterns; although the /kl/-<<clumsiness; thickness; stupidity>> phonaestheme appeared to have been somewhat less widely-perceived in Experiment 1 than in Experiments 2 and 3.

In addition to the above, it is worth comparing the number of speakers using /sl/, /sm/ and /sn/ in response to the three <<pejorative>> definitions. As discussed above, /sm/ was used by ten speakers in response to definition three, in which I had predicted /sm/ might be the most frequently-used onset. However, onsets /sl/ and /sn/ were not used at all in this question. By contrast, onsets /sl/ and /sn/ were both used multiple times in response to definition one (/sl/ being used five times and /sn/ three times); where I had predicted that /sn/ might be the most frequently-used onset. Notably, /sm/ was not used at all in response to the first definition. Finally, only /sl/ was used in response to the second <<pejorative>> definition, in which I predicted that this would be the most frequently-used of the three onsets. In combination, these findings suggest that there may be differences in connotation between the /sl/, /sm/ and /sn/, although it is not clear at this stage precisely what these

differences might be. It appears that onset /sm/ was favoured by English speakers in response to the <<pejorative condescension>> expressed by definition three; and it appears that onset /sl/ may have been preferred by English speakers for the <<pejorative promiscuous>> qualities of definition two. However, it is notable that fewer speakers used one of the <<pejorative>> onsets in response to this definition than in response to definitions one and three. Finally, it appears that /sl/ may also have been preferred in response to the <<pejorative animalistic>> connotations of definition one – a finding which contravened my expectations somewhat – although onset /sn/ also appears to have been associated with this meaning to some extent. The possible differences in connotation between these onsets is a point which would clearly benefit from further research in the future, as I discuss in Chapter 7.2.

On comparing the English data with the French and Polish data, it appears that the majority of the phonaesthemes were more productive – and may have been more widely-perceived patterns – in English. The only phonaesthemes to attain equal or higher frequencies of use in French or Polish than in English were /sl/ and /sn/. Onsets /sl/ and /sn/ appeared to be similarly productive in both English and French: Two speakers in each language coined words featuring /sl/ in response to the second <<pejorative>> definition; and five speakers in each language coined words featuring /sn/ in response to the first <<pejorative>> definition. While the productivity of /sl/ in French appears generally congruent with Experiments 1 and 2 (in which /sl/ was among the most productive phonaesthemes in French), the number of French speakers associating /sn/ with <<pejoration>> appears to be somewhat unexpected, as /sn/ was neither among the most or least productive phonaesthemes in French across Experiments 1 and 2. To some extent, then, it appears that French speakers may have perceived a greater level of association between onset /sn/ and <<pejoration>> in this experiment than in previous tests; as one might not expect a fifth of the respondent group to have chosen the same onset by chance, given the potential number of word-initial sequences available to speakers. In addition to the above, three Polish speakers coined onset /sl/ words in response to the second <<pejorative>> definition, exceeding the two English and two French speakers who made the same association. The fact that /sl/ was one of the most productive phonaesthemes among Polish speakers in this experiment is again somewhat unexpected, as this pattern appeared to have been one of the least productive phonaesthemes in Experiments 1 and 2 (attaining thirteen uses in each test). Again, then, it is possible that Polish speakers may have perceived some level of association between onset /sl/ and <<pejoration>>; although this pattern does not appear to have been as productive as the /sn/-

<<pejoration>> phonaestheme in French.

Aside from the /sn/ phonaestheme in French and the /sl/ phonaestheme in Polish, the remainder of the associations were more productive in English. As noted above, this suggests that the majority of the phonaesthemes may have been more widely-perceived patterns in English than in either other language. This would appear to be supported by the fact that considerably fewer of the onsets were used by two or more speakers in French and Polish than in English. In French, seven onsets were used twice or more in respondents' coined words. These included /gl/, /st/, /kɹ/, /gɹ/, /stɹ/, /sl/ and /sn/; the latter being used by multiple respondents for two of the <<pejorative>> definitions. In Polish, just four onsets were used twice or more. These included /st/, /kɹ/, /gɹ/ and /sl/. By comparison, every single onset was used by more two or more English speakers in response to its associated definition.

Notwithstanding this, however, the frequency of use attained by /st/ in Polish is particularly notable. Although the number of Polish speakers coining /st/ words was lower than the number of English speakers, six Polish speakers (one fifth of the cohort) also used /st/ in response to the <<strength and stoicism>> definition. This finding constitutes the highest productive use of any phonaestheme in either French or Polish in this experiment; and suggests that, while the results at-large point towards a more widespread perception of the patterns in English, a number of the phonaesthemes also appear to have been perceived outside of English. The degree of productivity attained by /st/ in this experiment could reflect the fact that <<strength and stoicism>> is the most common meaning among /st/ headwords in the vocabulary of Polish (cf. Table 14). This also appears to concur with the findings of Experiments 1 and 2, in which the /st/-<<strength and stoicism>> phonaestheme was consistently one of the most productive phonaesthemes in Polish; attaining the joint highest and third highest frequencies of productive use respectively.

My final stage in analysing the findings of this experiment was to compare the English data against the 1,000 random nouns and adjectives sampled from the BNC. As discussed above, the purpose of this was to reach a more informed interpretation of the findings, by investigating whether the productivity of the onsets could have been influenced by their frequency of occurrence in everyday language use. I began by examining the sample of nouns. Table 22, overleaf, shows the frequency with which each distinct onset recurred in the sample. For economy of space, the onsets are presented in three columns, ranked in

descending order from those most frequently-occurring:

Table 22: Frequency of word onsets in a sample of 1,000 random English nouns in the BNC

| **Frequency of discrete onsets in a sample of 1,000 random English nouns from the BNC** | | | | | |
|---|---|---|---|---|---|
| **Onset** | **Frequency of Recurrence** | **Onset** | **Frequency of Recurrence** | **Onset** | **Frequency of Recurrence** |
| /k/ | 81 | /g/ | 16 | /kɹ/ | 4 |
| /m/ | 65 | /n/ | 16 | /kw/ | 4 |
| /p/ | 58 | /j/ | 14 | /stɹ/ | 4 |
| /ɹ/ | 55 | /sk/ | 12 | /fɹ/ | 3 |
| /s/ | 55 | /v/ | 9 | /mj/ | 3 |
| /d/ | 46 | /kl/ | 8 | /θɹ/ | 3 |
| /w/ | 46 | /ʤ/ | 8 | /gl/ | 2 |
| /f/ | 40 | /gɹ/ | 8 | /vj/ | 2 |
| /t/ | 38 | /ʃ/ | 8 | /skɹ/ | 1 |
| /l/ | 35 | /pl/ | 7 | /sm/ | 1 |
| /b/ | 30 | /θ/ | 7 | /sn/ | 1 |
| /h/ | 29 | /bl/ | 5 | /spɹ/ | 1 |
| /tɹ/ | 25 | /bɹ/ | 5 | /sw/ | 1 |
| /pɹ/ | 24 | /dɹ/ | 5 | /z/ | 1 |
| /st/ | 22 | /fl/ | 5 | | |
| /tʃ/ | 20 | /sp/ | 5 | | |

These findings are important to the /sm/, /tw/, /sw/, /skw/, /stɹ/ and /sn/ phonaesthemes. This is because the questions testing these onsets asked for ‘a word to *name*’ the associated definitions (i.e., a noun-like word). All six phonaesthemes were used in speakers’ coined words; but could this have been influenced by the frequency with which speakers tend to encounter these prosodies? Encouragingly, Table 22 suggests that this may have been unlikely. Thirty-two onsets occurred more frequently in the sample than /stɹ/, which was found just four times. Forty onsets occurred more frequently in the sample than /sm/, /sw/ and

/sn/, which were found only once each in the 1,000 nouns. Finally, none of the nouns in the sample featured either onset /tw/ or onset /skw/. As such, it seems unlikely that the frequency with which speakers encounter these onsets might have influenced their productivity and preference among respondents. This strengthens the interpretation that English speakers appear to have perceived these patterns in this experiment; particularly the /sm/, /tw/, /sw/, /skw/ and /stɹ/ phonaesthemes.

A similar scenario is evident for the six remaining phonaesthemes. The definitions testing /gl/, /gɹ/, /kl/, /st/, /kɹ/ and /sl/ all asked respondents to coin 'a word to *describe*' a definition related to their associated meanings (i.e., adjective-like words). Table 23, overleaf, shows the frequency with which each distinct onset recurred in the sample of 1,000 random adjectives:

Table 23: Frequency of word onsets in a sample of 1,000 random English adjectives in the BNC

| **Frequency of discrete word onsets in a sample of 1,000 random English adjectives from the BNC** | | | | | |
|---|---|---|---|---|---|
| **Onset** | **Frequency of Recurrence** | **Onset** | **Frequency of Recurrence** | **Onset** | **Frequency of Recurrence** |
| /s/ | 64 | /fɹ/ | 14 | /dɹ/ | 4 |
| /l/ | 56 | /g/ | 13 | /θ/ | 3 |
| /ɹ/ | 56 | /t/ | 12 | /sl/ | 2 |
| /k/ | 48 | /kl/ | 12 | /hj/ | 2 |
| /m/ | 45 | /ʃ/ | 12 | /fl/ | 2 |
| /d/ | 40 | /v/ | 12 | /mj/ | 2 |
| /p/ | 35 | /tɹ/ | 9 | /pj/ | 2 |
| /ɹ/ | 34 | /ʤ/ | 8 | /pl/ | 2 |
| /f/ | 33 | /bl/ | 7 | /kw/ | 2 |
| /pɹ/ | 28 | /st/ | 7 | /skw/ | 2 |
| /b/ | 19 | /kɹ/ | 6 | /fj/ | 1 |
| /nj/ | 18 | /stɹ/ | 6 | /sk/ | 1 |
| /w/ | 17 | /sp/ | 6 | /sn/ | 1 |
| /n/ | 15 | /bɹ/ | 5 | /sw/ | 1 |
| /gɹ/ | 15 | /sm/ | 5 | | |
| /j/ | 15 | /ʧ/ | 4 | | |

As in the noun-like words, there were many other onsets recurring more frequently in the sample than the six phonaesthemes under test. Thirteen onsets recurred more frequently than /gɹ/, which was found in fifteen of the adjectives sampled. Seventeen onsets recurred more frequently than /kl/, which was found in twelve of the adjectives. Nineteen onsets recurred more frequently than /st/ (seven adjectives); twenty-one onsets recurred more frequently than /kɹ/ (six adjectives), and forty-two onsets recurred more frequently than /sl/ (two adjectives). There were no /gl/ adjectives found in the sample, making the frequency of productivity attained by /gl/ in this experiment appear all the more marked. Even though a number of

these onsets were more recurrent than the six studied in the sample of nouns, they were by no means the most frequent in the thousand adjectives. Again, then, it seems unlikely that the frequency with which speakers encounter these onsets might have affected their productivity and preference among respondents. The fact that all six phonaesthemes were used more than twice seems to be attributable to English speakers having perceived these associations in this experiment. This appears to be particularly true of the /gl/-<<light>> phonaestheme.

### 5.2.4) Experiment 4: Word-Ranking; Simplified (Semantic Differential Scales)

#### 5.2.4.1) Methodology

As discussed above (cf. § 5.1.8), I simplified the 'word-ranking' experiment from the pilot study into a test using a Semantic Differential scale for the main study. I implemented twelve questions of this type. In each question, I provided respondents with an audio recording of a single coined word, featuring one of the twelve onset phonaesthemes under study. I ensured that the coined word combined the phonaestheme with a different rime prosody to those used in Experiments 1 and 2, as discussed in § 5.2.2. I then invited respondents to listen to the coined word as many times as they wished, and place it on a five-point Semantic Differential scale (Babbie, 2013:178-9) depending on how much (or how little) they thought it would express a given meaning. At one end of the scale, I gave the meaning associated with the onset in English – i.e. the meaning under test (cf. § 1.1) – while at the other end I gave an antonymous meaning. In an attempt to control for any respondents recognising a pattern – a phenomenon termed 'response bias' – I varied the end of the scale at which I placed each onset's associated meaning (Oppenheim, 1966:206; Du Plooy, 1995:82-83).

For the purpose of clarification, consider the question testing the /kɹ/-<<crookedness>> phonaestheme. I provided respondents of all three language cohorts with an audio recording of the coined word /kɹuːg/, and a five-point Semantic Differential (henceforth 'SD') scale. On this scale, point 1 corresponded to the meaning of <<straightness>>, and point 5 to the meaning of <<crookedness>>; this being the meaning associated with onset /kɹ/ in English, and the meaning under test in the present study. I asked respondents to assign the coined word an SD-scale ranking from 1 to 5, depending on the meaning they thought it would express, given its phonetic structure. A screenshot is included of this question overleaf (see Appendix 18 for screenshots of the entire main study):

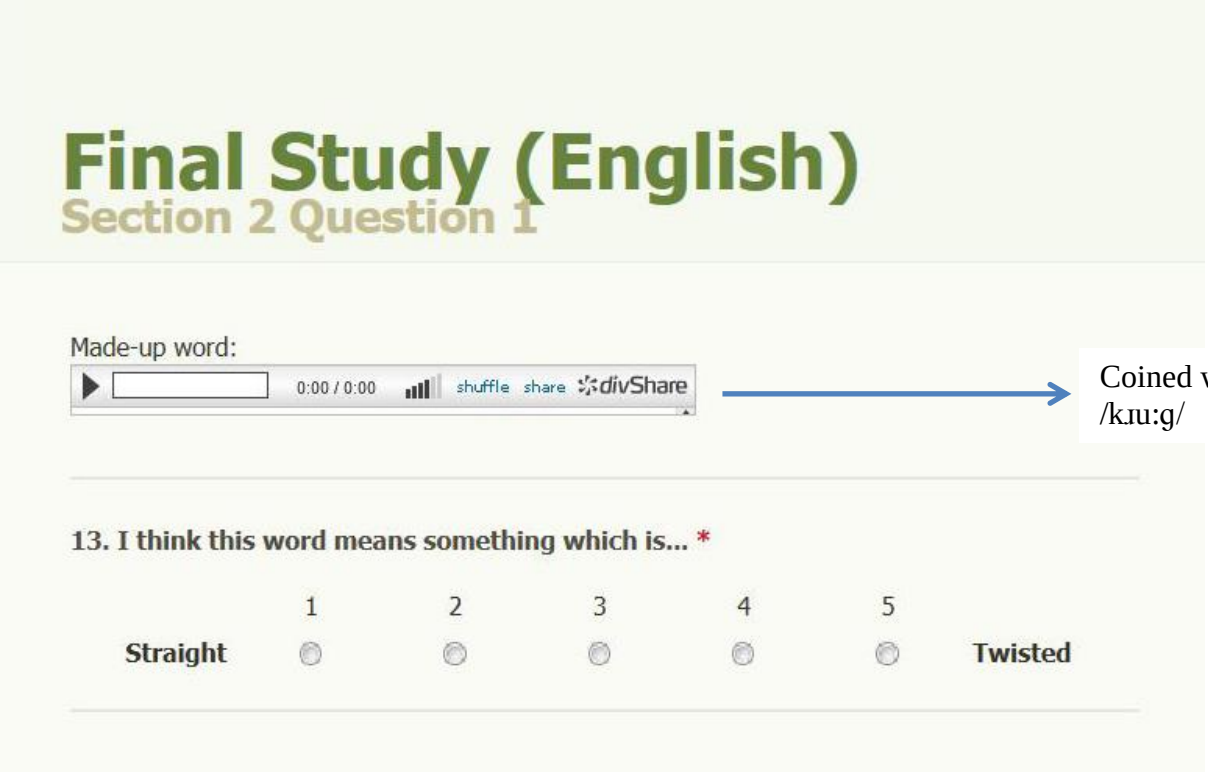



Figure 12: Screenshot of Semantic Differential scale question testing /kɹ/ and <<crookendess; curvilinearity>> in the main study.

I chose to use a five-point Semantic Differential scale for a number of reasons. First, unlike an even-numbered scale (such as a 4 or 6-point version), it offered respondents a noncommittal or "middle" option, in the event that they felt the coined word would not express either of the meanings on the scale (Cohen, Manion and Morrison, 2011:385-388). Further, it offered key affordances over a smaller three-point scale or a larger seven-point scale. Providing five rating dimensions rather than three allowed respondents to select a point biased towards one or other end of the scale without having to be totally committed (i.e. points 2 or 4). At the same time, providing five rating dimensions rather than a higher number was potentially more helpful to respondents. Johns (2010:6) suggests that the use of longer scales in Likert and Semantic Differential questions can cause difficulties for participants, because the differences in meaning between each discrete point on the scale become smaller and less definable as the number of points increases.

My analysis of the data from this experiment involved a number of stages. Taking each question in turn, I began by counting the number of respondents who assigned the coined word to each of the five positions on the scale. That is, I counted the number of speakers who ranked the coined word at position 1, then the number who ranked the coined

word at position 2, and so on through to position 5. I then calculated each of these frequencies as a percentage of the respondent group. For example, if six speakers ranked a given word at position 1 on the scale, this would constitute one fifth (20%) of the respondent cohort. I repeated this process for all twelve onsets in all three languages. These findings are displayed in Table 24, below:

Table 24: The number and percentage of English, French and Polish speakers who chose each point on the SD scale for the twelve phonaesthemes

| ***Phonaesthemes with associated meanings located at point 5.00 on the scale*** | | | | | | |
|---|---|---|---|---|---|---|
| **Phonaestheme** | **Number (%) English speakers choosing each scale point** | | **Number (%) French speakers choosing each scale point** | | **Number (%) Polish speakers choosing each scale point** | |
| **/sn/** | Point 1.00: | 0 (0%) | Point 1.00: | 2 (7%) | Point 1.00: | 1 (3%) |
| | Point 2.00: | 1 (3%) | Point 2.00: | 0 (0%) | Point 2.00: | 3 (10%) |
| | Point 3.00: | 0 (0%) | Point 3.00: | 6 (20%) | Point 3.00: | 4 (13%) |
| | Point 4.00: | 15 (50%) | Point 4.00: | 12 (40%) | Point 4.00: | 13 (43%) |
| | Point 5.00: | 14 (47%) | Point 5.00: | 10 (30%) | Point 5.00: | 9 (30%) |
| **/sm/** | Point 1.00: | 1 (3%) | Point 1.00: | 2 (7%) | Point 1.00: | 3 (10%) |
| | Point 2.00: | 0 (0%) | Point 2.00: | 4 (13%) | Point 2.00: | 3 (10%) |
| | Point 3.00: | 4 (13%) | Point 3.00: | 3 (10%) | Point 3.00: | 4 (13%) |
| | Point 4.00: | 16 (53%) | Point 4.00: | 18 (60%) | Point 4.00: | 13 (43%) |
| | Point 5.00: | 9 (30%) | Point 5.00: | 3 (10%) | Point 5.00: | 7 (23%) |
| **/kɹ/** | Point 1.00: | 0 (0%) | Point 1.00: | 0 (0%) | Point 1.00: | 6 (20%) |
| | Point 2.00: | 3 (10%) | Point 2.00: | 7 (23%) | Point 2.00: | 2 (7%) |
| | Point 3.00: | 3 (10%) | Point 3.00: | 0 (0%) | Point 3.00: | 6 (20%) |
| | Point 4.00: | 16 (53%) | Point 4.00: | 12 (40%) | Point 4.00: | 12 (40%) |
| | Point 5.00: | 8 (27%) | Point 5.00: | 11 (37%) | Point 5.00: | 4 (13%) |
| **/kl/** | Point 1.00: | 0 (0%) | Point 1.00: | 1 (3%) | Point 1.00: | 2 (7%) |
| | Point 2.00: | 3 (10%) | Point 2.00: | 9 (30%) | Point 2.00: | 2 (7%) |
| | Point 3.00: | 3 (10%) | Point 3.00: | 7 (23%) | Point 3.00: | 3 (10%) |
| | Point 4.00: | 18 (60%) | Point 4.00: | 6 (20%) | Point 4.00: | 11 (37%) |
| | Point 5.00: | 6 (20%) | Point 5.00: | 7 (23%) | Point 5.00: | 12 (40%) |
| **/sl/** | Point 1.00: | 0 (0%) | Point 1.00: | 5 (17%) | Point 1.00: | 6 (20%) |
| | Point 2.00: | 3 (10%) | Point 2.00: | 4 (13%) | Point 2.00: | 9 (30%) |
| | Point 3.00: | 5 (17%) | Point 3.00: | 9 (30%) | Point 3.00: | 3 (10%) |
| | Point 4.00: | 15 (50%) | Point 4.00: | 10 (33%) | Point 4.00: | 7 (23%) |
| | Point 5.00: | 7 (23%) | Point 5.00: | 2 (7%) | Point 5.00: | 5 (17%) |

Table Continues Overleaf

| *Phonaesthemes with associated meanings located at point 1.00 on the scale* | | | | | | |
|---|---|---|---|---|---|---|
| **/gɹ/** | Point 1.00: | 3 (10%) | Point 1.00: | 4 (13%) | Point 1.00: | 5 (17%) |
| | Point 2.00: | 13 (43%) | Point 2.00: | 13 (43%) | Point 2.00: | 7 (23%) |
| | Point 3.00: | 7 (23%) | Point 3.00: | 6 (20%) | Point 3.00: | 11 (37%) |
| | Point 4.00: | 6 (20%) | Point 4.00: | 5 (17%) | Point 4.00: | 3 (10%) |
| | Point 5.00: | 1 (3%) | Point 5.00: | 2 (7%) | Point 5.00: | 4 (13%) |
| **/sw/** | Point 1.00: | 11 (37%) | Point 1.00: | 0 (0%) | Point 1.00: | 3 (10%) |
| | Point 2.00: | 11 (37%) | Point 2.00: | 10 (33%) | Point 2.00: | 9 (30%) |
| | Point 3.00: | 4 (13%) | Point 3.00: | 8 (27%) | Point 3.00: | 4 (13%) |
| | Point 4.00: | 2 (7%) | Point 4.00: | 11 (37%) | Point 4.00: | 5 (17%) |
| | Point 5.00: | 2 (7%) | Point 5.00: | 1 (3%) | Point 5.00: | 9 (30%) |
| **/tw/** | Point 1.00: | 9 (30%) | Point 1.00: | 11 (37%) | Point 1.00: | 11 (37%) |
| | Point 2.00: | 16 (53%) | Point 2.00: | 11 (37%) | Point 2.00: | 9 (30%) |
| | Point 3.00: | 4 (13%) | Point 3.00: | 3 (10%) | Point 3.00: | 4 (13%) |
| | Point 4.00: | 1 (3%) | Point 4.00: | 2 (7%) | Point 4.00: | 3 (10%) |
| | Point 5.00: | 0 (0%) | Point 5.00: | 3 (10%) | Point 5.00: | 3 (10%) |
| **/stɹ/** | Point 1.00: | 14 (47%) | Point 1.00: | 12 (40%) | Point 1.00: | 9 (30%) |
| | Point 2.00: | 10 (30%) | Point 2.00: | 8 (27%) | Point 2.00: | 6 (20%) |
| | Point 3.00: | 4 (13%) | Point 3.00: | 3 (10%) | Point 3.00: | 3 (10%) |
| | Point 4.00: | 2 (7%) | Point 4.00: | 5 (17%) | Point 4.00: | 5 (17%) |
| | Point 5.00: | 0 (0%) | Point 5.00: | 2 (7%) | Point 5.00: | 7 (23%) |
| **/skw/** | Point 1.00: | 17 (57%) | Point 1.00: | 1 (3%) | Point 1.00: | 2 (7%) |
| | Point 2.00: | 8 (27%) | Point 2.00: | 6 (20%) | Point 2.00: | 5 (17%) |
| | Point 3.00: | 2 (7%) | Point 3.00: | 3 (10%) | Point 3.00: | 3 (10%) |
| | Point 4.00: | 3 (10%) | Point 4.00: | 11 (37%) | Point 4.00: | 9 (30%) |
| | Point 5.00: | 0 (0%) | Point 5.00: | 9 (30%) | Point 5.00: | 11 (37%) |
| **/st/** | Point 1.00: | 16 (53%) | Point 1.00: | 11 (37%) | Point 1.00: | 9 (30%) |
| | Point 2.00: | 11 (37%) | Point 2.00: | 9 (30%) | Point 2.00: | 11 (37%) |
| | Point 3.00: | 2 (7%) | Point 3.00: | 5 (17%) | Point 3.00: | 4 (13%) |
| | Point 4.00: | 1 (3%) | Point 4.00: | 5 (17%) | Point 4.00: | 4 (13%) |
| | Point 5.00: | 0 (0%) | Point 5.00: | 0 (0%) | Point 5.00: | 2 (7%) |
| **/gl/** | Point 1.00: | 19 (63%) | Point 1.00: | 10 (33%) | Point 1.00: | 16 (53%) |
| | Point 2.00: | 9 (30%) | Point 2.00: | 14 (47%) | Point 2.00: | 7 (23%) |
| | Point 3.00: | 1 (3%) | Point 3.00: | 5 (17%) | Point 3.00: | 1 (3%) |
| | Point 4.00: | 1 (3%) | Point 4.00: | 0 (0%) | Point 4.00: | 3 (10%) |
| | Point 5.00: | 0 (0%) | Point 5.00: | 1 (3%) | Point 5.00: | 3 (10%) |

Whilst compiling the table above, I looked back through Appendix 18, and took note of which end of the scale I had placed the meaning associated with each onset. (This information is included in the table. The first five phonaesthemes are those whose meaning associations I placed at point 5 on the scale; the next seven are those whose meaning associations I placed at point 1.) In analysing the results of this experiment, it was important that I knew whether I had placed each onset's meaning association at point 1 or point 5. This is because my analysis focused on whether the rankings given to each phonaestheme by the three language groups appeared biased or inclined towards particular ends of the SD scale. I

reasoned that if speakers tended to rank a phonaestheme towards one particular end of the scale, it would suggest they had perceived a level of association between the coined word (phonaestheme) and this meaning. To analyse the extent of this ranking 'inclination' in each case, I took each language group and each coined word (phonaestheme) at a time, and using the data in Table 24, calculated:

- The number and proportion of respondents who ranked the coined word at the 'lower' end of the scale (i.e. at points 1 and 2);
- The number and proportion of respondents who ranked the coined in the 'middle' of the scale (i.e. at the median point of 3); and
- The number and proportion of respondents who ranked the coined word at the 'upper' end of the scale (i.e. at points 4 and 5).

These findings are shown in Table 25, overleaf.

Table 25: The number and proportion of respondents in each language who ranked each coined word at the lower end, upper end and median point of the SD scale

| ***Phonaesthemes with associated meanings located at point 5.00 on the scale*** | | | | | | |
|---|---|---|---|---|---|---|
| **Phonaestheme** | **Number (%) English speakers choosing each rank category** | | **Number (%) French speakers choosing each rank category** | | **Number (%) Polish speakers choosing each rank category** | |
| **/sn/** | Lower (1+2): | 1 (3%) | Lower (1+2): | 2 (7%) | Lower (1+2): | 4 (13%) |
| | Median (3): | 0 (0%) | Median (3): | 6 (20%) | Median (3) : | 4 (13%) |
| | Upper (4+5): | 29 (97%) | Upper (4+5): | 22 (73%) | Upper (4+5): | 22 (73%) |
| **/sm/** | Lower (1+2): | 1 (3%) | Lower (1+2): | 6 (20%) | Lower (1+2): | 6 (20%) |
| | Median (3): | 4 (13%) | Median (3): | 3 (10%) | Median (3): | 4 (13%) |
| | Upper (4+5): | 25 (83%) | Upper (4+5): | 21 (70%) | Upper (4+5): | 20 (66%) |
| **/kɹ/** | Lower (1+2): | 3 (10%) | Lower (1+2): | 7 (23%) | Lower (1+2): | 8 (27%) |
| | Median (3): | 3 (10%) | Median (3): | 0 (0%) | Median (3): | 6 (20%) |
| | Upper (4+5): | 24 (80%) | Upper (4+5): | 23 (77%) | Upper (4+5): | 16 (53%) |
| **/kl/** | Lower (1+2): | 3 (10%) | Lower (1+2): | 10 (33%) | Lower (1+2): | 4 (13%) |
| | Median (3): | 3 (10%) | Median (3): | 7 (23%) | Median (3) : | 3 (10%) |
| | Upper (4+5): | 24 (80%) | Upper (4+5): | 13 (43%) | Upper (4+5): | 23 (77%) |
| **/sl/** | Lower (1+2): | 3 (10%) | Lower (1+2): | 9 (30%) | Lower (1+2): | 15 (50%) |
| | Median (3): | 5 (17%) | Median (3): | 9 (30%) | Median (3): | 3 (10%) |
| | Upper (4+5): | 22 (73%) | Upper (4+5): | 12 (40%) | Upper (4+5): | 12 (40%) |
| ***Phonaesthemes with associated meanings located at point 1.00 on the scale*** | | | | | | |
| **/gɹ/** | Lower (1+2): | 16 (53%) | Lower (1+2): | 17 (57%) | Lower (1+2): | 12 (40%) |
| | Median (3): | 7 (23%) | Median (3): | 6 (20%) | Median (3): | 11 (37%) |
| | Upper (4+5): | 7 (23%) | Upper (4+5): | 7 (23%) | Upper (4+5): | 7 (23%) |
| **/sw/** | Lower (1+2): | 22 (73%) | Lower (1+2): | 10 (33%) | Lower (1+2): | 12 (40%) |
| | Median (3): | 4 (13%) | Median (3): | 8 (27%) | Median (3) : | 4 (13%) |
| | Upper (4+5): | 4 (13%) | Upper (4+5): | 12 (40%) | Upper (4+5): | 14 (47%) |
| **/tw/** | Lower (1+2): | 25 (83%) | Lower (1+2): | 22 (73%) | Lower (1+2): | 20 (67%) |
| | Median (3) : | 4 (13%) | Median (3): | 3 (10%) | Median (3): | 4 (13%) |
| | Upper (4+5): | 1 (3%) | Upper (4+5): | 5 (17%) | Upper (4+5): | 6 (20%) |
| **/stɹ/** | Lower (1+2): | 24 (80%) | Lower (1+2): | 20 (67%) | Lower (1+2): | 15 (50%) |
| | Median (3): | 4 (13%) | Median (3): | 3 (10%) | Median (3): | 3 (10%) |
| | Upper (4+5): | 2 (7%) | Upper (4+5): | 7 (23%) | Upper (4+5): | 12 (40%) |
| **/skw/** | Lower (1+2): | 25 (83%) | Lower (1+2): | 7 (23%) | Lower (1+2): | 7 (23%) |
| | Median (3): | 2 (7%) | Median (3): | 3 (10%) | Median (3): | 3 (10%) |
| | Upper (4+5): | 3 (10%) | Upper (4+5): | 20 (67%) | Upper (4+5): | 20 (67%) |
| **/st/** | Lower (1+2): | 27 (90%) | Lower (1+2): | 20 (67%) | Lower (1+2): | 20 (67%) |
| | Median (3): | 2 (7%) | Median (3): | 5 (17%) | Median (3): | 4 (13%) |
| | Upper (4+5): | 1 (3%) | Upper (4+5): | 5 (17%) | Upper (4+5): | 6 (20%) |
| **/gl/** | Lower (1+2): | 28 (93%) | Lower (1+2): | 24 (80%) | Lower (1+2): | 23 (77%) |
| | Median (3): | 1 (3%) | Median (3): | 5 (17%) | Median (3): | 1 (3%) |
| | Upper (4+5): | 1 (3%) | Upper (4+5): | 1 (3%) | Upper (4+5): | 6 (20%) |

Having determined where I had placed the meaning associated with each onset, I was then able to determine the number (and proportion) of respondents who ranked each phonaestheme at a point on the scale **closer than the median point** to the end at which I had placed the meaning under test. (In other words, the number and proportion of respondents who ranked each phonaestheme at a point on the scale that was closer, or 'more inclined', to the meaning under test than to the opposite end of the scale (or exactly in the middle).) I was also able to calculate whether this proportion represented a majority or minority of the respondent cohort in each case. I reasoned that, if the majority of respondents had ranked the coined word closer than the median point (3) to the end of the scale where I had placed the meaning under test, it would suggest that the language group as a whole had perceived the phonaestheme to some extent. By contrast, if the majority of respondents had ranked the coined word either at the median point (3) or closer to the opposite end of the scale, it would suggest that the language group as a whole had not perceived the pattern. Rather, this would suggest that they had not perceived any kind of relationship between the coined word and either of the meanings provided, or had actually associated the onset with an antonymous meaning to the one under test.

I repeated this process for all twelve phonaesthemes in each language cohort. In analysing and interpreting the data, I began by studying the findings from English respondents, in common with Experiments 1-3. For each phonaestheme, I established whether a majority of respondents had ranked the coined word at a position on the scale that was inclined towards its relevant meaning association; and if so, to what extent. In so doing, I discussed what appeared to be the most and least productive phonaesthemes in this experiment (i.e. those ranked towards their meaning associations by the greatest and fewest numbers of speakers); and any notable patterns or trends in the data. I then studied the same in French and Polish, and compared the number of speakers in each cohort who ranked the phonaesthemes at points inclined towards their meaning associations. For any given phonaestheme, I reasoned that the language in which the greatest number of speakers had ranked the coined word towards its meaning association would be the language in which the pattern was most productive, and therefore perceived to the greatest extent.

### 5.2.4.2) Results

Table 26, below, shows the number and percentage of speakers in each language group who ranked each phonaestheme at a point on its scale closer than median (3) to the end at which I had placed its associated meaning.

Table 26: The number of speakers in each cohort ranking each onset towards the meaning under test

| **Phonaestheme** | **Total no. (%) associations with meaning under test** | | |
|---|---|---|---|
| **/sn/** | <<pejoration>> (points 4 or 5 on scale) | | |
| | Eng: 29 (97%) | Fr: 22 (73%) | Pol: 21 (70%) |
| **/sm/** | <<pejoration>> (points 4 or 5 on scale) | | |
| | Eng: 25 (83%) | Fr: 21 (70%) | Pol: 20 (67%) |
| **/kɹ/** | <<crookedness>> (points 4 or 5 on scale) | | |
| | Eng: 24 (80%) | Fr: 23 (77%) | Pol: 16 (53%) |
| **/kl/** | <<clumsiness; thickness; stupidity>> (points 4 or 5 on scale) | | |
| | Eng: 22 (73%) | Fr: 13 (43%) | Pol: 23 (77%) |
| **/sl/** | <<pejoration>> (points 4 or 5 on scale) | | |
| | Eng: 22 (73%) | Fr: 12 (40%) | Pol: 12 (40%) |
| **/gɹ/** | <<gripping>> (points 1 or 2 on scale) | | |
| | Eng: 16 (53%) | Fr: 17 (57%) | Pol: 12 (40%) |
| **/sw/** | <<fast/strong movement>> (points 1 or 2 on scale) | | |
| | Eng: 22 (73%) | Fr: 10 (33%) | Pol: 12 (40%) |
| **/tw/** | <<smallness>> (points 1 or 2 on scale) | | |
| | Eng: 25 (83%) | Fr: 22 (73%) | Pol: 20 (67%) |
| **/stɹ/** | <<straightness>> (points 1 or 2 on scale) | | |
| | Eng: 24 (80%) | Fr: 20 (67%) | Pol: 15 (50%) |
| **/skw/** | <<weakness; cuteness>> (points 1 or 2 on scale) | | |
| | Eng: 25 (83%) | Fr: 7 (23%) | Pol: 7 (23%) |
| **/st/** | <<strength; stoicism>> (points 1 or 2 on scale) | | |
| | Eng: 27 (90%) | Fr: 20 (67%) | Pol: 20 (67%) |
| **/gl/** | <<light>> (points 1 or 2 on scale) | | |
| | Eng: 28 (93%) | Fr: 24 (80%) | Pol: 23 (77%) |

The findings of this experiment appear to offer fairly clear evidence to suggest that all twelve phonaesthemes were productive in English. Every single coined word (onset) was associated with the meaning under test by a majority of respondents. That is to say, a majority of respondents ranked each coined word at a position on the scale that was closer than the median point to the end at which I had placed its meaning association. For a number of phonaesthemes, the number of English speakers who ranked the coined word closest to the relevant end of the scale appears to have been particularly marked. Twenty-nine of the thirty respondents (97%) ranked the /sn/ word at point 4 or 5 on the SD scale, with the meaning <<pejorative>> being labelled at point 5. Similarly, twenty-eight speakers (93%) ranked the /gl/ word at point 1 or 2 on the scale (the meaning <<light>> being labelled at point 1); and twenty-seven speakers (90%) ranked the /st/ word at point 1 or 2 (the meaning <<strength; stoicism>> being labelled at point 1). In general, the level of productivity attained by these phonaesthemes appears to be congruen wih the findings of previous tests. The /gl/-<<light>> phonaestheme was the most productive of the twelve patterns in Experiments 2 and 3, and the second most productive in Experiment 1. Therefore, the fact that twenty-eight speakers ranked the /gl/ coined word at a point on the scale inclined towards the meaning <<light>> appears to support these earlier findings, and does not seem particularly surprising. Similarly, the /st/-<<strength; stoicism>> phonaestheme was the third most productive pattern in Experiments 1 and 2, and the fourth most productive in Experiment 3. Therefore, the fact that twenty-seven speakers ranked the /st/ coined word at a point on the scale inclined towards this meaning does not seem surprising. The only finding which does not appear fully congruent with those of previous tests is the number of speakers who ranked the /sn/ word towards <<pejoration>>. In Experiments 1 and 3, the /sn/-<<pejoration>> phonaestheme was one of the least productive patterns under test; although its productivity rose in Experiment 2. The fact that this phonaestheme appeared relatively productive in two experiments (this test and Experiment 2), but relatively unproductive in Experiments 1 and 3, might suggest further evidence that onset /sn/ could have acquired an association with a particular **type** of <<pejorative>> meaning in English (cf. discussion of Experiment 3 findings, above); rather than the idea of <<pejoration>> at-large. However, at this stage, it is not clear whether this may be the case, or what this type of <<pejorative>> meaning might be. Again, then, the possible differences in connotation between onsets /sl/, /sm/ and /sn/ is a point which could benefit from further research in the future, as I discuss in Chapter 7.

Although the /gl/, /st/ and /sn/ phonaesthemes appeared to attain the highest degree of

productivity in this experiment, these are not the only patterns worthy of comment in the results above. As previously noted, all twelve phonaesthemes were ranked at points on the scale inclined towards the meanings under test by more than half the English speakers sampled. Indeed, with the exception of the /gɹ/-<<gripping>> pattern, every other phonaestheme was actually ranked towards its meaning association by seventy percent of English speakers or more. The extent to which the meanings under test were preferred over the antonymous meanings clearly seems to suggest that English speakers might have perceived these phonaesthemic patterns to some extent, and applied this perception productively. It is not immediately clear why the /gɹ/ phonaestheme appeared relatively less productive in this experiment; this pattern was the most productive association in Experiment 1, and was neither the most nor least productive association in Experiments 2 and 3. Perhaps this result may reflect the fact that <<gripping>> is not the most frequently-recurring meaning among English /gɹ/ headwords (cf. Chapter 4). In Chapter 4, I found that there were roughly three times more /gɹ/headwords in English which expressed something <<pejorative>> than the number that expressed a meaning of <<gripping>>. This could explain why English respondents might not have thought the /gɹ/ coined word sounded like it would express something related to <<gripping>>, yet still attributed onset /gɹ/ to a meaning of <<gripping>> in Experiments 1 and 2; here, <<pejoration>> was not one of the meaning associations available to choose. However, this would not seem to explain why four English speakers chose to coin a /gɹ/ word in response to the <<gripping>> definition in Experiment 3.

In general, the rankings attained by the coined words in French and Polish also appear to support the findings of previous experiments. As previously, it appears that both cohorts perceived associations between a number of the onsets and the meanings under test. In French, eight of the coined words were ranked towards their respective meanings by more than half the respondent cohort. These included the /sn/, /sm/, /kɹ/, /gɹ/, /tw/, /stɹ/, /st/ and /gl/ coined words respectively. In Polish, the same was true for seven of the coined words (including the /sn/, /sm/, /kɹ/, /kl/, /tw/, /st/ and /gl/ words); while the /stɹ/ word was ranked towards its associated meaning of <<straightness>> by exactly half the respondent cohort. At some level, then, this suggests that both cohorts may have perceived the /sn/-<<pejoration>>, /sm/-<<pejoration>>, /kɹ/-<<crookedness>>, /tw/-<<smallness>>, /st/-<<strength; stoicism>> and /gl/-<<light>> phonaesthemes to some extent. French speakers also appear to have perceived the /gɹ/-<<gripping>> and /stɹ/-<<straightness>> patterns; while Polish

speakers also appear to have perceived the /kl/-<<clumsiness; thickness; stupidity>> association. As in English, a number of these findings appear to be particularly marked. For instance, more than three-quarters of the French speakers sampled ranked the /gl/ and /kɹ/ words at points on the scales that were inclined towards their respective meanings of <<light>> and <<crookedness>>. Similarly, more than three-quarters of the Polish speakers sampled ranked the /gl/ and /kl/ words at points on the scales that were inclined towards their respective meanings of <<light>> and <<crookedness>>.

Although the above results suggest that a number of the patterns were perceived by French and Polish speakers, this perception does not appear to have extended across all twelve phonaesthemes, unlike in English. In both French and Polish, there were a number of coined words which were not ranked towards their associated meanings by a majority of respondents; and where the number of speakers associating the coined word with the meaning under test was in the minority. In French, the majority of speakers sampled did not rank the /kl/, /sl/, /sw/ and /skw words at points on the scales that were inclined towards the meanings under test[21]; in Polish, the same was true of the /sl/, /gɹ/, /sw/ and /skw/ coined words[22]. This suggests that neither cohort perceived the /sl/-<<pejoration>>, /sw/-<<fast or strong movement>> and /skw/-<<weakness; cuteness>> phonaesthemes in this experiment; while French speakers did not perceive the /kl/-<<clumsiness; thickness; stupidity>> pattern and Polish speakers did not perceive the /gɹ/-<<gripping>> pattern. What is more, comparing the English, French and Polish results side-by-side (as in Table 26, above) reveals that virtually every single phonaestheme attained a higher number of rankings towards its associated meaning in English than in either other cohort. All twelve patterns were ranked towards their meaning associations by a greater number of English speakers than Polish speakers, and every phonaestheme except /gɹ/-<<gripping>> was ranked towards its meaning association by a greater number of French speakers than English speakers; a finding congruent with Experiment 2, but not Experiments 1 or 3. (As noted above, it is unclear why this phonaestheme appears to have been relatively less productive in English than the other patterns in this experiment; though one explanation could be that <<gripping>> is not the most common meaning among English /gɹ/ headwords.) This is clearly demonstrable when

[21] Namely, <<clumsiness; thickness; stupidity>>, <<pejoration>>, <<fast or strong movement>> and <<weakness; cuteness>>

[22] The meanings under test being <<pejoration>>, <<gripping>>, <<fast or strong movement>> and <<weakness; cuteness>>, respectively.

the results from Table 26 are plotted onto a graph, as in Figure 13, overleaf. In combination, these findings seem to point towards a generally more convincing and more widespread level of perception of the patterns in English than in either other language. Once again, this appears to cohere with the findings of all previous experiments, which all suggest that the phonaesthemes tend to be more productive (and thus, by extension, more widely perceived) in English; although their productivity and perceptibility does not appear to have been confined entirely to English speakers.

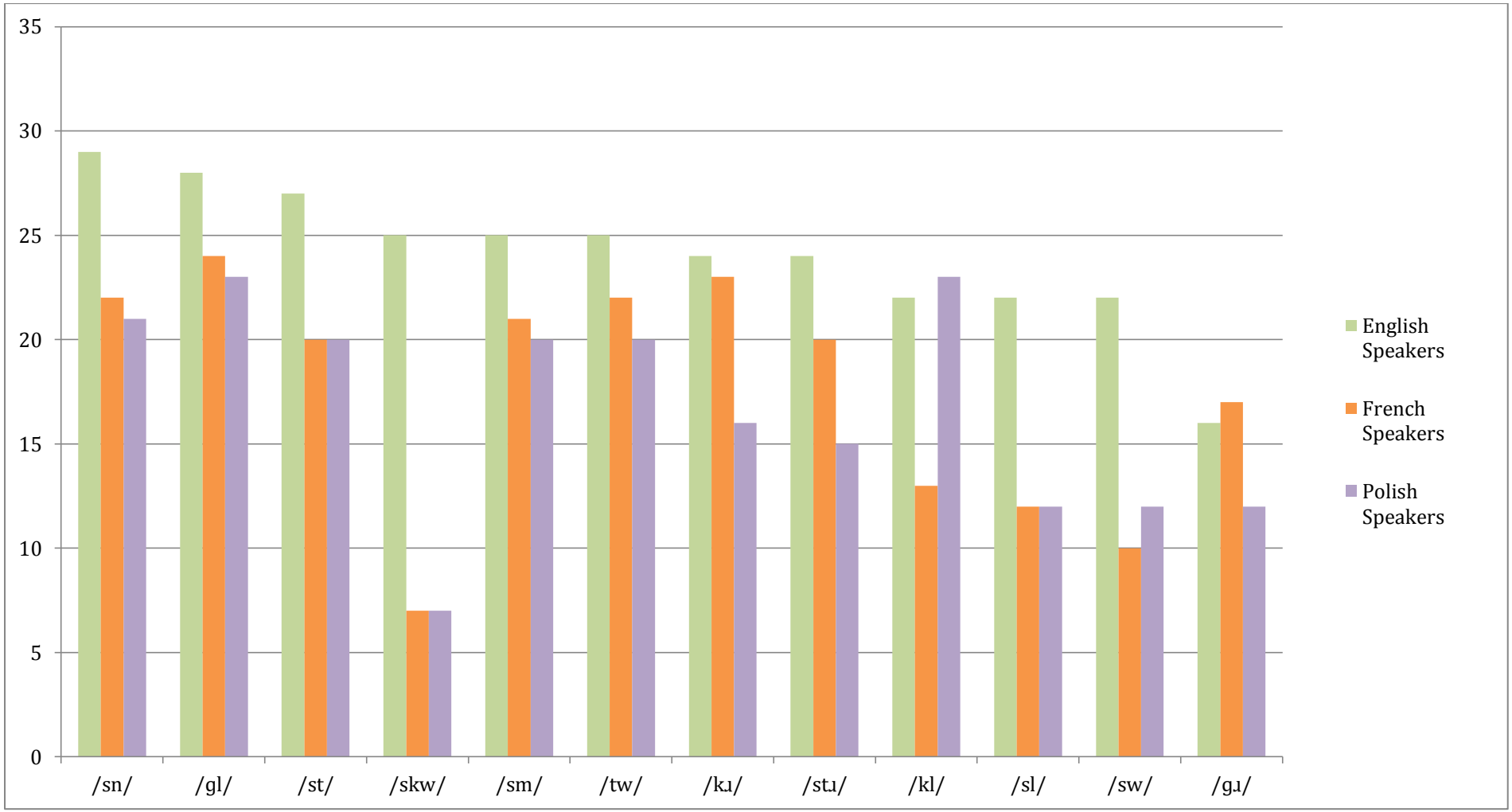



Figure 13: Graph showing the number of speakers in each cohort who ranked each onset at a point on the scale inclined towards the meaning under test

## 5.3) The Total Productivity of each Phonaestheme in English, French and Polish

To fully address Research Questions 3, 4 and 5, it is also necessary to determine what the findings of the four experiments show at-large as well as individually. To this end, this section calculates and discusses the overall productivity of each phonaestheme in English, French and Polish. The aims of this are, first, to determine the most and least productive phonaesthemes in each language across the four experiments; and second, to carry out a macro-level comparison of the twelve phonaesthemes across the three language cohorts.

Following this introduction, this section begins by describing how I calculated the overall productivity of the twelve phonaesthemes in each language. It then analyses the resulting findings, both within individual languages and across the three languages. The section then explains why I chose to supplement the overall comparison of the phonaesthemes in English, French and Polish with statistical testing. It introduces the specific statistical tests I performed, and analyses their findings. In conclusion, the section summarises how the findings of the statistical tests support the interpretations drawn from the figures of overall productivity.

### 5.3.1) Calculating and Comparing the Total Productivity of each Phonaestheme

The overall productivity of each phonaestheme is given in the form of three fractions (one in English, one in French and one in Polish). Each of these expresses the number of times the phonaestheme *was* used productively as a proportion of the number of times it *could have been* used productively. To clarify: In every experiment, each phonaestheme could have been productively associated with its respective meaning by thirty speakers in total. This gives a figure of 120 potential instances of productive use (thirty speakers multiplied by four experiments). I added together the number of times each phonaestheme was used productively across the four experiments, then expressed this number as a fraction (and percentage) of these 120 potential instances. Table 27, below, provides an illustration of this calculation for the /sw/-<<fast/strong movement>> and /sl/-<<pejoration>> phonaesthemes in English. Determining the productive frequencies of each phonaestheme was straightforward in Experiments 1-3, since I had previously recorded this information in Tables 18, 19 and 21, above. For the purpose of Experiment 4, I used the figures from Table

26, which showed the number of speakers awarding each phonaestheme a rank position closer than the median point to the end of the scale where I had placed its associated meaning.

Table 27: Calculating the total productivity of the /sw/-<<fast/strong movement>> and /sl/-<<pejoration>> phonaesthemes in English

| | No. productive associations with <<fast/strong movement>> | | | | |
|---|---|---|---|---|---|
| | Experiment 1 | Experiment 2 | Experiment 3 | Experiment 4 | **Total** |
| Onset /sw/ | 19 | 23 | 9 | 22 | **73** |
| Onset /sl/ | 11 | 22 | 5 + 2 (was used in 2 of the <<pejorative>> definitions) | 22 | **62** |
| **Thus total productivity of /sw/ phonaestheme in English = 73/120 (61%)** | | | | | |
| **Thus total productivity of /sl/ phonaestheme in English = 62/120 (52%)** | | | | | |

Table 28, overleaf, displays the overall figures of productivity for each phonaestheme in English, French and Polish.

Table 28: The total productivity of each phonaestheme in English, French and Polish (ranked by percentage in English)

| ONSET | TOTAL PRODUCTIVE USE BY SPEAKERS (%) | | |
|---|---|---|---|
| | ENGLISH | FRENCH | POLISH |
| /gl/ | 92/120 **(77%)** | 55/120 **(46%)** | 60/120 **(50%)** |
| /st/ | 80/120 **(67%)** | 54/120 **(45%)** | 62/120 **(52%)** |
| /sw/ | 73/120 **(61%)** | 38/120 **(32%)** | 45/120 **(38%)** |
| /kl/ | 69/120 **(58%)** | 41/120 **(34%)** | 46/120 **(38%)** |
| /tw/ | 69/120 **(58%)** | 52/120 **(43%)** | 58/120 **(48%)** |
| /skw/ | 67/120 **(56%)** | 37/120 **(31%)** | 35/120 **(29%)** |
| /gɹ/ | 66/120 **(55%)** | 57/120 **(48%)** | 44/120 **(37%)** |
| /sn/ | 65/120 **(54%)** | 66/120 **(55%)** | 42/120 **(35%)** |
| /sm/ | 65/120 **(54%)** | 66/120 **(55%)** | 50/120 **(42%)** |
| /stɹ/ | 63/120 **(53%)** | 38/120 **(32%)** | 43/120 **(36%)** |
| /sl/ | 62/120 **(52%)** | 44/120 **(37%)** | 43/120 **(36%)** |
| /kɹ/ | 55/120 **(46%)** | 48/120 **(40%)** | 43/120 **(36%)** |

There are a number of trends in this cumulative expression of the findings that are worthy of comment. The first of these is that the /gl/-<<light>> association was clearly the most productive phonaestheme overall among English speakers. This seems unsurprising given the results of the individual experiments. In three of the four experiments (2, 3, 4), onset /gl/ was more productive than any other phonaestheme; and in the remaining experiment (1), it was the second most-productive of the twelve phonaesthemes. In total, onset /gl/ was associated with <<light>> in 92 of the 120 instances in which it was possible for English speakers to have done so. This amounts to a figure of total productivity of 92/120 (77%); higher than any other onset in any other language cohort, and ten percent above the next most-productive phonaestheme overall (/st/ in English).

The second notable trend in the results relates to onset /kɹ/. The /kɹ/-<<crookedness>> phonaestheme attained the lowest overall productivity of any phonaestheme in English. In all, English speakers associated /kɹ/ with <<crookedness>> in 55 of the possible 120 instances where it was possible to do so. This particular phonaestheme is the only association to attain an overall productivity of less than fifty percent in English. This finding provides a clear demonstration of my earlier point that it is difficult to judge the overall productivity of a phonaestheme when analysing the findings of the individual experiments in turn: While /kɹ/ was the least productive phonaestheme in Experiment 1, it did not stand-out as being a particularly unproductive association on any other occasion. However, on looking back at the findings of experiments 1-4, I found that /kɹ/ had almost always been *among* the least-productive phonaesthemes in English. For instance, in Experiment 2, the /kɹ/-<<crookedness>> association was the fourth-least productive (cf. Table 19). Similarly, in Experiment 4, it was the joint third-least productive (cf. Appendix 21). Even in Experiment 3, where the phonaestheme attained its highest level of productivity relative to the other onsets, it was still less productive than /gl/, /sm/, /kl/, /sw/ and /st/ (cf. Table 21). As such, it appears that the /kɹ/-<<crookedness>> association might be a less widely-perceived phonaestheme in English than the others under test; particularly by comparison to the /gl/-<<light>> association. This may or may not reflect the fact that <<crookedness>> is not among the most frequent meanings of /kɹ/ headwords in English (cf. Table 14; see also § 6.1). For this reason, I discuss this phonaestheme in further detail in Chapter 6, with the aim of uncovering other possible factors that could explain this finding.

The third and final trend worthy of note in the English data relates to the /sl/-<<pejoration>> phonaestheme. As shown in Table 28, /sl/ was the second least productive association overall of the twelve patterns under test. This finding is notable, because /sl/ attained the highest type frequency of all twelve phonaesthemes in the vocabulary of English (cf. Table 13). On the strength of this evidence, it appears that the type frequency of a phonaestheme might not directly affect the perceptibility of the pattern within a language. If this was the case, it would seem reasonable to have expected /sl/ to be one of the most productive phonaesthemes overall. This is discussed further in § 5.4, below. I also consider /sl/ in further detail in Chapter 6, alongside /kɹ/, with the aim of uncovering possible explanations for this finding.

Comparing the English, French and Polish data in Table 28 reveals a very striking trend: With the exception of the /sm/ and /sn/ phonaesthemes in French, the overall productivity of every other pattern was greater in English than in either French or Polish. For several phonaesthemes (e.g. /gl/, /sw/, /skw/), the level of overall productivity attained in English is nearly double that seen in either of the other languages. Even the least productive phonaestheme in English (/kɹ/) still attained a higher overall productivity than the majority of the patterns in French and Polish. The extent to which the patterns appear to have been more productive overall in English is demonstrable when the percentage results from Table 28 (above) are plotted onto a graph, as shown by Figure 14 (overleaf). And even though the /sm/ and /sn/ phonaesthemes attained a higher level of overall productivity in French than in English, the difference between the two cohorts is only a single instance out of 120 possible uses (1%). This suggests that these patterns might be similarly productive in both languages, rather than markedly more productive in French. On balance, then, it appears that the majority of the phonaesthemes (that is, the /gl/, /st/, /sw/, /kl/, /tw/, /skw/, /gɹ/, /stɹ/, /sl/ and /kɹ/ phonaesthemes) may be more widely-perceived patterns in English than in French; that the /sm/ and /sn/ phonaesthemes might be similarly perceptible in English and French, and that all twelve of the phonaesthemes may be more widely-perceived in English than in Polish.

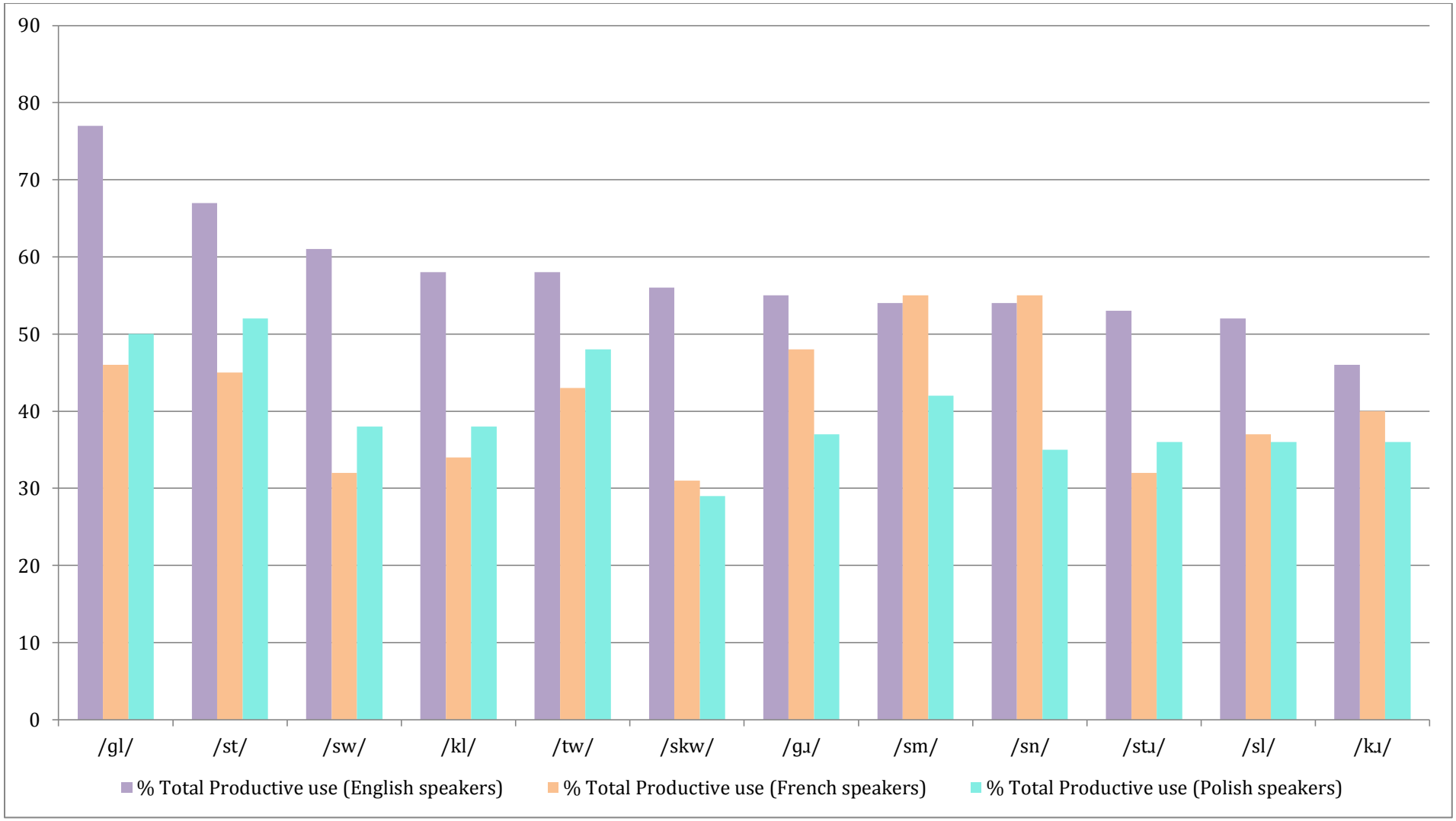



Figure 14: Graph showing the total productivity (%) of each phonaestheme among English, French and Polish speakers

### 5.3.2) Supplementing the Comparison with Statistical Testing

Having established that the phonaesthemes appeared to have been generally more productive (and, by extension, more widely-perceived) in English, I sought to supplement these findings with statistical testing. My aim here was to establish:

- Whether, in any of the individual experiments, there were any **statistically significant differences** between the productivity of the phonaesthemes in English, French and Polish, and if so;
- Whether these significant differences were consistent with the findings of the overall comparison (i.e., whether they were attributable to English speakers having shown greater use of the phonaesthemes than either or both other cohorts).

A 'statistically significant difference' refers to a difference in productivity across two or more languages that was mathematically unlikely to have occurred by chance.

In this analytical stage, I statistically tested the levels of productivity (in English, French, Polish) attained by each phonaestheme in Experiments 1 and 2[23].

#### 5.3.2.1) The Chi-Square test of the Data from Experiments 1 & 2

To compare the productivity of each phonaestheme between English, French and Polish speakers in Experiments 1 and 2, I used a Chi Square test. This test compared the number of respondents in each language who used the phonaestheme productively, and calculated the likelihood that any differences in numbers between the groups could have occurred by chance. In practice, I could only use the Chi-Square to compare the productivity of a phonaestheme between two language groups at any one time. If I had compared all three respondent cohorts simultaneously, the test would not have told me exactly which two cohorts were significantly different; only that there was a difference somewhere in the

[23] I could not perform statistical testing on the data generated by Experiment 3 and 4. In Experiment 3, statistical tests would have relied on the assumption that all possible 'answers' to the question (i.e. all possible word-onsets) had an equal chance of being chosen. As previously noted, this is unlikely to have been the case. In Experiment 4, I could theoretically have conducted a one-way ANOVA to compare the rank positions attained by each phonaestheme. However, it would have been inappropriate to perform this kind of test on ordinal data of this kind, since there is not necessarily an equal or regular difference between each of the points on the scale (cf. Cohen, Manion and Morrison, 2011:605).

findings. Consequently, I had to perform three Chi Square tests for each phonaestheme. The first Chi Square compared English and French respondents' use of the phonaestheme; the second English and Polish, and the third French and Polish. I repeated this process for each of the twelve phonaesthemes in both Experiments 1 and 2.

In all the statistical tests I performed in this investigation, I chose to use a significance level of 0.01. This means I only concluded that the productivity of a phonaestheme was 'significantly different' in two languages if the test yielded a statistic, known as a 'P' value, less than or equal to 0.01. A 'P' value of 0.01 or less ($P \leq 0.01$) indicates that the relationship between two data sets is sufficiently different that it was 99% unlikely to have occurred by chance (i.e. there is only a 1% likelihood that the difference between the two samples could have come about by chance). The SPSS software that I used to perform the calculations offered me a choice of 0.01 or 0.05 significance levels. The latter indicates that the relationship between two data sets is sufficiently different that it was 95% unlikely to have occurred by chance. I used the 0.01 level of significance in preference to the 0.05 level because of its increased strength. The 0.01 significance level offered me the greatest possible certainty that any difference in productivity between speaker cohorts was not likely to have been caused by chance. In the event of finding significant differences, this meant that I could conclude with the greatest possible certainty that the phonaestheme in question was indeed more productive (and by extension, more widely-perceived) in one language than another.

I did not formulate any hypotheses vis-à-vis the outcome of any of my statistical tests. This is because this analytical stage was simply investigative.

#### 5.3.2.3) Chi Square Results for Experiments 1 & 2

Table 29 (overleaf) summarises the results of every Chi Square test on the data from Experiments 1 and 2 through the use of ticks and crosses. A cross identifies that no significant differences were found between the three respondent cohorts, while a tick indicates that a statistically significant difference was found ($P \leq 0.01$). Next to each tick, the table summarises which of the specific groups were significantly different; whether it was the English and French speakers (denoted by 'E-F'), the English and Polish speakers (denoted by 'E-P') or the French and Polish speakers (denoted by 'F-P'):

Table 29: A summary of the findings of the Chi-Square tests

| **PHONAESTHEME** | **WORDS-TO-IMAGE (experiment one)** | **IMAGES-TO-WORD (experiment two)** |
|---|---|---|
| /sl/ | X | X |
| /sm/ | X | X |
| /sn/ | X | X |
| /kl/ | X | X |
| /kɹ/ | X | X |
| /gl/ | ✓ E-F, E-P | ✓ E-F |
| /gɹ/ | ✓ E-F, E-P | X |
| /st/ | X | ✓ E-F |
| /stɹ/ | X | ✓ E-F |
| /sw/ | X | ✓ E-F |
| /skw/ | X | X |
| /tw/ | X | X |

**KEY TO TABLE 29**

✓ = Statistically significant difference ($P \leq 0.01$)
X = No statistically significant difference ($P > 0.01$)
E = English cohort
F = French cohort
P = Polish cohort
\- = Statistical test

(thus '✓ E-F' refers to significant difference found by a statistical test comparing the responses of English and French speakers)

Table 29 reveals that there were comparatively few significant differences in productivity across the three experiments. Nevertheless, there is a clear trend in the results. Every single statistically significant difference occurred between English and another language. Moreover, when I consulted the output data from the tests which had found significant differences, I discovered that every significant difference had been caused by the phonaesthemes being more productive in English than the other language(s). (For the purpose of clarification, the output data from the tests which found significant differences is included as Appendix 22.) That is, on every occasion where a significance (P) value equal to or less than 0.01 was generated, it was because the productivity of the phonaestheme was greater among English speakers than among the non-English cohort(s). This finding strengthens the interpretations made above. Not only do the patterns appear to be generally more productive (and thus more widely-perceived) in English, but every significant difference in their productivity across the three languages is attributable to English speakers making a higher level of association

between the onsets and meanings than non-English speakers.

Having established the total productivity of the twelve phonaesthemes in each language, the next section in this chapter compares the productivity of the phonaesthemes in English with their type and token frequencies, in response to Research Question 5.

## 5.4) Comparing the Type, Token and Productive frequencies of the Phonaesthemes in English

This penultimate section in the present chapter investigates whether the productivity of the phonaesthemes is most affected by their type or token frequencies. This investigation is undertaken in response to Research Question 5. Research Question 5 asked about what sorts of factors might affect the extent to which the twelve phonaesthemes are perceived by speakers. In § 1.4.1, I considered that there could be a number of factors affecting speakers' perception of phonaesthemes; but that two reasonable possibilities could be the number of different words exhibiting a phonaestheme in the vocabulary of a language (type frequency), and the extent to which speakers encounter a phonaestheme in everyday language use (token frequency).

To perform this analysis, I compared the phonaesthemes' type and token frequencies in English (discussed in Chapter 4) alongside their overall productivity among English speakers. As discussed in Chapter 4, I undertook this part of the investigation in English only. I reasoned that if the phonaesthemes' type frequencies were most responsible for their productivity, then the figures of overall productivity should mirror the type frequencies more closely than the token frequencies. Phonaesthemes with higher type frequencies would have tended to attain higher overall productivity, and phonaesthemes with lower type frequencies would have tended to attain lower overall productivity. Conversely, if the token frequencies were most responsible for the phonaesthemes' productivity, then the figures of overall productivity should mirror the token frequencies more closely than the type frequencies. That is, phonaesthemes with higher token frequencies would have tended to attain higher overall productivity, and phonaesthemes with lower token frequencies would have tended to attain lower overall productivity.

Table 30, overleaf, shows the phonaesthemes' type and token frequencies alongside

their overall productivity. The leftmost column shows the phonaesthemes' type frequencies; the middle column shows their token frequencies and the rightmost column shows their overall productivity. Each finding is expressed as both a frequency and percentage. This is because it is most sensible to compare the three findings in percentage terms. It would be very difficult to make a meaningful comparison of the three data sets by frequency alone, because the frequencies represent different proportions in each case. Each type frequency represents a proportion of head lexemes featuring that particular onset in English; and this differs by onset. Each token frequency represents the number of occurrences in a thousand tokens; and each figure of total productivity represents the number of uses in 120 possible instances.

Table 30: Three-way comparison of the type frequency, token frequency and overall productivity of each phonaestheme in English (ranked in descending order of type frequency)

| ONSET | Type Frequency (%) | Token Frequency (%) | Total Productivity (%) |
|---|---|---|---|
| /sl/ | 63 (43%) | 201 (20%) | 62 (52%) |
| /st/ | 53 (13%) | 142 (14%) | 80 (67%) |
| /sn/ | 37 (49%) | 371 (37%) | 65 (54%) |
| /gl/ | 30 (31%) | 459 (46%) | 92 (77%) |
| /sm/ | 26 (38%) | 113 (11%) | 65 (54%) |
| /sw/ | 24 (26%) | 197 (20%) | 73 (61%) |
| /stɹ/ | 22 (19%) | 191 (19%) | 63 (53%) |
| /kɹ/ | 22 (7%) | 134 (13%) | 55 (46%) |
| /kl/ | 16 (9%) | 16 (1%) | 69 (58%) |
| /tw/ | 11 (31%) | 7 (<1%) | 69 (58%) |
| /gɹ/ | 10 (3%) | 23 (2%) | 66 (55%) |
| /skw/ | 7 (14%) | 55 (6%) | 67 (56%) |

Presenting the findings in this way makes it easy to compare the proportional type frequency, token frequency and overall productivity of each individual phonaestheme. For instance, one can see that the /sl/ phonaestheme attained a type frequency of forty-three percent; a token frequency of twenty percent, and a total productivity of fifty-two percent. Similarly, the /sw/ phonaestheme attained a type frequency of twenty-six percent, a token frequency of twenty percent and a total productivity of sixty-one percent. However, this form of presentation makes it more difficult to study patterns in the type, token and productive frequencies across

all twelve onsets (i.e., to compare whether the phonaesthemes attaining higher type frequencies also attained higher token frequencies and/or higher overall productivity, and so on). Therefore, for ease of interpretation, Table 31, below, presents the same information, but ranks each dimension (type, token, overall productivity) from highest to lowest by phonaestheme:

Table 31: The type frequency, token frequency and overall productivity of each phonaestheme in English (each ranked in descending order)

| **Type Frequencies (ranked highest-lowest by phonaestheme)** | **Token Frequencies (ranked highest-lowest by phonaestheme)** | **Total Productivity (ranked highest-lowest by phonaestheme)** |
|---|---|---|
| **/sl/** - 63 (43%) | **/gl/** - 459 (46%) | **/gl/** - 92 (77%) |
| **/st/** - 53 (13%) | **/sn/** - 371 (37%) | **/st/** - 80 (67%) |
| **/sn/** - 37 (49%) | **/sl/** - 201 (20%) | **/sw/** - 73 (61%) |
| **/gl/** - 30 (31%) | **/sw/** - 197 (20%) | **/kl/** - 69 (58%) |
| **/sm/** - 26 (38%) | **/stɹ/** - 191 (19%) | **/tw/** - 69 (58%) |
| **/sw/** - 24 (26%) | **/st/** - 142 (14%) | **/skw/** - 67 (56%) |
| **/stɹ/** - 22 (19%) | **/kɹ/** - 134 (13%) | **/gɹ/** - 66 (55%) |
| **/kɹ/** - 22 (7%) | **/sm/** - 113 (11%) | **/sm/** - 65 (54%) |
| **/kl/** - 16 (9%) | **/skw/** - 55 (6%) | **/sn/** - 65 (54%) |
| **/tw/** - 11 (31%) | **/gɹ/** - 23 (2%) | **/stɹ/** - 63 (53%) |
| **/gɹ/** - 10 (3% ) | **/kl/** - 16 (1%) | **/sl/** - 62 (52%) |
| **/skw/** - 7 (14%) | **/tw/** - 7 (<1%) | **/kɹ/** - 55 (46%) |

This three-way comparison reveals an unexpected but clear trend. Simply, the productivity of the phonaesthemes does not appear to have been directly influenced by either their type *or* token frequencies. Consider first the type frequency data in comparison to the overall productivity. Contrary to what could have been expected, many of the phonaesthemes with higher type frequencies appear to have attained comparatively low overall productivity; while many of the phonaesthemes with lower type frequencies appear to have attained comparatively high overall productivity. For example, the /sl/-<<pejoration>> phonaestheme attained the highest type frequency of all twelve onsets; yet attained the second lowest overall

productivity. Similarly, the /sm/-<<pejoration>> phonaestheme attained the fifth highest type frequency, but attained the joint fourth lowest overall productivity. A similar trend is also evident for /kɹ/. Conversely, the /kl/-<<clumsiness, thickness and stupidity>> phonaestheme attained the fourth lowest type frequency of all twelve patterns, yet attained the fourth highest productivity overall. A similar trend is evident for the /sw/, /skw/ and /tw/ phonaesthemes. Moreover, the phonaestheme that attained the highest overall productivity, /gl/, has only the fourth highest type frequency in the vocabulary. In short, then, there does not appear to be any clear relationship between the phonaesthemes' type frequencies and their productivity. Indeed, the only phonaestheme whose type frequency and overall productivity appears to bear any similarity is onset /st/. The /st/-<<strength and stoicism>> phonaestheme has the second highest type frequency of the twelve onsets, and was the second most-productive overall. However, this pattern appears to be the only clear counterexample to the prevailing trend.

On balance, then, it seems that the perceptibility of the phonaesthemes might not be influenced by the number of [English] words that exhibit the patterns. However, the data also suggests that the perceptibility of the phonaesthemes does not appear to be influenced by their token frequencies either. Once again, many of the results are in opposition to what one might reasonably expect. Many of the phonaesthemes with higher token frequencies attained comparatively low overall productivity; while many of the phonaesthemes with lower token frequencies attained comparatively high overall productivity. For example, the /sl/-<<pejoration>> phonaestheme attained the third highest token frequency but the second lowest overall productivity. Similarly, the /stɹ/-<<straightness and stretching>> phonaestheme attained the fifth highest token frequency but the third lowest overall productivity. The same trend is also evident for onsets /kɹ/ and /sn/. Conversely, the /kl/-<<clumsiness, thickness and stupidity>> phonaestheme attained the second lowest token frequency but the joint fourth highest overall productivity. Likewise, the /tw/-<<smallness>> phonaestheme attained the lowest token frequency but was also the joint fourth most productive overall. A similar trend is evident for onset /st/. Of all twelve phonaesthemes, only the token frequencies of /gl/ and /sw/ appear to display any clear similarity with their overall productivity. The /gl/-<<light>> phonaestheme attained the highest token frequency and the highest overall productivity; while the /sw/-<<fast/strong movement>> phonaestheme attained the fourth highest token frequency and third highest overall productivity. However, these patterns seem to represent the only clear counterexamples to the prevailing trend.

In short, then, there does not appear to be any demonstrable link between a phonaestheme's type *or* token frequency and the extent to which it is used productively. To confirm this interpretation, I used SPSS to conduct a series of Spearman-Rho tests. These measured the degree of correlation between the phonaesthemes' type frequencies, token frequencies and figures of overall productivity, and calculated whether any correlation between these variables was statistically significant. To increase the certainty with which I could draw conclusions, I conducted Spearman-Rho tests on both the percentage data and outright frequency data. Tables 32-35, below, show the results of this enquiry as follows:

- Table 32 shows the correlation between the percentage type frequencies and percentage overall productivity.
- Table 33 shows the correlation between the raw type frequencies and raw figures of overall productivity.
- Table 34 shows the correlation between the percentage token frequencies and percentage overall productivity.
- Table 35 shows the correlation between the raw token frequencies and raw figures of overall productivity.

Table 32: The Correlation between percentage Type Frequencies and percentage Overall Productivity (across all twelve phonaesthemes in English)

**Correlations**

| | | | PercentTypeFreq | PercentOverallProd |
|---|---|---|---|---|
| Spearman's rho | PercentTypeFreq | Correlation Coefficient | 1.000 | -.079 |
| | | Sig. (2-tailed) | . | .807 |
| | | N | 12 | 12 |
| | PercentOverallProd | Correlation Coefficient | -.079 | 1.000 |
| | | Sig. (2-tailed) | .807 | . |
| | | N | 12 | 12 |

Table 33: The Correlation between raw Type Frequencies and raw figures of Overall Productivity (across all twelve phonaesthemes in English)

**Correlations**

| | | | RawTypeFreq | RawOverallProductivity |
|---|---|---|---|---|
| Spearman's rho | RawTypeFreq | Correlation Coefficient | 1.000 | -.018 |
| | | Sig. (2-tailed) | . | .957 |
| | | N | 12 | 12 |
| | RawOverallProductivity | Correlation Coefficient | -.018 | 1.000 |
| | | Sig. (2-tailed) | .957 | . |
| | | N | 12 | 12 |

Table 34: The Correlation between percentage Token Frequencies and percentage Overall Productivity (across all twelve phonaesthemes in English)

**Correlations**

| | | | PercentTokenFreq | PercentOverallProd |
|---|---|---|---|---|
| Spearman's rho | PercentTokenFreq | Correlation Coefficient | 1.000 | .011 |
| | | Sig. (2-tailed) | . | .974 |
| | | N | 12 | 12 |
| | PercentOverallProd | Correlation Coefficient | .011 | 1.000 |
| | | Sig. (2-tailed) | .974 | . |
| | | N | 12 | 12 |

Table 35: The Correlation between raw Token Frequencies and raw figures of Overall Productivity (across all twelve phonaesthemes in English)

**Correlations**

| | | | RawTokenFreq | RawOverallProductivity |
|---|---|---|---|---|
| Spearman's rho | RawTokenFreq | Correlation Coefficient | 1.000 | -.018 |
| | | Sig. (2-tailed) | . | .957 |
| | | N | 12 | 12 |
| | RawOverallProductivity | Correlation Coefficient | -.018 | 1.000 |
| | | Sig. (2-tailed) | .957 | . |
| | | N | 12 | 12 |

In each of the tables 32-35, above, the Significance ('P') value is clearly greater than 0.01 ($P > 0.01$):

- The correlation between the percentage type frequencies and percentage overall productivity yields a 'Sig.' (P) value of 0.807 (Table 32);
- The correlation between the raw type frequencies and raw figures of overall productivity yields a 'Sig.' (P) value of 0.957 (Table 33);
- The correlation between the percentage token frequencies and percentage overall productivity yields a 'Sig.' (P) value of 0.974 (Table 34);
- The correlation between the raw token frequencies and raw figures of overall productivity yields a 'Sig.' (P) value of 0.957 (Table 35).

These findings support my interpretation of Table 31, in showing that there are no statistically-significant correlations between either the type or token frequencies of the phonaesthemes and their overall productivity; either in percentage terms or in outright frequencies.

The fact that there does not appear to be any demonstrable correlation between either the type or token frequencies of the phonaesthemes and their overall productivity, suggests there may be other factors which affect the perceptibility of these patterns besides their presence in the vocabulary of a language and the extent to which they are encountered by speakers. To this end, a further two factors are discussed in Chapter 6. I label these factors "multiple semantic recurrence" and "multiple phonetic recurrence". "Multiple semantic recurrence" refers to the idea that one onset prosody could recur with the expression of many different meanings; while "multiple phonetic recurrence" refers to the idea that one meaning could recur with multiple onset prosodies. Both situations could feasibly affect the extent to which speakers perceive an association between a given onset prosody and one particular meaning. To clarify: If an onset prosody recurred with many different meanings in a language, it might be less prone to acquiring a [connotative] association with any one of these than if it recurred with relatively few meanings. Similarly, if a given meaning recurred with many different onset prosodies, it might be less prone to acquiring a [connotative] association with any one of these than if it recurred with relatively few onset prosodies. Both factors are

systematically tested on five of the phonaesthemes in Chapter 6, to investigate whether they could account for the levels of productivity attained by these patterns.

The next and final section in this chapter summarises how the findings of the productive experiments address Research Questions 3-5.

### 5.5) Summary of Productive Findings

Research Question 3 asked: 'Do native English speakers appear to perceive the twelve phonaesthemes?' The findings of the experiments suggest that all twelve phonaesthemes may have been perceived by English speakers. In experiment 1 (words to image), eight of the onsets were productively associated with their respective meanings by a majority of respondents. These included /gɹ/, /gl/, /st/, /tw/, /skw/, /sw/, /stɹ/ and /kl/. In experiment 2 (images to word), all twelve meaning associations were productively associated with their respective onsets by half the respondent cohort or greater. In experiment 3 (open-ended definitions), all twelve prosodies were used by two or more speakers in response to their associated definitions. Finally, in experiment 4 (SD scale), every single coined word (onset) was associated with the meaning under test by a majority of respondents. That is, a majority of respondents ranked each coined word at a position on the scale that was closer than the median point to the end at which I had placed its meaning association. Overall, eight of the twelve phonaesthemes appear to have productive in every single experiment: the /gl/, /st/, /gɹ/, /tw/, /skw/, /sw/, /stɹ/ and /kl/ phonaesthemes. The /sm/ and /kɹ/ phonaesthemes were productive in three of the four experiments (tests 2, 3 and 4); while the /sl/ and /sn/ phonaesthemes were productive in two of the four experiments (tests 2 and 4). Insofar as productivity is an indicator of perception, it therefore appears that all twelve phonaesthemes may have been perceived to some degree by English speakers. Moreover, the consistency with which the majority of phonaesthemes were productive across the four experiments suggests that the syllable rimes combined with the onsets did not substantially affect the perceptibility of these patterns.

Despite the evidence that all twelve phonaesthemes may have been perceived by English speakers, both the individual experiments and the findings at-large suggest that some phonaesthemes might have been perceived more consistently and to a greater extent than others. Adding together the figures of productivity across all four experiments revealed that

the /gl/-<<light>> phonaestheme was the most productive overall. In total, /gl/ was associated with <<light>> in seventy-seven percent of all instances where this association could have been made. It was the most productive pattern in experiments 2, 3 and 4, and the second most productive in experiment 1. Conversely, the /kɹ/-<<crookedness>> phonaestheme was the least productive phonaestheme overall. In total, /kɹ/ was associated with <<crookedness>> in forty-six percent of instances where this association could have been made; thirty-one percent fewer than /gl/-<<light>>. This phonaestheme was also unfailingly among the least productive patterns across the four experiments. Again, then, insofar as productive use is an indicator of perception, it appears that the /gl/-<<light>> phonaestheme might be a more perceptible pattern in English than the /kɹ/-<<crookedness>> phonaestheme. To some extent, based on the figures of overall productivity, it appears that the /st/ and /sw/ phonaesthemes may also be fairly widely-perceived patterns in English (albeit not as widely-perceived as /gl/); while the /stɹ/, /sl/ and /sm/ phonaesthemes may be somewhat less widely-perceived patterns.

Research Question 4 asked: 'Do native speakers of languages other than English appear to perceive the twelve phonaesthemes?' The findings of the experiments suggest that a number of the phonaesthemes may have been perceived by native speakers of French and Polish (the two languages selected for comparison with English). In experiment 1 (words to image) five of the onsets were productively associated with their respective meanings by a majority of French and Polish speakers. These included /gɹ/, /st/, /sw/ and /tw/ in both languages; as well as /skw/ in French and /stɹ/ in Polish. Similarly, in experiment 2 (images to word), four phonaesthemes were used productively by a majority of each respondent cohort. A majority of French speakers productively associated the meanings <<light>>, <<pejoration>>, <<clumsiness/stupidity>> and <<gripping>> with onsets /gl/, /sl/, /kl/ and /gɹ/; while a majority of Polish speakers productively associated <<light>>, <<strength/stoicism>>, <<pejoration>> and <<smallness>> with onsets /gl/, /st/, /sm/ and /tw/. In experiment 3 (open-ended definitions), seven of the onsets were used twice or more in French respondents' coined words. These included /gl/, /st/, /kɹ/, /gɹ/, /stɹ/, /sl/ and /sn/. Of these, /sl/ and /sn/ were used more frequently than any other onsets. By comparison, four of the onsets were used twice or more in Polish speakers' coined words. These included /st/, /kɹ/, /gɹ/ and /sl/. Of these, /gl/, /gɹ/ and /st/ were used more frequently than any other onsets. Finally, in experiment 4, a majority of French speakers awarded nine of the twelve phonaesthemes mean rank positions that were closer [than the median point] to the end of the

scale at which I had placed the onsets' meaning associations. These included /sl/, /sm/, /sn/, /kl/, /kɹ/, /gl/, /st/, /tw/ and /stɹ/. In Polish, the same was true for eight of these phonaesthemes; the only exception being /stɹ/. In each experiment, then, there was evidence that multiple phonaesthemes were productive (and thus perceived) among both French and Polish speakers.

However, despite this evidence, it appears from both the individual experiments and the findings at-large that the phonaesthemes tended to be generally more widely-perceived in English. This is evident on a number of levels. First, the figures of total productivity indicate that every single phonaestheme was more productive overall in English than Polish; and that ten of the phonaesthemes were more productive overall in English than in French. Only the /sm/-<<pejoration>> and /sn/-<<pejoration>> phonaesthemes attained similar (in this case, marginally higher) levels of productivity in French as in English. Moreover, the findings of the statistical tests (Chi Squares) indicate that every single significant difference between the respondent cohorts was attributable to a higher level of productivity among English speakers. In addition to the above, there appear to have been fewer phonaesthemes productive across all four experiments in French and Polish than there were in English. In English, eight phonaesthemes were productive in all four experiments; with a further two phonaesthemes being productive in three of the four tests. However, in French, no phonaesthemes were productive across all four experiments; the most consistently productive were /st/, gl/, /sl/, and /gɹ/, which were productive in three of the four tests. In Polish, /st/ was productive across all four experiments, but no other phonaestheme was productive in more than two tests. Finally, in all four experiments, a majority of the phonaesthemes appear to have been more productive in English than in either other language. In Experiment 1, seven of the twelve phonaesthemes (/gl/, /gɹ/, kl/, /st/, /tw/, /skw/, /stɹ/) were more productive in English than in either other cohort. Similarly, in Experiment 2, ten of the twelve phonaesthemes (/gl/, /sw/, /sl/, /sn/, /st/, /kl/, /sm/, /kɹ/, /skw/, /stɹ/) attained their highest levels of productivity in English. In Experiment 3, only two of the twelve phonaesthemes (/sl/ and /sn/) attained equal or higher frequencies of use in French or Polish than in English; the remainder were more productive in English. And in Experiment 4, virtually every single phonaestheme attained a higher number of rankings towards its associated meaning in English than in either other cohort. All twelve patterns were ranked towards their meaning associations by a greater number of English speakers than Polish speakers, and every phonaestheme except /gɹ/-<<gripping>> was ranked towards its meaning association by a greater number of French

speakers than English speakers. On balance, then, it appears that the twelve phonaesthemes may have been more widely and more consistently-perceived patterns in English; even though their perception does not appear to have been confined exclusively to English speakers.

Research Question 5, the final question addressed in this thesis, asked: 'What sorts of factors might affect the perception of the twelve phonaesthemes?' In § 1.4.1, I suggested that there could be a number of factors affecting speakers' perception of the phonaesthemes; but that two very likely possibilities could be the phonaesthemes' type and token frequencies. Type frequency refers to the number of different words exhibiting the phonaesthemes in the vocabulary of a language. Token frequency refers to the frequency with which speakers tend to encounter the phonaesthemes in everyday language use. However, while both factors *could* have affected the perceptibility and thus productivity of the phonaesthemes, the data suggests that neither factor appears to have substantially done so – at least in English. Put simply, there did not appear to be any demonstrable link between either the type or token frequencies of the phonaesthemes and the extent to which the patterns were productive. Many of the phonaesthemes with higher type frequencies (such as /sl/, /sm/ and /kɹ/) attained comparatively low overall productivity; while many of the phonaesthemes with lower type frequencies (such as /kl/, /sw/ and /skw/) attained comparatively high overall productivity. Similarly, many of the phonaesthemes with higher token frequencies (such as /sl/, /stɹ/ and /kɹ/) attained comparatively low overall productivity; while many of the phonaesthemes with lower token frequencies (such as /kl/ and /tw/) attained comparatively high overall productivity. These interpretations were supported by a series of Spearman-Rho tests, which showed that there were no significant correlations (either positive or negative) between the phonaesthemes' type or token frequencies and their overall productivity. As a result of these findings, it is not possible to form a complete answer to Research Question 5 at this stage. For this reason, the next chapter discusses two further factors that may have affected the perceptibility and productivity of the phonaesthemes. These factors, termed 'multiple semantic recurrence' and 'multiple phonetic recurrence', were briefly outlined at close of § 5.4. It is to these factors that the discussion now turns; and on this point that the present chapter concludes.

# CHAPTER 6: FURTHER DISCUSSION

This chapter has two goals. The first is to consider two other factors that could affect the perceptibility of the phonaesthemes. In § 5.4, I compared the phonaesthemes' type and token frequencies against their productivity in English, and found there did not appear to be any demonstrable link between either their type or token frequencies and the extent to which the patterns were productive (and thus perceptible). This suggests that there may be other factors which affect the perceptibility of phonaesthemes besides their presence in the vocabulary of a language and the extent to which speakers encounter these patterns. At the close of § 5.4, I proposed a further two possible factors, which I termed "multiple semantic recurrence" and "multiple phonetic recurrence". "Multiple semantic recurrence" refers to the idea that one onset prosody could recur with the expression of many different meanings. "Multiple phonetic recurrence" refers to the idea that the expression of one particular meaning could recur with multiple onset prosodies. Both situations could feasibly affect the extent to which speakers perceive an association between a given onset prosody and one particular meaning. If an onset prosody recurred with many different meanings in a language, it might be less prone to acquiring a [connotative] association with any one of these than if it recurred with relatively few meanings. Similarly, if a given meaning recurred with many different onset prosodies, it might be less prone to acquiring a [connotative] association with any one of these than if it recurred with relatively few onset prosodies. In this chapter, I systematically test both factors on five of the twelve phonaesthemes in English, to investigate whether they could account for the levels of productivity attained by these patterns.

In addition to the above, the second goal of this chapter is to discuss the limitations of the present study. Broadly, these limitations fall into three categories. These include:

- The relative lack of previous research into both the perceptibility of phonaesthemes and the twelve patterns under investigation, and the difficulties this caused for designing my experiments and interpreting their findings;
- The scope of the study (including the time available for the project), which had implications for the numbers of respondents sampled and the specific patterns I chose to study;
- Limitations of the particular methodologies I used to investigate the twelve patterns.

Following this introduction, the present chapter is divided into two subsections. The first subsection begins by discussing the particular phonaesthemes on which I test multiple semantic recurrence and multiple phonetic recurrence; as well as my rationale for choosing to test these particular patterns in further detail. It then goes on to investigate multiple semantic recurrence and multiple phonetic recurrence, and concludes by summarising the findings of these investigations. The second and final subsection of the chapter then presents my critical discussion of the limitations of the present study. In light of these limitations, a number of possible directions for future research are proposed in Chapter 7.

## 6.1) Exploring Multiple Semantic Recurrence and Multiple Phonetic Recurrence

### 6.1.1) The Choice of Phonaesthemes

In this chapter, I test the ideas of multiple semantic recurrence and multiple phonetic recurrence on the /gl/, /sl/, /kɹ/, /stɹ/ and /sw/ phonaesthemes. My rationale for using these particular phonaesthemes was as follows:

- The /gl/-<<light>> phonaestheme: As shown in Chapter 5, this phonaestheme was the most productive of all twelve associations in English. I reasoned that if multiple semantic recurrence and/or multiple phonetic recurrence were factors affecting the perception and productivity of the phonaesthemes, then they might be able to account for the high overall productivity attained by this phonaestheme.
- The /sl/-<<pejoration>> phonaestheme: As discussed in Chapter 5, this phonaestheme was one of the least productive overall in English; despite the fact that its type frequency was the highest of all twelve phonaesthemes, and the fact that this association is one of the most frequently-discussed in the literature. Again, I reasoned that if multiple semantic recurrence and/or multiple phonetic recurrence had affected the perception and productivity of the phonaesthemes, then they might be able to account for the relatively low overall productivity attained by this phonaestheme.
- The /kɹ/-<<crookedness>> phonaestheme: In Chapter 5, I found that this was the least productive phonaestheme in English. In § 5.3.1, I suggested that its low overall productivity might have been attributable to the fact that <<crookedness>> was only the fifth most common semantic field of /kɹ/ headwords in English; that there were

four meanings that recurred more frequently with onset /kɹ/. However, on reflection, this seems unlikely. The meaning associations of the other eleven onsets were either the first or second most common semantic fields among their headwords; yet there were still clear differences in productivity between these phonaesthemes. This suggests that there might have been another factor or other factors responsible for the lower productivity of /kɹ/. To this end, I decided to investigate whether multiple semantic recurrence and/or multiple phonetic recurrence could account for the low overall productivity of this phonaestheme.

- The /stɹ/-<<straightness>> and /sw/-<<fast/strong movement>> phonaesthemes: I included these purely as random samples, to investigate whether multiple semantic recurrence and/or multiple phonetic recurrence were also congruent with the overall productivity of these patterns. The /stɹ/ phonaestheme was less productive overall than the /gl/ and /sw/ phonaesthemes, but was more productive than the /sl/ and /kɹ/ phonaesthemes. The /sw/ phonaestheme was less productive overall than the /gl/ phonaesthemes, but was more productive than the /sl/, /stɹ/ and /kɹ/ phonaesthemes.

### 6.1.2) Multiple Semantic Recurrence

The first of the two factors under test in this chapter is multiple semantic recurrence. As noted above, this term refers to the idea that a single onset prosody could recur with the expression of many different meanings within a language. It is possible that the perceptibility (and thus productivity) of the phonaesthemes could have been affected by the number of different meanings with which their onsets recurred. The reason for this is that, if a prosody recurs with many different meanings, it might be less prone to acquiring a [connotative] association with any one of these than if it recurred with relatively few meanings. Conversely, if a prosody recurs with relatively few different meanings, it might stand more chance of acquiring an association with one of these meanings within the speech community. The plausibility of this scenario is supported by both Usage-Based Phonology (Bybee, 1985; 2001) and Probabilistic Phonology (Pierrehumbert, 2003), discussed in Chapter 3. In Usage-Based terms, some prosodies may be recurrently activated with a greater number of [different] semantic schema than others. In such cases, the sheer number of different schema could reduce the likelihood of speakers learning an association between the prosody and just one particular [semantic] pattern. By contrast, other prosodies may be recurrently activated with a much smaller number of [different] semantic schema; perhaps even only one or two.

With fewer competing schema, speakers might be more likely to learn an association between these prosodies and one particular semantic pattern. Similarly, in Probabilistic terms; if a speaker tended to experience a certain prosody recurring with a small number of different meanings, they would be likely to calculate a high degree of probability that the prosody would be found with one of these meanings in subsequent linguistic encounters. However, if they experienced the prosody recurring with many different meanings, the probability of it recurring with any one of these in the future would be lower.

To test whether the perceptibility and productivity of the aforementioned phonaesthemes might have been affected by multiple semantic recurrence, I first needed to calculate the number of different meanings that recur with these onsets in English. I then needed to compare these figures with the overall productivity attained by these phonaesthemes. If the phonaesthemes had been affected by the number of different meanings recurring with their onsets as outlined above, I predicted that the following trends would be found:

- The phonaesthemes whose onsets recur with fewer different meanings would have tended to attain higher overall productivity;
- The phonaesthemes whose onsets recur with a greater number of different meanings would have tended to attain lower overall productivity.

Testing for evidence of these trends was relatively straightforward. I had already collected data that I could use to calculate the number of different meanings recurring with the five onsets (/gl/, /sw/, /sl/, /stɹ/ and /kɹ/). For this purpose, I re-used the lists of semantic fields that I had compiled for the first vocabulary-based experiment in Chapter 4 (cf. § 4.2). Using this data, I took each phonaestheme in turn, and counted the number of times every discrete semantic field was found in its list of headwords. This allowed me to establish which semantic fields were recurrent (i.e. found more than once) with each of the five onsets. I then simply counted the total number of recurring semantic fields for each onset. The results of this investigation are shown in Table 36, overleaf. (For the purpose of comparison, the lists of semantic fields of the /gl/, /sw/, /sl/, /stɹ/ and /kɹ/ headwords are provided in Appendix 23.)

Table 36: The recurring semantic fields of /gl/, /sw/, /sl/, /stɹ/ and /kɹ/ headwords in English

| Onset prosody | Recurring semantic fields (no. words exhibiting each semantic field) |
|---|---|
| /gl/ | light (30); medical (14); chemicals/chemistry (8); pejoration/depression (8); biology (6); worldliness (6); body/bodily functions (5); smoothness (5); adhesion/stickiness (3); clothing (3); ice/coldness (3); movement (3); nature (3); political (3); smallness (3); transparency (3) abrasion (2); construction (2); delicateness (2); fauna (2); glutinous consistency (2); storage (2); tools (2) |
| | **Total no. recurring semantic fields for /gl/: 23** |
| /sl/ | pejoration (63); shape (11); transport (9); bodily functions (8); food/drink/consumption (7); smallness (7); nationality/ethnicity (6); nature (6); speed (6); clothing (5); construction (5); liquidity (5); medical (5); fauna (4); geology (4); weather (4); ease/efficiency (3); mining (3); nauticalia (3); skill/dexterity (3); alcohol (2); architecture (2); colour (2); gap/space (2); measurement (2); photography (2); slowness (2); soil (2); sports (2); topography/hills (2) |
| | **Total no. recurring semantic fields for /sl/: 30** |
| /stɹ/ | straightness/stretching/linearity (23); pejoration (11); constriction/restriction (9); thinness (8); music (7); food/drink/consumption (5); marks/impressions (5); morality/immorality (5); medical (5); work (5); clothing (4); construction (4); layers/layering (4); liquidity (4); light (4); security (4); classification (3); control (3) effort (3); flowing (3); geology (3); prominence (3); removal (3); weather (3); appearance (2); bacteria (2); biology (2); chemicals/chemistry (2); cloudiness (2); colour (2); intermittency (2); isolation (2); murder (2); modern (2); neck/necks (2); physical science (2); protesting (2); roads/transport (2); Scotland (2); skill/dexterity (2); striking/hitting (2); urban (2); wisdom (2) |
| | **Total no. recurring semantic fields for /stɹ/: 43** |
| /sw/ | fast/strong movement (26); food/drink/consumption (14); nature (11); clothing (8); fauna (8); pejoration (8); ethnicity/identity (7); nostalgia (7); liquidity (6); warmth/heat (6); fabric (5); taste (5); electrical (4); flight (4); propel/propulsion (4); protection/safety/security (4); shape (4); accessibility (3); change/difference (3); love/romance (3); medical (3); sport/exercise (3); summertime (3); swan (3); wrap/enclose (3); bog/marsh (2); colour (2); flamboyance (2); luxury/opulence (2); pleasantness/pleasing (2); possessions (2); power (2); profanity (2); sound/noise (2); transport (2) |
| | **Total no. recurring semantic fields for /sw/: 35** |

Table Continues Overleaf

| Onset prosody | Recurring semantic fields (no. words exhibiting each semantic field) |
|---|---|
| /kɹ/ | nature (46); food/drink/consumption (37); pejoration (34); flora (24); crookedness/curvedness (22); fauna (21); construction (19); material/fabric (17); medical (17); commerce/finance (13); geology/geography/topography (13); sound/noise (12); emotion (11); science (9); breeding/cultivation (8); farming (8); shape (7); sport (7); clothing (6); dice/gaming (6); legality/illegality (6); machinery (6); motion/movement (6); nationality/ethnicity (6); negativity (6); religion/religious accoutrements (6); skill/dexterity (6); transport (5); anatomy/body parts (4); death (4); destruction/damage (4); gap/break (4); health (4); lifestyle (4); royalty (4); safety (4); speed (4); stiffness/strength (4); academia (3); chemicals (3); colour (3); crystalline (3); dairy (3); drawing (3); eccentricity (3); fashion (3); history/historical (3); judging/judgments (3) morality (3); nautical (3); temperature (3); transparency (3); alcohol (2); appealing (2); belief (2); biology (2); brain (2); building (2); burning (2); collision (2); compression (2); container (2); credibility (2); crockery (2); cutting (2); equestrian (2); existence (2); furniture (2); gambling (2); glass (2); humanity (2); industry (2); language (2); light (2); mental health (2); oil (2); paper (2); patterns (2); physique (2); political (2); sight (2); smallness (2); stationery (2); storage (2); stupidity (2); support/structure (2); trade (2); travel (2); warfare (2); water (2); weapons (2); woman/female (2); zoology (2) |
| | **Total no. recurring semantic fields for /kɹ/: 93** |

Having established the number of recurring semantic fields of the five onsets, Table 37, below, compares these figures to their total productivity.

Table 37: Comparing the number of recurring semantic fields of onsets /gl/, /sw/, /sl/, /stɹ/ and /kɹ/ with the overall productivity of these phonaesthemes

| Onset Prosody (phonaesthemic association) | No. recurring semantic fields in English | Overall Productivity of phonaestheme (%) |
|---|---|---|
| /gl/ (light) | 23 | 77 |
| /sl/ (pejoration) | 30 | 52 |
| /sw/ (fast/strong movement) | 35 | 61 |
| /stɹ/ (straightness) | 43 | 53 |
| /kɹ/ (crookedness) | 93 | 46 |

The data in Tables 36 and 37, above, suggests that the perceptibility and productivity of the phonaesthemes might have been affected to some degree by multiple semantic recurrence. In the majority of cases, the phonaesthemes whose onsets recur with fewer different meanings tended to attain higher overall productivity; while the phonaesthemes whose onsets recur with a greater number of different meanings tended to attain lower overall productivity. For instance:

- Onset /kɹ/ recurs with the greatest number of different meanings (semantic fields) – ninety-three in total – and attained the lowest overall productivity.
- Onset /stɹ/ recurs with fewer different meanings than /kɹ/ (forty-three in total), but a greater number of different meanings than /sl/. As predicted, the overall productivity attained by this phonaestheme was higher than /kɹ/ but lower than /sl/.
- Onset /gl/ recurs with the fewest number of different meanings. As noted in Chapter 5 and as is clearly shown in Table 37, this phonaestheme attained the highest overall productivity.

In general, then, it appears that the perception and productivity of the phonaesthemes may have been affected to some degree by multiple semantic recurrence. It would appear that phonaesthemes such as /gl/, which recurs with fewer different meanings, might have had more chance to become associated with one particular meaning in the speech community than phonaesthemes like /kɹ/, which recurs with a greater number of different meanings. The general trend is shown when the data from Table 37 is plotted onto a graph (cf. Figure 15, overleaf). The graph shows that, *in general*, as the number of different meanings increases, the overall productivity of the phonaestheme tends to decrease:

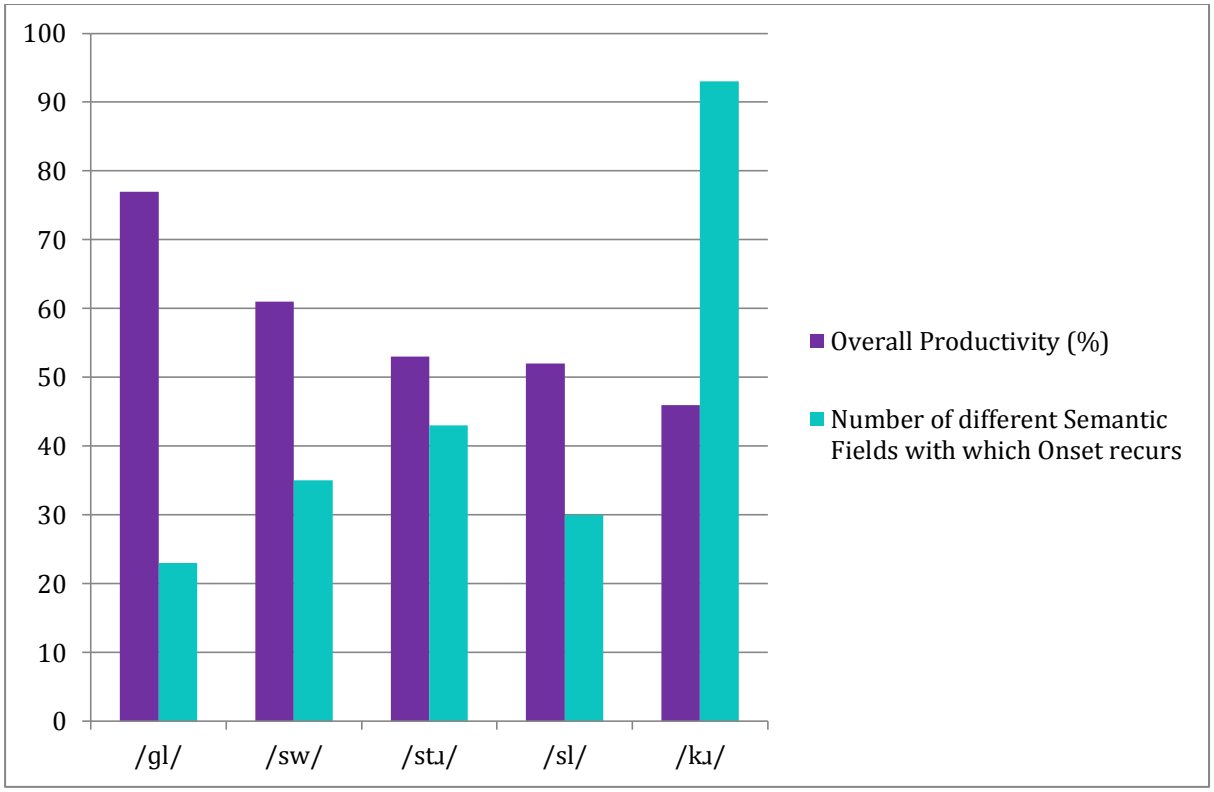



Figure 15: Graph showing the overall productivity of the /gl/, /sw/, /sl/, /stɹ/ and /kɹ/ phonaesthemes, and the number of different semantic fields recurring with each of these onsets in English

There are, however, two counterexamples to this trend: the /sl/ and /sw/ phonaesthemes. Onset /sl/ was the second least-productive of the five onsets; attaining an overall figure of productivity of 52% - higher than the /kɹ/ phonaesthemes but lower than the /gl/, /sw/ and /stɹ/ phonaesthemes. To have been fully compatible with the dominant trend, /sl/ would have needed to recur with less than ninety-three different semantic fields, but more than forty-three; the frequencies attained by the least productive phonaestheme (/kɹ/) and next most productive phonaestheme (/stɹ/). However, Table 37 and Figure 15 both show that this was not the case: I found that onset /sl/ recurred with thirty different semantic fields – some thirteen fewer than /stɹ/– even though it was less productive than /stɹ/ overall. If multiple semantic recurrence had been the only factor influencing the productivity of the phonaesthemes, then theoretically the /sl/ phonaestheme should have attained a higher overall productivity than the /stɹ/ phonaestheme, since it recurred with fewer [different] semantic fields. A similar scenario is seen for onset /sw/. The /sw/-<<fast or strong movement>> phonaestheme was the second most-productive of the five patterns; attaining an overall figure of productivity of 61% - higher than the /stɹ/, /sl/ and /kɹ/ phonaesthemes but lower than the

/gl/ phonaestheme. To have been fully compatible with the overall trend, /sw/ would have needed to recur with more than twenty-three different semantic fields but fewer than thirty; twenty-three being the number of semantic fields associated with /gl/, and thirty being the number recurring with the [much less productive] onset /sl/. However, Table 37 and Figure 15 both show that this was not the case; onset /sw/ actually recurs with five more semantic fields than onset /sl/, even though it was more productive than /sl/overall.

On balance, then, while multiple semantic recurrence may have been one factor affecting speakers' perception of the phonaesthemes, it does not appear to have been the only factor responsible for the productivity they attained. To this end, the discussion now turns to the second factor outlined for investigation in this chapter; the idea labelled 'multiple phonetic recurrence'.

### 6.1.3) Multiple Phonetic Recurrence

The second factor under test in this chapter is multiple phonetic recurrence. As previously noted, this term refers to the idea that a particular meaning could recur with many different onset prosodies within a language. It is possible that the perceptibility (and thus productivity) of the phonaesthemes could have been affected by the number of different prosodies with which the twelve meaning associations recurred. The reasoning behind this is the same as for multiple semantic recurrence: If a meaning recurs with many different [onset] prosodies, it might be less prone to acquiring a connotative association with any one of these than if it recurred with relatively few prosodies. Conversely, if a meaning recurs with relatively few onset prosodies, it might stand more chance of acquiring an association with one of these prosodies within the speech community. Once again, the plausibility of this scenario is supported by both Usage-Based (Bybee, 1985; 2001) and Probabilistic (Pierrehumbert, 2003) models. In Usage-Based terms, one particular meaning may be recurrently activated with a greater number of [different] phonetic schema than other meanings. In such cases, the number of different schema could reduce the likelihood of speakers learning an association between this meaning and just one phonetic pattern. By contrast, another meaning may be recurrently activated with a much smaller number of [different] phonetic schema. With fewer competing schema, speakers might be more likely to learn an association between this meaning and one particular phonetic pattern. Similarly, in Probabilistic terms; if a speaker tended to experience a certain meaning recurring with a small

number of different prosodies, they would be likely to calculate a relatively high degree of probability that the meaning would co-occur with one of these prosodies in subsequent linguistic encounters. However, if they experienced the same meaning recurring with many different prosodies, the assumption of probability that it would co-occur with any one of these in the future would be lower.

To test whether the perceptibility and productivity of the five phonaesthemes might have been affected by multiple phonetic recurrence, I needed to calculate the number of different onsets recurring with each of their meaning associations [in English]. This involved using a thesaurus. For this purpose, I used Roget's thesaurus rather than a conventional type, because Roget's version lists items conceptually rather than alphabetically. For example, it lists all items related to the theme of <<light>> together, and all items related to <<straightness>> together, and so on. My idea was to use the thesaurus to look-up the meaning associations of the five phonaesthemes (<<crookedness>>, <<light>>, <<straightness>>, <<fast/strong movement>> and <<pejoration>>); make a list of all the words related to each concept, and sort these lists by onset. Using these lists, I could then count-up how many different onsets (excluding zero onsets) recurred with each of the five meanings, and compare these figures with the overall productivity of the phonaesthemes. This would reveal whether the phonaesthemes attaining higher overall productivity tended to be the ones whose meanings recurred with fewer different onsets.

In practice, this method worked well for the /gl/, /stɹ/, /sl/ and /kɹ/ phonaesthemes. However, I realised early-on that it was not suitable for investigating multiple phonetic recurrence in relation to the /sl/ phonaestheme. The problem here was that I could not just look for the concepts <<pejoration>> or <<unpleasantness>> and list the words related to these ideas. This would only have sampled words related to the idea of unpleasantness or pejoration itself (such as 'nasty', 'horrible', 'contemptuous', and so on); rather than all words expressing culturally pejorative or unpleasant meanings. To remedy this, I compiled a list of all the <<pejorative>> words I encountered in everyday language use over a period of six weeks. Each time I heard a word that I felt was broadly <<pejorative>>, I typed it into a dedicated list on my cell phone. Using this method, I recorded words from a variety of contexts; including the mass media, online communication and in literature. For the purpose

of this exercise, I used the same operationalization of <<pejoration>> that I had applied when studying the dictionary headwords in Chapter 4[24]. Having compiled this list, I then followed the same procedure as for the other phonaesthemes; I sorted the list by onset and counted the number of discrete word-onsets in the sample. Although this method was not ideal, and may have caused me to overlook some <<pejorative>> words, it was still preferable to having only sampled words related to the **idea** of <<pejoration>>.

The results of this investigation shown in Tables 38-42, below. These tables are organised as follows:

- Table 38 shows the different onsets recurring with the meaning <<light>> in English, and examples of words featuring each of these onsets.
- Table 39 shows the different onsets recurring with <<fast/strong movement>> in English, and examples of words featuring each onset.
- Table 40 shows the same for <<straightness/stretching>>.
- Table 41 shows the same for <<crookedness/curvedness>>.
- Table 42 shows the same from my collection of <<pejorative>> words.

Following this, Table 43 then compares the number of onsets recurring with each meaning against the overall productivity of their respective phonaesthemes.

[24] Specifically, "any action, quality, person or thing which can cause any type of harm or bring about an effect that is culturally perceived as disadvantageous to anyone or anything; any expression which can in any sense be used to offend, criticise, insult or degrade; any action, quality, person or thing which is in any sense culturally perceived as 'unpleasant' at the time of writing".

Table 38: Onsets found to recur with the expression of <<light>> in English (cf. Roget, 2004:163-165)

| **Onset** | **English words related to <<light>> that feature this onset** |
|---|---|
| /b/ | beam, burnish(ed), burn, bedazzle(d), bedim, becloud, bolt, bulb, beacon |
| /bl/ | blaze, blinding, blushing, blink, black(ness), blur(red), bleary, blackout |
| /bɹ/ | bright(ness), brilliance |
| /d/ | daylight, dawn, diffraction, dispersion, dance, dazzle, dark, dim(mer), dusk(y), dull(ness), definition, dingy, dip, dun |
| /dʒ/ | jet, jet-black |
| /f/ | fire(light), fulgent, phosphoresce(nt), fair, fade(out), faint(ness), fuzzy(ness), film(y), fog(gy), photo(graph/luminosity/sensitive), firefly |
| /fl/ | flare(ing), flame(ing), flicker(ing), flush(ing), flash(ing), flood(lit), fluoresce(nt), flat |
| /g/ | gaslight, ghostly, garish, gay |
| /gl/ | gloss, glint, glimmer, glisten, gleam, glister, gloaming, glitter, glow, glass(y), glare, glimpse |
| /k/ | chiaroscuro, colour, contrast, kindle, candle |
| /kl/ | clear, cloud, clean |
| /h/ | halo, heat, heavens, haze(y) |
| /l/ | light, luminous/luminance/luminosity, lightning, lamp, lambent, lucid, lurid, lustre(ous), lantern, lighthouse, limelight, lacklustre, leaden |
| /m/ | moon(light), morning, magnitude, matt, misty, muddy, meteor, moonless |
| /n/ | noon, naked, nimbus, nocturnal, neon |
| /ɹ/ | resplendent, refulgent, rainbow, ray, radiant, refract(ion), reflect(ion), radioactivity, rise |
| /p/ | polish(ed), pale(ness), polarize(d), penumbra |
| /s/ | sunlight, soft, sun(ny), sunset, sundown, sunrise, sombre, silhouette, sunbeam, scintillating, searchlight, signal |
| /ʃ/ | sheen, shine, shaft, shimmer(ing), chatoyant, shoot(-out), shade, shed, shadow |
| /spl/ | splendour/splendid |
| /st/ | star(light), starless |
| /sp/ | spectrum, spark(ling) |
| /w/ | wax, wane, wan |
| **Total no. onsets recurring with <<light>> (excluding zero onset): 23** | |

Table 39: Onsets found to recur with the expression of <<fast/strong movement>> in English (cf. Roget, 2004: 110-111)

| **Onset** | **English words related to <<fast/strong movement>> that feature this onset** |
|---|---|
| /b/ | bolt, burst, bound, burn(-up), bowl |
| /d/ | dive, dash, dart, dartle |
| /fl/ | flash, fly, flit |
| /h/ | hasten, hurtle, hare, hurry, hustle |
| /k/ | career, careen, cut |
| /l/ | lunge, lively, lope, leg(-it), lick, leap |
| /p/ | pelt, pound, pounce |
| /ɹ/ | romp, run, rush, rip, race, rattle |
| /sk/ | scamper, scorch, scud, skim, skirr, scurry, skelter, scuttle, scoot, skedaddle |
| /sp/ | spurt, speed, spank, spur |
| /spɹ/ | sprint, spring |
| /st/ | start, steam, storm, stampede |
| /stɹ/ | stream, streak, stride |
| /sw/ | swoop, swoosh, sweep, swipe, swift |
| /ʃ/ | shoot, shift |
| /tɹ/ | trample, tread |
| /tʃ/ | charge, chase |
| /θɹ/ | thrust, thrash |
| /w/ | whizz, whoosh, whirl, whisk |
| /z/ | zip, zoom, zing, zap |
| **Total no. onsets recurring with <<fast/strong movement>> (excluding zero onset): 20** | |

Table 40: Onsets found to recur with the expression of <<straightness and stretching>> in English (cf. Roget, 2004: 82; 97)

| **Onset** | **English words related to <<straightness>> and <<stretching>> that feature this onset** |
|---|---|
| /k/ | Continue, continual, continuing |
| /d/ | Direct, disperse(d) |
| /dɹ/ | Drawn-out; drawl(ing) |
| /f/ | Flat, flattened |
| /l/ | Linear, linear, long, lengthy, lanky, leggy, level, longitudinal |
| /p/ | Perpendicular, pulled-out |
| /pɹ/ | Protract(ed), prolong(ed) |
| /ɹ/ | Rectilinear, rectilineal, rigid, right |
| /stɹ/ | Straight, straight-lined, straightness, stretch, stretching, stretched, stretched-out, string, string/strung-out |
| /spɹ/ | Spreadeagle, spread-out, sprawl(ed), sprawling |
| **Total no. onsets recurring with <<straightness and stretching>> (excluding zero onset): 10** | |

Table 41: Onsets found to recur with the expression of [literal and metaphorical] <<crookedness>> in English (cf. Roget, 2004: 96-97; 288-289; 399)

| **Onset** | **English words related to <<crookedness>> that feature this onset** |
|---|---|
| /b/ | bent, bowed, biased, buckled, bandy, batter(ed), botch(ed), bony, bifurcate(d), billowy, busty, |
| /d/ | disproportion(ate/al), disfigured, defective, deform(ed), deviate(d), domed, deceive(d), disingenuous, dishonest, devious, dark, doubtful, dubious, dishonourable |
| /dʒ/ | jagged, jointed, juggle(r) |
| /f/ | fold(ed), forked, fishy |
| /fɹ/ | fractal, fraudulent |
| /h/ | hunched, hook(ed), hypocritical |
| /k/ | contorted/contortion, convoluted, cock-eyed, curved, corner(ed), concave, convex, curled, conical, kinked, cunning, contrive(d), cagey |
| /kɹ/ | crooked, crazy, crinkle(d), crumple(d), crotchet, crook, cranny, crescent, crafty, creep(ing) |
| /l/ | lopsided, looped, lying |
| /m/ | malformed, mutated, mangled, misshapen, Machiavellism |
| /n/ | gnarled, knotted, nook, notched |
| /p/ | perverse/perverted, parabolic |
| /pɹ/ | project(ing), protuberant/protuberance, prevaricate |
| /ɹ/ | wry, wrest, recess(ed), resourceful, wrong |
| /s/ | scythed, serrated, sinuous, circumflex, subtle, serpent, sinister, suspicious |
| /sk/ | skew(ed), scalene, skilled, scheme(r) |
| /sl/ | sly, sleight, slick |
| /st/ | stagger(ed), stealth(y) |
| /ʃ/ | chevron, sharp, shifty, shameless, shady |
| /t/ | turn(ed), tortuous |
| /tɹ/ | trick(ed), trickster |
| /tw/ | twisted, twined |
| /w/ | warped, wavy, wily, wit, wangle, winding |
| **Total no. onsets recurring with <<crookedness>> (excluding zero onset): 23** | |

Table 42: A selection of <<pejorative>> English words and their onset prosodies

| **Onset** | **Examples of <<pejorative>> English words featuring this onset** |
|---|---|
| /b/ | barf, bum, buggery, bitch, bomb, base, bigoted, burst |
| /bl/ | blowout, blot, blaggard, bloody, blaspheme |
| /bɹ/ | brainless, brigand, brat, brawl, brute |
| /k/ | conniving, cancer, chaos, kidnap, cock, kill, cowardice, cad |
| /kl/ | clout, clingy, clot, clink, clutter |
| /kɹ/ | crass, crap, cranky, crybaby, crummy, cruel, crook, crony, crud, crucify |
| /d/ | death, degrade, devil, damnation, dastardly, deviant, demeaning |
| /dɹ/ | drunk, drag, dreadful, dreary, drudge, dropout |
| /f/ | feeble, faggot, fart, fat, foist, fug, fiend, fuck |
| /fl/ | flea, flabby, flaming, flatulent, floozy |
| /fɹ/ | frump, frigid, freak, freeloader, frighten |
| /g/ | gimp, goon, ghastly, gag, gammy, garbage, gargoyle, ghetto |
| /gl/ | glib, gloat, gloop |
| /gɹ/ | grim, gruesome, grime, grizzly, grimace |
| /h/ | horrible, hideous, harsh, hijack, heinous, hang, halfwit, whore, hell, homicide |
| /dʒ/ | jerk, jackass, jealousy, jab, jail, jargon, junk |
| /l/ | lunatic, layabout, loser, leer, lecherous, licentious |
| /m/ | mean, miser, moronic, malevolent, monster, munter, malfunction, misery, murder |
| /n/ | niggardly, noxious, nosy, nasty, nincompoop, numpty |
| /p/ | pig, poofter, pansy, poo, perverse, paedophile, pedant, pain, pungent |
| /pl/ | plebeian, plagiarise, plaque, plod |
| /pɹ/ | prissy, prig, prude, prat, prate |
| /ɹ/ | rude, rot, runt, remedial, ruin, rubbish, rat, ratbag |
| /s/ | cellulite, suffer, sucker, sodomy, sabotage, sick |
| /sk/ | scum, skank, skunk, sket, scoff, scag, scold, scant, scourge, skimp, scat, scar |
| /skɹ/ | scrawny, screw, scram, scrap, scream, scrape, scratch |
| /sl/ | slime, sleaze, slap, sly, slip, slump, slum, slut, sloven, slag, slapper, slog, slither |
| /sm/ | smirk, smug, smirch, smoke, smarmy, smelly, smack, smear, smite, smog |
| /sn/ | snoop, sniper, snake, snag, snap, snicker, snide, sneer, sniff, snitch, snob, snooty |
| /sp/ | spoil, spit, spew, spite, spinster |
| /spɹ/ | sprog, sprain |
| **Table Continues Overleaf** | |

| Onset | Examples of <<pejorative>> English words featuring this onset |
|---|---|
| /st/ | sty, stink, stupid, steal |
| /stɹ/ | strangle, strife, streetwalker, stress, stripper, strumpet |
| /sw/ | swastika, swear, sweat, sweatshop, swindle, swine, swizz |
| /ʃ/ | shit, shame, shoot |
| /t/ | tart, tosser, tit, tacky, tardy, tarnish |
| /θ/ | thug, thick, thump, thief |
| /θɹ/ | throw, thrash |
| /tɹ/ | tramp, treason, trump, trounce, trite |
| /tw/ | twit, twonk, twaddle |
| /v/ | vile, vulgar, vicious, vindictive, viper, vituperative, vitriolic |
| /w/ | woe, wacko, warble |
| /j/ | yob, usurp |
| **No. onsets recurring with <<pejorative>> meanings in this sample (excluding zero onset): 43** | |

Table 43: The number of onsets recurring with each meaning compared to the total productivity of the phonaesthemes (ranked in ascending order of multiple phonetic recurrence)

| Meaning | No. recurring onsets in English | Total productivity of respective phonaestheme (%) |
|---|---|---|
| straightness/stretching | 10 | /stɹ/ = 53 |
| fast/strong movement | 20 | /sw/ = 61 |
| crookedness | 23 | /kɹ/ = 46 |
| light | 23 | /gl/ = 77 |
| pejoration | at least 43 | /sl/ = 52 |

From these results, it does not appear that the perceptibility and productivity of the five phonaesthemes was substantially affected by multiple phonetic recurrence. If it had been, I would have expected to find that the most productive phonaesthemes were those whose meanings tended to recur with the fewest discrete onsets, as noted above. As such, I would

have expected to find that:

- <<light>> recurred with the **fewest** number of different onsets. This is because the /gl/-<<light>> phonaestheme was the most productive overall.
- <<fast/strong movement>> recurred with a greater number of different onsets than <<light>>, but fewer than the remaining phonaesthemes. This is because the /sw/-<<fast/strong movement>> phonaestheme was less productive than the /gl/ phonaestheme, but more productive than the remaining phonaesthemes.
- <<straightness>> recurred with a greater number of different onsets than <<light>>, and <<fast/strong movement>>, but fewer than <<pejoration>> and <<crookedness>>. This is because the /stɹ/-<<straightness>> phonaestheme was less productive than the /gl/ and /sw/ phonaesthemes, but more productive than the /sl/ and /kɹ/ phonaesthemes.
- <<pejoration>> recurred with a greater number of different onsets than <<light>>, <<fast/strong movement>> and <<straightness>>, but fewer than <<crookedness>>. This is because the /sl/-<<pejoration>> phonaestheme was less productive than the /gl/, /sw/ and /stɹ/ phonaesthemes, but more productive than the /kɹ/ phonaestheme.
- <<crookedness>> recurred with the **greatest** number of different onsets. This is because the /kɹ/-<<crookedness>> phonaestheme was the least productive overall.

However, Table 43 shows that this is simply not the case. For instance, I found that both <<light>> and <<crookedness>> recurred with exactly the same number of onsets. If multiple phonetic recurrence was a factor that had substantial affected the perceptibility and productivity of the phonaesthemes, then the overall productivity of these patterns should theoretically have been relatively similar. However, the /gl/-<<light>> phonaestheme and /kɹ/-<<crookedness>> phonaesthemes were polar opposites in terms of their overall productivity: Onset /gl/ attained the highest level of overall productivity, while onset /kɹ/ attained the lowest.

It is not only the <<light>> and <<crookedness>> findings which suggest that multiple semantic recurrence may not have affected the productivity of the phonaesthemes. Of the five meanings under test, I found that <<straightness/stretching>> recurred with the fewest number of distinct onsets in English. If multiple phonetic recurrence had been a factor

which had substantially affected the perceptibility and productivity of the phonaesthemes, then I would have expected to find that the /stɹ/-<<straightness/stretching>> phonaestheme was one of the most productive patterns, as noted above. However, as shown in Table 43, this pattern was neither the most nor least productive; attaining a figure of overall productivity lower than that of /gl/ and /sw/ but greater than that of /sl/ and /kɹ/. Further, I found that <<pejoration>> recurred with by far the greatest number of onsets, even despite the fact that I may have overlooked some <<pejorative>> words in my sample. In theory, then, the /sl/-<<pejoration>> phonaestheme should have been the least productive pattern among these five phonaesthemes; particularly by comparison to the /stɹ/-<<straightness>> phonaestheme. However, Table 43 again shows that this was not the case: the /sl/-<<pejoration>> phonaestheme was only the second least productive of the five patterns; attaining greater overall productivity than /kɹ/-<<crookedness>>, and only 1% less productivity than /stɹ/-<<straightness>>, despite the fact that <<straightness>> recurred with the fewest number of different onsets of all.

In short, then, there does not appear to be any clear relationship between the number of onsets with which the meanings recur and the productivity attained by the five phonaesthemes. The phonaestheme whose meaning recurs with the least number of onsets (/stɹ/) attains neither the highest nor lowest figure overall productivity. Conversely, the phonaestheme whose meaning recurs with by far the greatest number of onsets (/sl/) attains only the second lowest level of productivity of the five phonaesthemes sampled. Moreover, the meaning associations of the most and least productive phonaesthemes (/gl/ and /kɹ/) recur with an identical number of different onsets. In sum, these findings suggest that multiple phonetic recurrence does not appear to have been one of the main factors affecting the perceptibility and productivity of the phonaesthemes. This is not to suggest that multiple phonetic recurrence does not (or could not) affect the perceptibility and productivity of any phonaestheme in any language of the world; indeed, it might feasibly have affected any or all of the phonaesthemes that I did not subject to further testing. However, it does not appear that multiple phonetic recurrence was a factor substantially affecting the perceptibility and productivity of the /gl/, /sw/, /sl/, /stɹ/ or /kɹ/ phonaesthemes.

### 6.1.4) Summary of Findings pertaining to Multiple Semantic Recurrence and Multiple Phonetic Recurrence

The first goal of this chapter was to consider two other factors that might have affected the perceptibility and productivity of the phonaesthemes. In § 5.4, I compared the phonaesthemes' type and token frequencies against their productivity in English, and found there did not appear to be any clear link between either their type or token frequencies and the extent to which the patterns were productive (and thus perceptible). Many of the phonaesthemes with higher type frequencies (such as /sl/, /sm/ and /kɹ/) tended to attain comparatively low overall productivity; while many of the phonaesthemes with lower type frequencies (such as /kl/, /sw/ and /skw/) tended to attain comparatively high overall productivity. Similarly, many of the phonaesthemes with higher token frequencies (such as /sl/, /stɹ/ and /kɹ/) tended to attain comparatively low overall productivity; while many of the phonaesthemes with lower token frequencies (such as /kl/ and /tw/) tended to attain comparatively high overall productivity. As a result of these findings, I investigated whether the perceptibility and productivity of the phonaesthemes might instead have been affected by the number of different meanings with which their onsets recurred ('multiple semantic recurrence'); and/or the number of different onsets with which their meanings recurred ('multiple phonetic recurrence'). While it is plausible that both of these factors could have been influential, it does not appear that either factor wholly determined the perceptibility and productivity of the phonaesthemes. To some extent, it appears that the phonaesthemes could have been affected by multiple semantic recurrence. The results of my enquiry suggest that, as the number of different meanings recurring with an onset increases, the overall productivity of the phonaestheme tends to decrease. Nevertheless, at least two phonaesthemes appeared incongruent with this trend; suggesting that multiple semantic recurrence may not have been the only factor affecting the perceptibility and productivity of the patterns. However, it does not appear that the phonaesthemes were substantially affected by multiple phonetic recurrence. My enquiry found no demonstrable link between the productivity of the phonaesthemes and the number of onsets with which their meanings recurred. It may be that speakers' perception of the phonaesthemes, and the consequent productivity of these patterns, was actually influenced by a variety of factors; or by a particular factor that I did not test. This is clearly a question warranting further research in the future, as is discussed in Chapter 7. At present, however, it seems that the only clear way to establish the perceptibility of a phonaestheme is to conduct productive experiments like those discussed in Chapter 5.

This chapter now concludes with a discussion of the limitations of the present study.

## 6.2) Limitations of the Present Study

As previously noted, the second goal of this chapter is to discuss, objectively and in detail, the limitations of the present study. Broadly, these limitations fall into three broad categories. These include:

- The relative lack of previous research into both the perceptibility of phonaesthemes and the twelve patterns under investigation, and the difficulties this caused for designing my experiments and interpreting their findings;
- The scope of the study (including the time available for the project), which had implications for the numbers of respondents sampled and the specific patterns I chose to study;
- Limitations of the particular methodologies I used to investigate the twelve patterns.

In this section, each of these limitations is discussed in turn. As a result of these limitations, a number of methodological and investigative cues are identified – both for my own work and for any subsequent studies in the field – when I outline possible directions for future research in Chapter 7.

### 6.2.1) Limitations Concerning the Absence of Previous Research

The first limitation with the present study concerns the relative lack of previous research into both the perceptibility of phonaesthemes and the twelve patterns studied in this thesis, and the difficulties this causes for designing my experiments and interpreting their findings. As discussed in § 2.6, there is by no means extensive support or discussion in the literature of the twelve patterns explored in this thesis. With the notable exceptions of /gl/-<<light>> and /sl/-<<pejoration>>, the majority of these phonaesthemes are discussed by just two to three scholars in the field; while several are only identifiable in the work of Firth. What is more, the limited existing works to have studied these patterns are entirely vocabulary-based. That is, they have made claims that these phonotactic sequences have acquired certain connotations in English simply from studying the patterns in the vocabulary.

No study has yet investigated whether speakers actually perceive associations between these phonotactic sequences and meanings (connotations). Part of the reason for this phenomenon having received such little attention could be the inherent difficulty in investigating the idea of 'perception', as discussed in § 1.4.1. Indeed, as noted in § 1.4.2, it is not possible to 'feel' or experience people's cognitive reactions to such phonetic and semantic stimuli, or the extent of any affinities they may perceive between these stimuli; let alone measure these reactions directly. In the present study, I used the idea of productivity as an indirect way of operationalising and measuring the extent to which speakers perceived the patterns. However, as I noted in § 1.4.1, even this method was almost entirely novel; there are certainly no established ways of testing this phenomenon, or any conventional methodologies or strategies that I could have applied or adapted. Indeed, I had only a single precedent (Abelin, 1999) to motivate the design of my experiments. Even though I undertook a detailed critical evaluation of this previous study, and used its strengths and weaknesses to inform the design of my own investigation, my experiments were still pioneering in many respects. In practice, this meant that I had no guarantee as to whether my experiments provided the most effective contexts for the perception of the phonaesthemes; or whether other methods might have been equally if not more effective. As discussed in §§ 3.3 and 5.1, I strove to maximise the validity and reliability of my tests wherever possible. However, the only way I was able to judge whether my experiments had actually provided effective contexts for the perception of phonaesthemes was to investigate whether the different test types tended to generate congruent results; both in the pilot and main studies (cf. Chapter 5)..

The other main limitation arising from the lack of previous research was that it made interpreting my results somewhat more difficult. This was because I had no previous data against which to compare my findings; no way of contextualizing my results. As noted in § 3.3, this was one of the main factors which motivated my decision to compare the phonaesthemes across different languages; to help reach a more informed interpretation of the English data by obtaining other points of reference against which to compare each finding (see further discussion in § 6.2.2, below). The consequence of there being no previous studies against which to compare my results, and no established ways of testing or measuring the perception of phonaesthemes, was that I could not claim or even suggest that any given result in this study 'proved' that a phonaestheme had been perceived (or not). Indeed, I had no way of knowing or even estimating how many speakers would need to have been sampled, or how many times the patterns would need to have been used productively in order to generate

'proof' about whether the phonaesthemes were perceived or not (cf. §§ 3.2.2.2; 3.2.3). Rather, given the exploratory nature of my research, my analysis had to focus on numerical tendencies and trends, offering appropriately mitigated suggestions as to what these findings appeared to show. For instance, in Chapter 5, I compared the productivity of the phonaesthemes within experiments, across experiments, within each language cohort and across the three language cohorts, and the focus of these comparisons was on whether the phonaesthemes (and if so, which ones) appeared to have been more or less productive in each case. Comparing the phonaesthemes in this way (i.e. within and across experiments and within and across languages), and describing the general trends in the data, was the limit of what I could have attempted at this early stage of enquiry in this field.

### 6.2.2) Limitations Arising from the Scope of the Study

In addition to the lack of previous research in the field, another factor which caused difficulties in drawing conclusions from my data was the size of my respondent cohorts. To a large extent, the number of respondents I sampled was dictated by the temporal constraints of the study. As noted above, and as discussed in § 3.3, one of the main factors motivating my choice to study the phonaesthemes in other languages was to assist in interpreting the [English] data. I reasoned that studying the phonaesthemes in other languages would provide a useful basis against which to compare and interpret each finding. With multiple points of reference to compare each finding, it would be possible to reach a more informed interpretation of the data.

However, while exploring the productivity of the phonaesthemes in multiple languages offered this key advantage, this decision also introduced a further limitation to my study, with regards to the number of respondents sampled in each language (cf. § 3.3). By studying the patterns in two other languages besides English, I was not able to sample as many respondents as if I had simply focused on studying the patterns in English alone. This was largely due to constraints of time. I first had to arrange for the productive experiments (discussed in Chapter 5) to be professionally translated into French and Polish; a process which took several weeks. I then had to produce a separate online survey for each language cohort. Due to the complexity of embedding images and/or audio recordings of coined words in each question, this amounted to several further weeks spent writing .html code. In combination, these processes left me with less time to recruit and sample respondents without

overrunning the deadlines of the project than if I had restricted my enquiry to English alone. In practice, I sampled thirty speakers of each language group, on the basis that this number represented a twofold increase in sample size over the largest respondent groups in Abelin's (1999) Swedish study; the only existing precedent in the literature. In addition, I tried to ensure where possible that each survey was distributed to a diverse range of informants; both male and female, with a variety of different ages, education levels and socioeconomic backgrounds. In combination, I reasoned that these decisions would strike a balance between maximising the generalizability of my findings and the certainty with which I could interpret the results, and what was feasible in the time available.

However, even though I took the above steps to maximise the generalizability of my findings and the certainty with which I could interpret my results, sampling just thirty speakers in each language (rather than, say, a hundred) ultimately reduced the extent to which I could claim my findings were representative of the languages at-large, and the certainty with which I could claim that they had not come about by chance (cf. discussion of Abelin's 'Test 1a', § 3.2.2). When combined with the fact that my experiments were very much exploratory, with no established precedents and no extant findings against which their results could be compared (see above), the consequence of my [relatively small] sample sizes was that I had to be extremely tentative in interpreting the data. As noted above and in §§ 3.2.2.2 and 3.2.3, my analysis had to focus on the general trends in the results; whether individual phonaesthemes were more or less productive across experiments and between languages, and what these trends appeared to show about the productivity of the patterns. Beyond this, I was not able to draw any more definitive conclusions. In Chapter 7, I suggest that a future project could revisit the most and least productive phonaesthemes from this study and explore their productivity over a much larger group of English respondents. Studying the patterns over a single, larger respondent cohort would go some way towards addressing the limitations of the present study with regards sample size, and would increase both the generalizability of findings and the certainty with which one could draw interpretations and conclusions from the data. Such a study would lose the benefit of having other points of reference (i.e. languages) against which to compare its English findings, but could investigate, with greater certainty and generalizability, which of the phonaesthemes were more and less productive both within and across different experiments.

The second key limitation imposed by the [temporal] scope of my thesis was that it constrained the number of phonaesthemes I was able to investigate. As discussed in § 1.4.2, the length of the project meant that I was not able to study every phonaestheme – or even every word-initial phonaestheme – identified by scholars in the vocabulary of English. Rather, I had to restrict my enquiry to a more easily-manageable sample. One consequence of this is that there may well be other [onset] phonaesthemes that are even more productive than /gl/– either in English and/or French and Polish – which I was not able to explore in the present study (cf. discussion of /fl/-<<movement>> phonaestheme below and in Chapter 7.2). This means there is clearly further exploration required in this field in the future (cf. Chapter 7). However, the other consequence of my having to restrict the phonaesthemes tested is that I had to make a decision about which patterns to include and exclude from my study. In the case of one particular phonaestheme, it appears that my decision to exclude it from the study may have been premature. As noted in § 2.6.1, I decided not to investigate the /fl/ phonaestheme in the present thesis. My reason for this was that, although a number of subsequent scholars have associated /fl/ with <<movement>> in English (e.g. Sturtevant, 1947; Knowles, 1987: Ch.5; Allan, 2001:136), it was not at all clear that this was the connotation interpreted by Firth when he first cited an /fl/ phonaestheme in his work (1935:39). Of the three words provided by Firth to demonstrate the /fl/ phonaestheme (‘flick’, ‘flake’ and ‘fluke’), only two were congruent with the idea of <<movement>> (‘flick’ and ‘flake’); and even ‘flake’ was somewhat tenuous. As such, I argued that Firth may have envisaged a different meaning association among /fl/ words; one which was imperceptible from his choice of exemplifying words and lack of an explicit description of the phonaestheme. As a result of Firth’s less than convincing evidence for this phonaestheme in comparison to the other associations he posited, I decided to exclude it from the study, and to focus instead on a number of the remaining phonaesthemes cited in the literature.

However, on reflection, it appears this decision may have been somewhat premature; and that it might actually have been worthwhile studying the /fl/ phonaestheme after all. On closer reading of the literature, I found that there was actually more substantial scholarly support for an /fl/-<<movement>> phonaestheme than I had first appreciated. At least three scholars have posited the existence of an /fl/-<<movement>> phonaestheme in English; including Sturtevant (1947), Liberman (1990), and Allan (2001:136). Indeed, this pattern is actually more well-documented in the literature than a number of the patterns I chose to explore (e.g. /kl/-<<clumsiness; thickness; stupidity>>; /skw/-<<weakness; cuteness>>).

Therefore, given that an /fl/-<<movement>> phonaestheme does not appear congruent with the exemplar words provided by Firth (1935:39), it would seem worthwhile investigating whether [English] speakers **do** in fact perceive an association between this onset and a meaning of <<movement>>; or whether they tend to associate it with some other meaning. In response to this, I suggest in Chapter 7 that it would be worthwhile exploring this pattern further in a future study; particularly through the use of more open-ended experiments where speakers are totally unprompted and are free to associate the /fl/ prosody with any meaning of their choice (and indeed, the meaning of <<movement>> with any onset prosody of their choice).

### 6.2.3) Methodological Limitations

The final limitations of the present study are those concerning the particular ways in which I tested the phonaesthemes. These include limitations with the methodologies of my image-based experiments, as well as limitations caused by my decision to explore the phonaesthemes using only an online survey. This section discusses each of these in turn.

As noted in § 5.1, I chose to use a survey to test the phonaesthemes because it offered the greatest number of advantages (relative to the goals of my study) of the various techniques discussed in the literature. Unlike introspection, a survey allowed me to explore the perception of phonaesthemes among multiple, real-life speakers; rather than relying solely on my own intuition. Using a survey also presented fewer logistical difficulties than ethnographic or interview-based approaches. It allowed me to collect data from afar, eliminating the need for face-to-face or one-to-one contact with the individuals I sampled. In turn, this meant I did not need to travel to French and Polish speech communities, or acquire linguistic competence in French or Polish within an impractically short time. In addition, unlike psychological or psycholinguistic methodologies, a survey allowed me to conduct experiments that were appropriate to my research questions; namely, experiments to measure the extent to which speakers associated particular prosodies with particular meanings (and vice versa); and the numbers of speakers to do so in each case. Having determined that a survey was the most appropriate methodology, I chose to use an **online** survey, rather than a conventional paper-based approach, because it allowed me to incorporate both recordings and sound files into my questions by means of embedded hyperlinks. As the purpose of my study was to test the connotations ascribed to coined words on the basis of their **phonetic**

**structures**, it was crucial to provide these words in spoken form rather than written, to ensure that every informant was responding to identical phonetic stimuli. Notably, there would have been no way to feasibly incorporate spoken words in a paper-based survey.

However, although an online survey offered the advantages above, it would probably have been useful to combine this method with some form of open-ended face-to-face data collection methodologies (such as interviews or focus groups), as these could have allowed me to explore qualitatively, with participants, the reasons *why* they may have given particular answers, and *why* they may have associated particular prosodic forms with particular meanings. This would have been particularly insightful with respect to the image-based questions, and how these were interpreted by speakers (see further discussion below). These methodologies would have presented too many logistical issues in the present study; as discussed in § 5.1, this was one of the factors motivating my decision to use only the [online] survey to explore the phonaesthemes. However, it would almost certainly be useful to combine a survey with these types of methods if a similar study was attempted in the future – particularly if this study was only investigating the patterns in English (cf. Chapter 7.2). This is because there were a number of findings which, if clarified in a more open-ended way, could have increased the certainty with which I was able to interpret the data. For instance, did [English] speakers really perceive any differences between onsets /sl/, /sm/ and /sn/ in terms of their <<pejorative>> connotations? And if so, how were the connotations of each onset different? Why was /gl/ so consistently productive [in English] by comparison to the other patterns? What might the factors have been which led to speakers making such a regular association between this sequence and one particular meaning? Why did /kɪ/ tend to be less productive [in English] than the other patterns? Are <<weakness; cuteness>> and <<clumsiness; thickness; stupidity>> the most appropriate glosses of the connotations (if any) that tend to be ascribed to onsets /skw/ and /kl/? And how did speakers really interpret the images in Experiments 1 and 2? Did they identify each of the meanings I had intended to portray? Or did they tend to make other interpretations? Would there have been more appropriate ways of portraying the twelve meaning associations in the visual mode? An open-ended, qualitative methodology, such as a series of interviews or focus groups, would have allowed me to explore each of these (and other) questions in detail with my respondents, and would have offered greater clarity and certainty to my analysis and interpretation of the data. It is for this reason that I suggest combining a survey method with these approaches (in particular, the use of focus groups) when I outline possible directions for future research in

As well as the limitations of relying exclusively on an online survey, there were also a number of limitations with the methodologies of my image-based experiments. As noted in §§ 5.1.1 and 5.1.7, there are no established methodologies for testing the perceptibility of phonaesthemes. As a result, I used multiple choice experiments as one way to test the patterns, since multiple choice tests had seemed to provide effective contexts for the perception of phonaesthemes in Abelin's (1999) Swedish study (cf. Chapter 3.2.2.3). Two of my four tests in this study (Experiments 1 and 2) used multiple choice formats. In Experiment 1, I asked respondents to choose a particular sound sequence (coined word) to name or describe a given meaning. In Experiment 2, I asked respondents to choose a particular meaning to attribute to a given sound sequence (coined word). In each of these tests, I chose to portray the meanings associated with the twelve phonaesthemes using **images**, rather than linguistic contexts such as definitions or descriptions. There were several factors motivating this decision. The first was so that I could be absolutely certain I had not primed respondents to think about any of the phonaesthemes by unintentionally or unavoidably using the prosodies when presenting the twelve meanings. The second was to keep the survey questions as varied and stimulating as possible for respondents, in order to retain their attention and willingness to cooperate for as long as possible, and avoid any possible survey fatigue (cf. §§ 5.1.1; 5.1.7). Presenting the phonaesthemes' meanings using images, rather than definitions or descriptions, was intended to provide stimulation through the visual mode.

While choosing to present the phonaesthemes' meanings through images afforded the advantages discussed above, this decision was not without its disadvantages. Indeed, one of the consequences arising from this decision – the potential for misinterpretation – was arguably the main limitation of the present study. If I had used linguistic contexts (such as definitions or descriptions) to present the phonaesthemes' meanings in these questions, I could have deliberately constructed these linguistic contexts so as to be confident that respondents were all reacting to the same meanings that I had intended. However, using images did not offer this guarantee. Even though I modified Experiment 1 to direct participants towards specific parts of the images (cf. § 5.1.8), I had no guarantee that respondents would react to these parts of the images, or interpret the same kind of meaning(s) from them, as I had intended. After all, any image has the potential to be interpreted in different ways by different people. Moreover, in Experiment 2, it would have been unfeasibly

complex to direct participants to specific parts of each image, as discussed in § 5.1.8. This meant that I could not even be sure whether participants had responded to the same parts of the images I had intended in these questions. For these reasons, I had to exercise caution in making any claims about the phonaesthemes when analysing the findings of my image-based experiments, as discussed in §§ 5.1.1; 5.1.2; 5.2.1.1 and 5.2.2.1. In practice, Experiments 1 and 2 appeared to have generated largely congruent findings with Experiments 3 and 4 – particularly in English – which suggests that the majority of respondents may have successfully interpreted the meanings I had intended to convey (cf. §§ 5.1.1; 5.3); but the potential for misinterpretation in these questions meant that I still had to be cautious in drawing claims from their findings (cf. §§ 5.2.1.1; 5.2.2.2).

In hindsight, it would probably have been helpful to conduct one or more focus groups – both when designing my experiments and analysing their findings – to establish whether the images I had chosen were suitable ways of representing each meaning association. In the focus group(s) conducted at the 'design' stage, I could have asked participants to discuss how each meaning association would be best represented through an image or images, and why. I could also have shown participants the images I intended to use, and asked them to discuss whether they felt these images were suitable to represent the intended meanings, or whether it would be advisable to choose alternative images. Perhaps one could offer participants a series of images intended to depict each meaning, and ask informants to 'rank' these in terms of how suitable they felt each one would be. (Naturally, any focus groups performed at this stage would necessarily have sampled a different group of respondents to those surveyed for the productive experiments.) Then, in the focus groups conducted at the 'analytic' stage, I could have asked the respondents who completed the survey about what meaning(s) they had interpreted from each image, and what they thought had led them to choose the answers they had provided in each question. I could also have taken the opportunity to explore, with the participants, any themes or trends that had emerged during the analysis of the data, such as those listed on page 243 (above); an opportunity which I did not have in the present study. Focus groups would have been an ideal research method for these purposes, for a number of reasons. They would have offered a relatively quick means of sampling multiple people; they would have allowed me to clarify any necessary points with respondents quickly and simply; and they would have generated a large amount of rich and useful qualitative data vis-à-vis the suitability of images. Naturally, the logistics of this exercise would have made such focus groups extremely difficult in the

present study (cf. § 5.1). However, as I note in Chapter 7.2, this approach would be one I would advocate adopting in any future study in this area; particularly if one was only sampling speakers of a single language.

The next and final chapter concludes the thesis, and outlines a number of possible areas for further research in the future. In so doing, it discusses possible steps which could be taken in order to address the limitations of the present study discussed above.

# Chapter 7: Conclusion

This final chapter provides a brief review of the thesis, and a summary of its main findings. It then outlines a number of directions for future research. It concludes with a brief summary of the contribution made by the present study to the existing literature and to the knowledge about phonaesthemes at-large.

## 7.1) A Review of the Thesis and its Findings

In this thesis, I studied 12 word-initial phonaesthemes identified by scholars in the vocabulary of English:

1. The association between onset /sl/ and <<pejoration>>;
2. The association between onset /sm/ and <<pejoration>>;
3. The association between onset /sn/ and <<pejoration>>;
4. The association between onset /stɹ/ and <<straightness and stretching>>;
5. The association between onset /st/ and <<strength; stoicism>>;
6. The association between onset /gl/ and <<light>>;
7. The association between onset /gɹ/ and <<gripping>>;
8. The association between onset /tw/ and <<smallness >>;
9. The association between onset /sw/ and <<fast or strong movement>>;
10. The association between onset /skw/ and <<weakness and cuteness>>;
11. The association between onset /kl/ and <<clumsiness; thickness; stupidity>>;
12. The association between onset /kɹ/ and <<crookedness; curvilinearity>>.

I began by identifying a number of gaps in the literature pertaining to these phonaesthemes. Prior to this thesis, existing studies had not provided a clear, empirical measure of the extent to which these twelve patterns were present in the vocabulary of English. Most had simply cited a list (or lists) of words demonstrating one or more of the phonaesthemes, and a list of words to illustrate the presence of counterexamples. No study had calculated the proportion of words featuring the twelve onsets which also express the meanings outlined above; or the proportion of words which do not. Nor had any previous study established whether the twelve prosodies recur with the aforementioned meanings more frequently than with any other meaning(s) in the vocabulary of English; or if they do, how much more frequently. And no

study had investigated whether these phonaesthemes were also evident in the vocabularies of other languages; or whether they were unique to the vocabulary of English. What is more, no study had investigated whether these phonaesthemes were patterns that were actually perceived by speakers; either in English or any other language. It was not even clear whether the phonaesthemes were patterns that ordinary speakers (i.e. non-linguists) *could* perceive; or what sorts of factors might affect this perceptibility.

In response to these gaps in the literature, I posed the following Research Questions:

Research Question 1:
To what extent are the twelve phonaesthemes observable in the present-day vocabulary of English?

Research Question 2:
To what extent are the twelve phonaesthemes observable in the present-day vocabularies of languages other than English?

Research Question 3:

Do native English speakers appear to perceive the twelve phonaesthemes?

Research Question 4:
Do native speakers of languages other than English appear to perceive the twelve phonaesthemes?

Research Question 5:
What sorts of factors might affect the perception of the twelve phonaesthemes?

In addressing these Research Questions, I compared the presence and perception of the phonaesthemes in French and Polish alongside English. To address Research Questions 1 and 2, I used a two-stage vocabulary-based experiment. In the first stage of this experiment, I used dictionaries of comparable length to analyse the vocabularies of English, French and Polish and establish the phonaesthemes' type frequencies in each language. Type frequency refers to the number of head lexemes featuring each onset that also expressed its meaning

association listed above. In the second stage of the experiment, I then determined the three most common semantic fields expressed by words bearing each onset in English, French and Polish. In combination, these methods enabled me to generate a detailed quantitative picture of the extent to which the twelve phonaesthemes were present in the vocabularies of English, French and Polish. By calculating the phonaesthemes' type frequencies, I was able to compare the number and proportion of words that exhibited the twelve phonaesthemes in each language. By conducting the semantic field analysis, I was able to reach a more informed interpretation of this numerical data, by establishing the extent to which each onset recurred with the meaning under test relative to any other frequently-recurring meanings.

The findings from this experiment revealed clear evidence of all twelve phonaesthemes in the vocabulary of English; despite the fact that some were more pervasive associations (i.e. had higher type frequencies) than others. All twelve onsets recurred with the meanings under test, and thus exhibited the necessary conditions to be termed 'phonaesthemes' (cf. § 2.3). The most frequently-recurring of the twelve phonaesthemes was the /sl/-<<pejoration>> association, which recurred in 63 of the 148 /sl/ head lexemes. This was followed by the /st/-<<strength and stoicism>> phonaestheme (53 recurrences) and the /sn/-<<pejoration>> phonaestheme (37 recurrences). Conversely, the least frequently-recurring phonaesthemes were the /skw/-<<weakness and cuteness>> association (7 recurrences); the /gɹ/-<<gripping>> association (10 recurrences) and the /tw/-<<smallness>> association (11 recurrences). But despite the fact that some phonaesthemes attained higher type frequencies than others, ten of the twelve prosodies recurred more frequently with the meanings under test than with any other semantic fields. The only exceptions to this were onsets /kɹ/ and /gɹ/; though the meaning association of /gɹ/ (<<gripping>>) was nevertheless the second-most common semantic field among /gɹ/ headwords.

The findings from this experiment also revealed that many of the pattersn were evident in the vocabularies of French and Polish. In French, seven of the twelve onsets recurred with the meanings under test; these included /sl/, /sn/, /st/, /stɹ/, /gl/, /gɹ/ and /kɹ/. Similarly, in Polish, nine of the twelve onsets recurred with the meanings under test; these included /sl/, /sm/, /sn/, /st/, /stɹ/, /gl/, /gɹ/, /kl/ and /kɹ/. Only the /sw/, /skw/ and /tw/ phonaesthemes were not present to some extent in either language (these attained type frequencies of one or zero in both French and Polish). Notably, the /st/-<<strength and stoicism>> and /kɹ/-<<crookedness>> associations appeared to be particularly pervasive in

the vocabularies of both languages. Indeed, the type frequency of /kɹ/ in Polish was over double that in English. To a large extent, these findings were supported by the semantic field analysis. This revealed that, in both languages, four of the onsets (/sl/, /sn/, /st/, /gl/) recurred more frequently with the meanings under test than with any other meanings. For a further two onsets (/stɹ/ and /kɹ/), the meanings under test were among the most frequently-recurrent in the onsets' headwords. However, eleven of the twelve associations were ultimately more pervasive in the vocabulary of English. Every single phonaestheme attained a higher type frequency in English than in French; and only the /kɹ/-<crookedness>> phonaestheme attained a higher type frequency in Polish than in English. In addition, there were a greater number of phonaesthemes whose meaning associations were not among the three most common semantic fields of their onsets in French and Polish. This was true of /kl/, /gɹ/ and /sw/ in both languages; as well as /skw/ in French and /tw/ in Polish. This contrasted with English, where, for every phonaestheme except /kɹ/, the meanings under test were within the three most recurrent semantic fields of their onsets' headwords. On balance, then, while there was evidence of a number of the phonaesthemes in the vocabularies of French and Polish – particularly /st/ and /kɹ/ – there was clearer evidence of all but one phonaestheme in English.

To address Research Questions 3 and 4, I used four different productive experiments, each administered to the same cohort of 30 English speakers, 30 French speakers and 30 Polish speakers by way of an online survey. Each experiment featured twelve questions; one for every phonaestheme under study. In each question of the first experiment, I provided speakers with three coined words (in spoken form) and a single image. The image depicted a meaning associated with one of the phonaesthemes under test. Only one of the coined words featured the corresponding phonaestheme. The others featured different, randomly-chosen word onsets. I asked speakers to choose one of the words to name or describe a specified part of the image. In the second experiment, I reversed this process. In each question of this test, I provided speakers with a single coined word (in spoken form), bearing one of the twelve phonaesthemes under test. To control for any possible effects of syllable rimes, I combined each phonaestheme with a different rime prosody to the one used in the first experiment. In addition to the coined word, I also provided speakers with three images; of which only one depicted the meaning associated with the phonaestheme in the coined word. I asked speakers to choose whichever image they felt was best named or described by the coined word. The third experiment was more open-ended. In each question of this test type, I provided speakers with a definition related to one of the phonaesthemes. I asked speakers to coin any word of

their choosing to name or describe specified definition. Finally, in each question of the fourth experiment, I provided speakers with a single coined word bearing one of the phonaesthemes under test. I also provided a Semantic Differential scale. At one end of the scale, I placed the meaning associated with the phonaestheme in the coined word. At the other end, I placed an antonymous meaning. I asked speakers to place the coined word on the scale depending on how much (or how little) they thought it would express the meaning associated with its onset.

In analysing the findings of these experiments, I considered the data on both an experiment-by-experiment basis and at-large. In considering the findings at large, I calculated an overall figure of productivity for each phonaestheme in each language; an expression of how many times the pattern was used productively as a proportion of the number of times it could have been used productively. I used these figures to determine which phonaesthemes appeared to have been most and least productive in each language across the four experiments; and to carry out a macro-level comparison of the twelve phonaesthemes across the three language cohorts. In Experiments 1 and 2, I used statistical (Chi-Square) tests of statistical significance to supplement this overall comparison of the phonaesthemes across English, French and Polish.

The findings of the experiments suggest that all twelve phonaesthemes appear to have been perceived by English speakers; even though some of the patterns may have been perceived more consistently and to a greater extent than others. In Experiment 1 (words to image), eight of the onsets were productively associated with their respective meanings by a majority of the respondents. These included /gɹ/, /gl/, /st/, /tw/, /skw/, /sw/, /stɹ/ and /kl/. In Experiment 2 (images to word), all twelve meaning associations were productively associated with their respective onsets by half the respondent cohort or more. In Experiment 3 (open-ended definitions), all twelve prosodies were used by two or more speakers in response to their associated definitions. Finally, in Experiment 4 (SD scale), every single coined word (onset) was associated with the meaning under test by a majority of respondents. That is, a majority of respondents ranked each coined word at a position on the scale that was closer than the median point to the end at which I had placed its meaning association. Overall, eight of the twelve phonaesthemes were productive in every single experiment: the /gl/, /st/, /gɹ/, /tw/, /skw/, /sw/, /stɹ/ and /kl/ phonaesthemes. The /sm/ and /kɹ/ phonaesthemes were productive in three of the four experiments (tests 2, 3 and 4); while the /sl/ and /sn/ phonaesthemes were productive in two of the four experiments (tests 2 and 4). Insofar as

productivity is an indicator of perception, it therefore appears that all twelve phonaesthemes may have been perceived to some degree by English speakers.

However, despite the evidence that all twelve phonaesthemes appear to have been perceived to some extent in English, both the individual experiments and the findings at-large suggest that some phonaesthemes might have been perceived more consistently and to a greater extent than others. The figures of total productivity revealed that the /gl/-<<light>> phonaestheme was the most productive overall in English. In total, /gl/ was associated with <<light>> in seventy-seven percent of all instances where this association could have been made. It was the most productive pattern in experiments 2, 3 and 4, and the second most productive in experiment 1. Conversely, the /kɹ/-<<crookedness>> phonaestheme was the least productive phonaestheme overall. In total, /kɹ/ was associated with <<crookedness>> in forty-six percent of instances where this association could have been made; thirty-one percent fewer than /gl/-<<light>>. This phonaestheme was also unfailingly among the least productive patterns across the four experiments. Again, then, insofar as productive use is an indicator of perception, it appears that the /gl/-<<light>> phonaestheme may be a more widely-perceived pattern in English than the /kɹ/-<<crookedness>> phonaestheme. To some extent, based on the figures of overall productivity, it appears that the /st/ and /sw/ phonaesthemes may also be fairly widely-perceived patterns in English (albeit not as widely-perceived as /gl/); while the /stɹ/, and /sl/ phonaesthemes may be somewhat less widely-perceived patterns.

The findings also indicate that a number of the phonaesthemes may have been perceived by native speakers of French and Polish. In Experiment 1 (words to image) five of the onsets were productively associated with their respective meanings by a majority of French and Polish speakers. These included /gɹ/, /st/, /sw/ and /tw/ in both languages; as well as /skw/ in French and /stɹ/ in Polish. Similarly, in Experiment 2 (images to word), four phonaesthemes were used productively by a majority of each respondent cohort. These included /gl/, /sl/, /kl/ and /gɹ/ in French, and /gl/, /st/, /sm/ and /tw/ in Polish. In Experiment 3 (open-ended definitions), seven of the onsets were used twice or more in French respondents' coined words. These included /gl/, /st/, /kɹ/, /gɹ/, /stɹ/, /sl/ and /sn/. By comparison, four of the onsets were used twice or more in Polish speakers' coined words. These included /st/, /kɹ/, /gɹ/ and /sl/. Finally, in Experiment 4, a majority of French speakers awarded nine of the twelve phonaesthemes mean rank positions that were closer [than the

median point] to the end of the scale at which I had placed the onsets' meaning associations. These included /sl/, /sm/, /sn/, /kl/, /kɹ/, /gl/, /st/, /tw/ and /stɹ/. In Polish, the same was true for eight of these phonaesthemes; the only exception being /stɹ/. In each experiment, then, there was evidence that multiple phonaesthemes appeared to have been used productively (and thus appear to have been perceived) by both French and Polish speakers.

However, despite this evidence, it appears from both the individual experiments and the findings at-large that the phonaesthemes tended to be generally more widely-perceived in English. This is evident on a number of levels. First, the figures of total productivity indicate that every single phonaestheme was more productive overall in English than Polish; and that ten of the phonaesthemes were more productive overall in English than in French. Only the /sm/-<<pejoration>> and /sn/-<<pejoration>> phonaesthemes attained similar (in this case, marginally higher) levels of productivity in French as in English. Moreover, the findings of the statistical tests (Chi Squares) indicate that every single significant difference between the respondent cohorts was attributable to a higher level of productivity among English speakers. In addition to the above, there appear to have been fewer phonaesthemes productive across all four experiments in French and Polish than there were in English. In English, eight phonaesthemes were productive in all four experiments; with a further two phonaesthemes being productive in three of the four tests. However, in French, no phonaesthemes were productive across all four experiments; the most consistently productive were /st/, gl/, /sl/, and /gɹ/, which were productive in three of the four tests. In Polish, /st/ was productive across all four experiments, but no other phonaestheme was productive in more than two tests. Finally, in all four experiments, a majority of the phonaesthemes appear to have been more productive in English than in either other language. In Experiment 1, seven of the twelve phonaesthemes (/gl/, /gɹ/, kl/, /st/, /tw/, /skw/, /stɹ/) were more productive in English than in either other cohort. Similarly, in Experiment 2, ten of the twelve phonaesthemes (/gl/, /sw/, /sl/, /sn/, /st/, /kl/, /sm/, /kɹ/, /skw/, /stɹ/) attained their highest levels of productivity in English. In Experiment 3, only two of the twelve phonaesthemes (/sl/ and /sn/) attained equal or higher frequencies of use in French or Polish than in English; the remainder were more productive in English. And in Experiment 4, virtually every single phonaestheme attained a higher number of rankings towards its associated meaning in English than in either other cohort. All twelve patterns were ranked towards their meaning associations by a greater number of English speakers than Polish speakers, and every phonaestheme except /gɹ/-<<gripping>> was ranked towards its meaning association by a greater number of French

speakers than English speakers. On balance, then, it appears that the twelve phonaesthemes may have been more widely and more consistently-perceived patterns in English; even though their perception does not appear to have been confined exclusively to English speakers.

Finally, to address Research Question 5, I compared the phonaesthemes' type frequencies with their overall productivity and a representation of their token frequencies. The aim of this comparison was to establish whether the productivity (and, by extension, the perception) of the phonaesthemes appeared to have been most substantially affected by the number of different words in the language that exhibited the patterns, or by the extent to which speakers tend to be exposed to these patterns in everyday language use. I chose to address this research question in English alone, because there was no existing motivation to predict that the phonaesthemes would be present in vocabularies of French and Polish, or perceived by French and Polish speakers. To generate a representation of the phonaesthemes' token frequencies (i.e. the frequency with which speakers tend to encounter these patterns in everyday language use), I used the British National Corpus. I sampled 1,000 random words featuring each onset, and counted the number of words in each sample that expressed the meaning associated with their onset.

While both the type and token frequencies of the phonaesthemes could have affected their perceptibility and productivity, the findings of the three-way comparison suggested that neither factor appeared to have done so. There did not appear to be any demonstrable link between either the type or token frequencies of the phonaesthemes and the extent to which the patterns tended to be productive. Many of the phonaesthemes with higher type frequencies tended to attain comparatively low overall productivity; while many of the phonaesthemes with lower type frequencies tended to attain comparatively high overall productivity. Similarly, many of the phonaesthemes with higher token frequencies tended to attain comparatively low overall productivity; while many of the phonaesthemes with lower token frequencies tended to attain comparatively high overall productivity. These interpretations were supported by a series of Spearman-Rho tests, which showed that there were no significant correlations (either positive or negative) between the phonaesthemes' type or token frequencies and their overall productivity. As a result of these findings, I investigated in Chapter 6 whether the perceptibility and productivity of the phonaesthemes might instead have been affected by the number of different meanings with which the onsets

recurred (‘multiple semantic recurrence’); and/or the number of different onsets with which the meaning associations recurred (‘multiple phonetic recurrence’). I began by selecting a sample of five phonaesthemes for further testing. These included /gl/, /sw/, /sl/, /stɹ/ and /kɹ/. To investigate the possible effect of multiple semantic recurrence, I used the semantic field analysis prepared for Research Question 1. I counted the number of discrete semantic fields recurring with the five onset prosodies, and compared these figures to the overall productivity of the phonaesthemes. To investigate the possible effect of multiple phonetic recurrence, I used Roget’s thesaurus to establish the number of different word-onsets recurring with the five meaning associations. Once again, I then compared these figures to the overall productivity of the phonaesthemes.

The results of my further investigation suggested that the perceptibility and productivity of the phonaesthemes did not appear to be wholly determined by either multiple semantic recurrence or multiple phonetic recurrence. To some extent, it appears that the perceptibility and productivity of the phonaesthemes could have been affected by multiple semantic recurrence. I found that as the number of different meanings recurring with an onset increased, the overall productivity of the phonaestheme tended to decrease. Nevertheless, at least two phonaesthemes were incongruent with this trend, suggesting that multiple semantic recurrence might not have been the only factor affecting the perceptibility and productivity of the patterns. However, it was less clear that there was any correlation between the number of onsets recurring with each meaning and the levels of productivity attained by the phonaesthemes. This is not to suggest that multiple phonetic recurrence might not be capable of affecting the perceptibility and productivity of any phonaestheme whatsoever; it just does not appear to have been a factor substantially affecting the /gl/, /sw/, /sl/, /stɹ/ or /kɹ/ phonaesthemes.

### 7.2) Directions for Future Research

Having now reviewed the thesis and its main findings, this penultimate section outlines a number of directions for future research, with reference to the limitations of the present study discussed in Chapter 6.5.

Perhaps the most appropriate direction for a future research project in this area would be to start by exploring the perceptibility of some or all the phonaesthemes among a much

larger sample of speakers. One idea could be to return to the patterns which appeared most and least productive in this thesis (namely, /gl/, /st/, /sw/ and /kɹ/, /stɹ/, /sl/ respectively), and explore the perceptibility of these patterns in English only. The present study found that the twelve patterns generally appeared more regularly and more widely perceived in English; so a future study would not necessarily require the context of other languages against which to compare the findings in order to interpret the data. Rather, it could simply compare the differences in productivity (perception) of the phonaesthemes within this one cohort, and across the various experiments used. By exploring the phonaesthemes in English only, one could sample a much larger respondent group, since it would not be necessary to translate any of the data collection methodologies into other languages, or recruit respondents from other countries. Indeed, in the same time as I took to produce and translate the survey and then sample thirty English, French and Polish speakers in this project, one could probably sample around 150-200 English speakers relatively feasibly. In turn, this would address the limitations of present study vis-à-vis size of respondent groups sampled, and would substantially increase the generalizability of findings and the certainty with which one could make conclusions from the data.

If one was to attempt a future study of the type described above, and one was to use image-based questions as a way of testing the perceptibility of phonaesthemes, it would seem most appropriate to combine these methods with the use of focus groups. As discussed in Chapter 6, this mixed-methods approach would, in hindsight, have been very helpful in my own study; both in the design of my image-based experiments and the analysis of their findings. Conducting focus group(s) at the 'design' stage would help to establish which images were thought to be the most suitable ways of representing each meaning association by the speech community being sampled. For instance, one could ask participants to discuss how they felt each meaning association would be best represented through an image or images, and why. One could also show participants the images they intended to use, and ask the participants to discuss whether they felt these images were suitable to represent the intended meanings, or whether it would be advisable to choose alternative images. Perhaps one could offer participants a series of images intended to depict each meaning, and ask informants to 'rank' these in terms of how suitable they felt each one would be. Then, in the focus groups conducted at the 'analytic' stage, one could ask the specific respondents who completed the survey about what meaning(s) they had interpreted from each image, and what they thought had led them to choose the answers they had provided in each question.

Notably, using focus groups in this way would help to address the limitations of the present study vis-à-vis not knowing how participants were likely to react to images. This approach would offer the researcher some insight into how best to represent the desired meanings visually, such that the given speech community would be most likely to interpret them as intended. It would also offer the researcher the opportunity to check whether the actual respondents sampled had interpreted the desired meanings; and to explore what had motivated the respondents to provide the answers they had given. In combination, these processes would substantially increase the certainty with which one could make claims from any image-based tests. In the later focus groups – those conducted at the 'analytic' stage – one could also take the opportunity to explore and clarify, with the participants, any themes or trends that had emerged during the analysis of the data. For instance, in the present study, would have been helpful to explore (among other topics) whether speakers had perceived a difference in the <<pejorative>> connotations of /sl/, /sm/ and /sn/; as well as the possible reasons for /gl/ attaining comparatively high productivity overall and /kɹ/ attaining comparatively low productivity (cf. § 6.2.3). Notably, the use of focus groups would not present as many logistical difficulties in a future work [as it did in the present study] if one was sampling only speakers of one language (in this case, English).

In addition to sampling a larger respondent cohort and using focus groups alongside any image-based experiments, another possible direction for future research would be to explore the /fl/ phonaestheme in more detail. In Chapter 6, I suggested that although I had excluded the hypothesised /fl/-<<movement>> phonaestheme from present study, it might actually have been worthwhile studying this pattern after all. On closer reading of the literature, I found that there was actually more substantial scholarly support for an /fl/-<<movement>> phonaestheme than I had first appreciated. At least three scholars have posited the existence of an /fl/-<<movement>> phonaestheme in English; including Sturtevant (1947), Liberman (1990), and Allan (2001:136). Indeed, this pattern is actually more well-documented in the literature than a number of the patterns I chose to explore (e.g. /kl/-<<clumsiness; thickness; stupidity>>; /skw/-<<weakness; cuteness>>). Therefore, given that an /fl/-<<movement>> phonaestheme does not appear congruent with the exemplar words provided by Firth (1935:39), it would seem worthwhile investigating whether [English] speakers **do** in fact perceive an association between this onset and a meaning of <<movement>>; or whether they tend to associate it with some other meaning. In view of this, one suggestion could be to explore /fl/ alongside the other phonaesthemes chosen for

testing on a larger English-only respondent cohort (see above). It might be particularly appropriate to test this phonaestheme using more open-ended experiments than in the present study, as these would provide contexts where speakers were totally free to associate the /fl/ prosody with any meaning of their choice (and indeed, the meaning of <<movement>> with any onset prosody of their choice). In reality, the same applies to all phonaesthemes chosen for further exploration, since open-ended experiments are likely to provide more convincing evidence as to whether the patterns are productive (i.e. perceived) or not (cf. discussion in § 5.1.4). As such, it would seem appropriate to include at least one other open-ended experiment in any future study exploring these patterns. One option could be to retain the open-ended test from the present study, where speakers are provided with a definition and asked to coin a word in response to it; and combine this with another open-ended test where speakers are given a coined word and asked to invent a definition for it.

As well as exploring the /fl/ phonaestheme, I suggested in Chapter 5 that it would be worth exploring in a future study the different connotations ascribed to onsets /sl/, /sm/ and /sn/. It appears that there may be differences in the <<pejorative>> connotations ascribed to /sl/, /sm/ and /sn/ in English, but it is not clear at this stage precisely what these differences might be. To recap: In Experiment 3 of the present study (open-ended definitions), I provided three different <<pejorative>> definitions to test onsets /sl/, /sm/ and /sn/. These included; 'a large, vicious animal which silently stalks its prey and attacks when least expected'; 'a dodgy nightclub frequented by people of low moral standards', and 'the satisfied look worn by someone who feels they are better than everyone else'. As noted in § 5.2.3.1, the first definition was intended to evoke connotations of animalistic violence; the second to evoke connotations of promiscuity and the third to evoke connotations of gloating and condescension. I chose these particular connotations because of their congruence with the words provided by Firth (and other scholars) to exemplify the /sl/, /sm/ and /sn/ phonaesthemes. The <<pejorative animalistic>> associations of the first definition appeared to be congruent with many of Firth's (1930; 1935) exemplar words for the /sn/ phonaestheme, including 'sneak', 'snap' and 'snatch'. The <<pejorative promiscuous>> connotations of the second definition appeared to concur with many of Firth's (1930; 1935) and Crystal's (2002) exemplar words for the /sl/ phonaestheme, such as 'slut', 'slag', 'sleazy' and 'slattern'. Finally, the associations of <<pejorative condescension>> evoked in the third definition appeared to be congruent with many of Firth's (1930; 1935) exemplar words for the /sm/ phonaestheme, including 'smirch', smirk' and 'smug'. As such, I reasoned that

speakers might tend to coin words featuring onset /sn/ in response to definition one, by analogy to the various /sn/-<<pejorative animalistic>> words in English. I reasoned that speakers might tend to coin words featuring onset /sl/ in response to definition two, by analogy to the various /sl/-<<pejorative promiscuous>> words in English. Finally, I reasoned that speakers might tend to coin words featuring onset /sm/ in response to definition three, by analogy to the various /sm/-<<pejorative condescending>> words in English. Upon analysing the results, I found that /sm/ was used by ten English speakers in response to definition three, in which I had predicted /sm/ might be the most frequently-used onset. However, onsets /sl/ and /sn/ were not used at all. By contrast, onsets /sl/ and /sn/ were both used multiple times in response to definition one (/sl/ being used five times and /sn/ three times). Notably, /sm/ was not used at all in response to this definition. Finally, only /sl/ was used in response to the second <<pejorative>> definition. In combination, these findings suggest that there may be differences in connotation between the /sl/, /sm/ and /sn/. It appears that onset /sm/ was favoured by English speakers in response to the <<pejorative condescension>> expressed by definition three; and it appears that onset /sl/ may have been preferred by English speakers for the <<pejorative animalistic>> and <<pejorative promiscuous>> qualities of definitions one and two respectively. As a result of these findings, I suggested that the possible differences in connotation between onsets /sl/, /sm/ and /sn/ is a point which would benefit from further research in the future.

To investigate possible differences in connotation between onsets /sl/, /sm/ and /sn/, in English, it would seem most appropriate to use open-ended rather than closed-ended experiments, as in the possible future tests for /fl/, discussed above. This is because these types of test would provide contexts where speakers were totally free to associate the three prosodies with any meanings of their choice. One method of open-ended test which could be used would be to provide different coined words featuring the three prosodies (in spoken form), and ask participants to invent definitions for these coined words. This could be administered either through another online survey or, perhaps more profitably, through focus groups. The advantage of using focus groups over a survey would be that they would allow participants the freedom to discuss each onset at length, and 'to react to and build on the responses of other group members' (Stewart and Shamdasani, 2015:45-46; cf. advantages of focus groups in § 5.1.7). In turn, this could generate ideas and suggestions that would not otherwise have been revealed by one-to-one sampling, such as in a survey. Having asked participants to invent definitions for the coined words featuring the three prosodies, one could

then look for trends in the results (or focus group answers if this method was used). For example, it may be that the /sl/ coined words tended to elicit definitions relating to one semantic field, while the /sn/ coined words tended to elicit definitions relating to another. This would seem a more appropriate way of testing these phonaesthemes than trying to guess in advance what types of <<pejorative>> meanings might be associated with each onset, then providing definitions relating to these meanings and asking respondents to coin words to name or describe them.

In addition to the above, I noted in Chapter 6.5 that it may be worth exploring multiple semantic recurrence and multiple phonetic recurrence in greater detail in a future study. In Chapter 6, I did not test these phenomena on all twelve phonaesthemes explored in this thesis. Although it was appropriate to test and discuss these themes in some detail following the findings in § 5.4, they were not part of the central research aims of the study. Rather, they evolved organically from my analysis of the data. However, having established that multiple semantic recurrence might have affected speakers' perception of the phonaesthemes to some extent, but that multiple phonetic recurrence did not appear to have affected speakers' perception of the phonaesthemes, it could be worth investigating both factors further in a future work. Specifically, it could be worth testing whether either (or both) factors could have accounted for the levels of productivity attained by the seven phonaesthemes I was unable to test in Chapter 6 (namely, /sm/, /sn/, /st/, /kl/, /gɹ/, /tw/, /skw/). If one was to conduct such a study in the future, one could either re-use the productive data gathered in the present study and apply the same methodologies as in § 6.2 on the remaining phonaesthemes; or one could conduct new productive experiments, testing the productivity of all twelve phonaesthemes on a larger English respondent cohort, and using this data to explore multiple semantic/phonetic recurrence from scratch. If multiple semantic recurrence had affected the perceptibility of the remaining seven phonaesthemes (or the twelve patterns if starting anew), I would expect to find that the most productive patterns would be those whose onsets tended to recur with the fewest number of different meanings. If multiple phonetic recurrence had influenced the perceptibility of these phonaesthemes, I would expect to find that that the most productive patterns would be those whose meaning associations tended to recur with the fewest number of different onsets.

In the process of testing multiple semantic recurrence and multiple phonetic recurrence in more detail, it might also be worth investigating the possible effects of both

factors in terms of token frequencies. In Chapter 6, I studied both factors by type frequency but not token. To clarify: In studying multiple semantic recurrence, I investigated the possible effect caused by the number of different meanings with which each onset recurred **in the vocabulary of English**. Similarly, in studying multiple phonetic recurrence, I investigated the possible effect caused by the number of different prosodies with which each meaning recurred **in the vocabulary of English**. As such, the figures of multiple semantic/phonetic recurrence I calculated reflected the type frequencies of these factors; the extent to which the phenomena were exhibited in the vocabulary of the language (cf. Chapter 4). However, it would seem equally plausible that the **token** frequencies of these factors could also have influenced the productivity of the phonaesthemes. That is to say, the extent to which speakers associated the twelve onsets and meanings could also have been affected by: (i) the number of different meanings that speakers tend to encounter with each onset **in everyday language use**; and (ii) the number of different onsets that speakers tend to encounter with the expression of each meaning **in everyday language use**. Investigating this possibility would be a complex issue, and would have been beyond the remit of the present study, but could nevertheless be attempted in a future investigation. The first stage in such a project could be to select a small corpus of phonaesthemes; ideally the most and least productive patterns in the present study (/gl/, /st/, /stɪ/ and /kɪ/). The next stage would be to use the BNC to generate a large-scale random sample of concordance lines. In this sample, one would need to calculate: (i) the number of different meanings recurring with each onset; and (ii) the number of different onsets recurring with the corresponding meaning associations[25]. Having analysed the corpus sample, one could then compare the findings against the overall productivity of the phonaesthemes (either using the data from the present study or any 'new' data, if the productivity of the patterns was subsequently explored using a larger group of English speakers). If multiple semantic token recurrence had affected the perceptibility of the patterns, I would expect to find that the most productive phonaesthemes would be those whose onsets tended to recur with the fewest number of different meanings in the sample (representing everyday language use). If multiple phonetic token recurrence had affected the perceptibility of the patterns, I would expect to find that the most productive phonaesthemes would be those whose meaning associations tended to recur with the fewest number of different onsets in the sample.

[25] Part of the complexity of the study would be to determine how I could identify each discrete 'meaning' occurring in the corpus sample.

One final possibility for future study would be to investigate whether I could identify any other factors [beyond multiple semantic/phonetic recurrence] that might affect speakers' perception of phonaesthemes. For instance, it could be possible that the education level of respondents might affect their perception of such patterns. Speakers with higher levels of education might feasibly have had more opportunity to encounter phonaesthemes than speakers with lower levels of education; in such areas as literary texts, for example (cf. Short, 1996:114-124). This have might have led these respondents to calculate higher levels of probability [than respondents with lower levels of education] that certain prosodies and meanings would be found together. For a similar reason, it is also possible that speakers' ages might affect their perception of phonaesthemic patterns. Older speakers could have had more opportunity to encounter phonaesthemes in their linguistic experiences; which might lead them to calculate higher levels of probability that certain prosodies and meanings would be found together. Equally, older speakers might perceive certain phonaesthemic patterns which are no longer pervasive or even identifiable in the vocabulary of modern-day English. In practice, it would be relatively straightforward to explore both these factors in a future study. One could begin by selecting a small corpus of the most productive phonaesthemes identified in this thesis: /gl/, /st/, /sw/ and perhaps /kl/ (cf. Table 28). One could then use similar types of experiments as in Chapter 5 (but sampling a larger number of speakers) to test their productivity among native English speakers of different ages and education levels. This could all be tested using a single, anonymous [online] survey. One would simply need to ask for respondents' age and highest academic qualification at the beginning of the survey. Providing the same number of respondents were sampled in each age and education category, one could then compare the productivity attained by the phonaesthemes across each of these independent variables, and thus investigate whether either factor appeared to have affected speakers' perception of the patterns. For example, if the education level of respondents had affected their perception of the phonaesthemes as outlined above, then I would expect to find a positive correlation between speakers' education level and their productive use of the phonaesthemes: As one were to increase, I would expect that the other would tend to increase. Similarly, if the age of respondents had affected their perception of the phonaesthemes as outlined above, I would expect to find a positive correlation between speakers' age and their productive use of the phonaesthemes.

On balance, it is clear that there are many possible directions for future research in this field. Principally, it would be worth exploring the productivity of these patterns over a

much larger English respondent cohort. Should image-based questions be used in such a study, it would be worth using a mixed-methods (i.e. qualitative and quantitative) approach, to increase the certainty with which one could make claims from the findings. In addition, there are other onset phonaesthemes – most notably /fl/ – whose perceptibility is yet to be tested; and there possible differences between phonaesthemes (namely /sl/, /sm/ and /sn/) which would benefit from further exploration. There are also other factors which might affect the perceptibility of phonaesthemes that are still to be investigated. At present, however, it is worth concluding this study by reflecting on its contributions to the field of enquiry. The next and final section provides an overview of how this thesis adds to the existing knowledge about phonaesthemes.

### 7.3) Contributions to the Field of Enquiry

This thesis constitutes an important contribution to the current literature pertaining to phonaesthemes, and develops the existing knowledge about these relationships on a number of levels. This thesis provides the first thorough quantitative investigation into the presence of the twelve onset phonaesthemes in the vocabulary of English. Unlike the existing studies in the literature, this investigation does not just calculate the number of different words exhibiting each phonaestheme; it also calculates the proportion of headwords constituted by each phonaesthemic pattern, and establishes whether the twelve onsets recur with their meaning associations more frequently than with any other meanings. In so doing, it reveals clear evidence of all twelve phonaesthemes in the vocabulary of English. It reveals that the /sl/-<<pejoration>> phonaestheme is the most pervasive pattern of those under test; while the /tw/-<<smallness>> phonaestheme is the least pervasive. This thesis also provides the first quantitative investigation into the presence of the twelve phonaesthemes in the vocabularies of French and Polish. In so doing, it provides clear evidence of many (though not all) of the phonaesthemes in the vocabularies of French and Polish; particularly the /kɹ/-<<crookedness>> and /st/-<<strength/stoicism>> phonaesthemes. Notably, it reveals that the twelve phonaesthemes are almost exclusively more pervasive in the vocabulary of English.

In addition to the above, the present study also provides the very first evidence to suggest that speakers can perceive phonaesthemic patterns. Its multiple productive experiments offer quantitative evidence to suggest that English speakers perceived all twelve phonaesthemes under test; that is, identified an affinity between the meaning associations and

the twelve prosodies that was not present with other prosodies, and identified an affinity between the prosodies and the twelve meaning associations that was not present with other [different] meaning associations. Notably, its findings suggest that phonaesthemic patterns may vary in the extent to which they are perceptible. The /gl/-<<light>> phonaestheme appeared to have been the most frequently and strongly perceived pattern in English; while the /kɹ/-<<crookedness>> phonaestheme appeared to have been the least frequently perceived. The study also provides quantitative evidence to suggest that a number of the phonaesthemes may also be perceptible in other languages; namely, French and Polish. This appears to be particularly true of the /gɹ/-<<gripping>> phonaestheme in French and the /st/-<<strength/stoicism>> phonaestheme in Polish. However, the data suggests that the twelve patterns tend to be perceived more consistently in English than in Polish; and that the majority of the patterns tend to be perceived more consistently in English than in French.

Finally, the present study provides the very first evidence about the sorts of factors that might affect speakers' perception of phonaesthemes. The perception of the twelve phonaesthemes [among English speakers] does not appear to have been influenced by the number of different words exhibiting the patterns in the vocabulary of the language; nor by the frequency with which speakers tend to encounter these patterns. Neither does it appear to have been affected by the number of different onsets with which each meaning association recurs (multiple phonetic recurrence). At some level, speakers' perception of the phonaesthemes may have been influenced by the number of different meanings with which the onset prosodies recur (multiple semantic recurrence). However, further research in this field is needed in the future to determine whether – and if so, what sorts – of other factors could influence the perceptibility of these sound-meaning patterns.

**WORD COUNT: 85351**

## References


Abbuhl, R, Gass, S and Mackey, A (2013) "Experimental research design". In Podesva, R J and Sharma, D (eds.) (2013) *Research Methods in Linguistics*. Cambridge: Cambridge University Press.

Abelin, A (1999) *Studies in Sound Symbolism*. PhD thesis: Göteborg University.

Ager, S (2014) "French Pronunciation". Available online at http://www.omniglot.com/writing/french.htm, accessed 17/03/2014.

Ahlner, F and Zlatev, J (2010) "Cross-modal iconicity: A cognitive semiotic approach to sound symbolism". *Sign Systems Studies* 38, p.298-348.

Aitchison, J (2012) *Words in the Mind: An Introduction to the Mental Lexicon*. Oxford: John Wiley and Sons Ltd.

Akamatsu, N (2006) "Literacy Acquisition in Japanese-English Bilinguals". In Joshi, R M and Aaron, P G (eds.) *Handbook of Orthography and Literacy*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Allan, K (2001) *Natural Language Semantics*. Oxford: Wiley-Blackwell.

Allott, R (1995) "Sound Symbolism". Available online at http://www.percepp.com/soundsmb.htm, accessed 06/12/2012.

Anderson, E R (1998) *A Grammar of Iconism*. Cranbury, N.J., London and Ontario: Associated University Presses, Inc.

Attridge, D (2004) *Peculiar Language*. London: Routledge.

Auer, E T Jr. and Luce, P A (2005) "Probabilistic Phonotactics in Spoken Word Recognition". In Pisoni, D B and Remez, R E (eds.) *The Handbook of Speech Perception*. Oxford: Blackwell.

Babbie, E (2013) *The Practice of Social Research*. Belmont, CA: Wadsworth.

Backhouse, A (1992) "Connotation". In Bright, W (ed.) *The International Encyclopedia of Linguistics*. New York: Oxford University Press.

Bajpai, N (2011) *Business Research Methods*. Noida and New Delhi: Dorling Kindersley India.

Bauer, L (1983) *English Word-formation*. Cambridge: Cambridge University Press.

Bauer, L (1998) *Vocabulary*. Abingdon and New York: Routledge.

Ben-Nun, P (2008) "Respondent Fatigue". In Lavrakas, P J (ed.) *Encyclopedia of Survey Research Methods*. Los Angeles, London, New Delhi, Singapore and Washington, DC: Sage.

Bergen, B (2004) "The Psychological Reality of Phonaesthemes". *Language*, 80, 2:290-311.

Bethlehem, J and Biffignandi, S (2012) *Handbook of Web Surveys*. Hoboken, NJ: John Wiley & Sons, Inc.

Betjeman, J (2006) *John Betjeman: Collected Poems*. London: John Murray.

Biber, D (1993) "Representativeness in corpus design". *Literary and Linguistic Computing*, 8, 4: 243-257.

Blalock, H M Jr. (1974) *Measurement in the Social Sciences: Theories and Strategies*. Chicago: Aldine Publishing Company.

Blank, A and Koch, P (1999) *Historical Semantics and Cognition*. Berlin: Walter de Gruyter & Co.

Blevins, J (2007) "Interpreting Misperception: Beauty is in the Ear of the Beholder". In Solé, M J, Beddor, P S and Ohala, M (eds.) *Experimental Approaches to Phonology*. Oxford: Oxford University Press.

BNC (2009) "What is the BNC?" Available Online at http://www.natcorp.ox.ac.uk/corpus/index.xml, accessed 23/07/2012.

Bolinger, D (1950) "Rime, Assonance and Morpheme Analysis". *Word* 6: 117-136.

Bolinger, D (1965) *Forms of English: Accent, Morpheme, Order*. Cambridge, MA: Harvard University Press.

Bolinger, D (1968) *Aspects of Language*. New York: Harcourt, Brace and World.

Bolinger, D and Sears, D A (1981) *Aspects of Language*. New York: Harcourt, Brace and Jovanovich.

Börkan, B (2010) "The mode effect in mixed-mode surveys: Mail and web surveys". *Social Science Computer Review*, 28: 371-380.

Bowles, H (1995) "The Semantic Properties of the Phonaestheme". *Studi italiani di linguistica teorica ed applicata*, 1: 91-106.

Brinton, L J (2004) *The Structure of Modern English*. Amsterdam: John Benjamins.

Brooks, M Z (1975) *Polish Reference Grammar*. The Hague: Mouton & Co.

Brady, H E (1990) "Dimension analysis of ranking data". *American Journal of Political Science*, 34: 1017-1048.

Brown, A (1991) *Pronunciation Models*. Kent Ridge: Singapore University Press.

Brown, C G (2013) *Postmodernism for Historians*. Abingdon: Routledge.

Brown, R W (1958) *Words and Things*. New York: Free Press.

Bryant, P, Howard, D, Lock, G and Philander, Z (2009) *Systems Design and Analysis*. Cape Town: Pearson Education South Africa.

Burdett, K W (2002) "New Paradigms: A Rule-and-Feature Based, Morpholexical Analysis of the Spanish Verbal System". University of Georgia: PhD Thesis.

Burke Johnson, R and Christensen, L (2014) *Educational Research*. Thousand Oaks, London, New Delhi and Singapore: Sage Publications, Inc.

Burling, R (2000) "Motivation, Conventionalization, and Arbitrariness in the Origin of Language". In King, B J (ed.) *The Origins of Language: What Nonhuman Primates Can Tell Us*. Oxford: James Currey.

Bybee, J (2001) *Phonology and Language Use*. Cambridge: Cambridge University Press.

Carey, T (2011) "We Need to Talk…the four words that could kill your marriage". *Daily Mail Online*. Available at http://www.dailymail.co.uk/femail/article-2029793/We-need-talk--words-KILL-marriage.html, accessed 21/11/13.

Carr, P (2008) *A Glossary of Phonology*. Edinburgh: Edinburgh University Press.

Carroll, D W (2008) *Psychology of Language*. Belmont, CA: Thomson Higher Education.

Carroll, L (2001) *Jabberwocky and Other Poems*. Mineola, NY: Dover Publications, Inc.

Carpenter, H (ed.), Tolkien, J R R and Tolkien, C (1981) *Letters of J.R.R. Tolkien*. London: George Allen & Unwin.

Charoenruk, N and Stange, M (2014) "Respondent Processing of Multiple Images throughout a Web Survey". *Proceedings of the 69th Annual Conference of the American Association for Public Opinion Research (AAPOR)*. Anaheim, CA: May 2014.

Chomsky, N and Halle, M (1968) *The Sound Pattern of English*. Cambridge, MA: MIT Press.

Clark, J, Yallop, C and Fletcher, J (2007) *An Introduction to Phonetics and Phonology*. Oxford, UK, Malden, MA and Carlton, Victoria: Blackwell.

Clements, G N and Hume, E V (1995) "The Internal Organization of Speech Sounds". In Goldsmith, J A (ed.) *The Handbook of Phonological Theory*. Oxford, UK and Cambridge, MA: Blackwell. p.245-306.

Cohen, L, Manion, L and Morrison, K (2011) *Research Methods in Education*. Abingdon: Routledge.

Cole, J (2014) *Web Survey Creator Reference Guide*. Sydney: Dipolar Pty Limited.

Converse, P E (1964) “The nature of belief systems in the mass public”. In Apter, D E (ed.) *Ideology and Discontent*. New York: Free Press.

Cook, J D, Hepworth, S J, Wall, T D and Warr, P B (1981) *The Experience of Work*. San Diego: Academic Press.

Couper, M P and Bosnjack, M (2010) “Internet Surveys”. In Marsden, P V and Wright, J D (eds.) *Handbook of Survey Research.* Bingley: Emerald Group Publishing Limited.

Cruttenden, A (2008) *Gimson’s Pronunciation of English*. London: Hodder Education.

Crystal, D (1995a) “Still Finicking with Trifles, Largely”. Otto Jespersen Memorial Lecture.

Crystal, D (1995b) “Phonaesthetically Speaking”. *English Today* 42, 11: 8-11.

Crystal, D (2001) *Language and the Internet*. Cambridge: Cambridge University Press.

Crystal, D (2002) *The English Language*. London: Penguin Books Ltd.

Crystal, D (2003) *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press.

Crystal, D (2006) *Words, Words, Words*. Oxford: Oxford University Press.

Crystal, D (2011) *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell.

Currie Hall, K (2011) “Phonological Relationships: A Probabilistic Model”. Available online at http://www.ling.ohio-state.edu/~kchall/Hall_2011_MOT_poster.pdf , accessed 09/08/2012.

Day, H N (1867) *The Art of Discourse: A System of Rhetoric*. New York: Charles Scribner and Company.

De Cuypere, L (2008) *Limiting the Iconic*. Amsterdam: John Benjamins.

Deignan, A (2005) *Metaphor and Corpus Linguistics*. Amsterdam and Philadelphia: John Benjamins.

Diffloth, G (2006) “i:big, ɑ:small”. In Hinton, L, Nichols, J and Ohala, J J (eds.) *Sound Symbolism*. Cambridge: Cambridge University Press. p.107-114.

Dingemanse, M (2008) “Three misconceptions about ideophones”. Available online at http://ideophone.org/three-misconceptions-about-ideophones/ , accessed 02/03/2012.

Dixon, R M W (1982) *Where Have All the Adjectives Gone? And Other Essays in Semantics and Syntax.* Berlin: Walter de Gruyter & Co.

Doke, C M (1935) *Bantu Linguistic Terminology*. London: Longmans and Green.

Dörnyei, Z (2007) *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.

Dörnyei, Z and Taguchi, T (2010) *Questionnaires in Second Language Research: Construction, Administration, and Processing*. New York and London: Routledge.

Dumont, K (2008) "Research Methods and Statistics". In Nicholas, L (ed.) *Introduction to Psychology*. Cape Town: UCT Press.

Du Plooy, T (1995) "Measurement in Communication Research". In du Plooy, G M (ed.) *Introduction to Communication Research (Vol.2)*. Kenwyn, South Africa: Juta & Co. Ltd.

Durkin, P (2009) *The Oxford Guide to Etymology*. Oxford: Oxford University Press.

Edwards, J E, Thomas, M D, Rosenfeld, P and Booth-Kewley, S (1997) *How to Conduct Organizational Surveys*. Thousand Oaks, London and New Delhi: Sage Publications, Inc.

Ehrlich, H J (1964) "Instrument error and the study of prejudices". *Social Forces*, 43: 197-206.

Elig, T W and Frieze, I H (1979) "Measuring causal attributions for success and failure". *Journal of Personality and Social Psychology,* 37: 621-634.

Ellis, N C (2002) "Frequency effects in language processing and acquisition". *Studies in Second Language Acquisition*, 24: 143-188.

Faber, P B and Usón, R M (1999) *Constructing a Lexicon of English Verbs*. Berlin: Mouton de Gruyter.

Fabiszak, M (2005) "Semantic and Lexical Change". In Ammon, U, Dittmar, N, Mattheier, K J and Trudgill, P (eds.) *Sociolinguistics Volume 2: An International Handbook of the Science of Language and Society*. Berlin: Walter de Gruyter GmbH.

Fazly, A, Alishahi, A and Stevenson, S (2010) "A Probabilistic Computational Model of Cross-Situational Word Learning". *Cognitive Science* 34: 1017-1063.

Finegan, E (2012) *Language: Its Structure and Use*. Boston, MA: Thomson-Wadsworth.

Firth, J R (1930) *Speech*. In Firth, J R (1964) *The Tongues of Men and Speech*. London: Oxford University Press.

Firth, J R (1935) "The Use and Distribution of Certain English Sounds". In Firth, J R *Papers in Linguistics 1934-1951*. London: Oxford University Press.

Firth, J R (1948) "Sounds and Prosodies". In Firth, J R *Papers in Linguistics 1934-1951*.

London: Oxford University Press.

Firth, J R (1951a) “Modes of Meaning”. In Firth, J R *Papers in Linguistics 1934-1951*. London: Oxford University Press.

Firth, J R (1951b) “General Linguistics and Descriptive Grammar”. In Firth, J R *Papers in Linguistics 1934-1951*. London: Oxford University Press.

Firth, J R (1957) "A Synopsis of Linguistic Theory 1930-1955". In Palmer, F (ed.) (1968) *Selected Papers of J R Firth.* Harlow: Longman.

Fischer, A (1999) “What, if Anything, is Phonological Iconicity?” In Fischer, O and Nänny, M (eds.) *Form Miming Meaning: Iconicity in Language and Literature*. Amsterdam: John Benjamins.

Fischer, O and Nänny, M (2001) “Introduction: Veni, Vedi, Vici”. In Fischer, O and Nänny, M (eds.) *The Motivated Sign: Iconicity in Language and Literature 2*. Amsterdam: John Benjamins.

Fónagy, I (2000) *Languages within Language: An Evolutive Approach*. Amsterdam and Philadelphia: John Benjamins.

Fontaine, L M (2008) “A Systemic Functional Approach to Referring Expressions: Reconsidering Postmodification in the Nominal Group”. PhD thesis: Cardiff University.

Fougeron, C and Smith, C L (1999) “Illustrations of the IPA: French”. In *The Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.

Frege, G (1892) “On sense and reference”. In Moore, A W (ed.) (1993) *Meaning and Reference*. Oxford: Oxford University Press.

Gee, J P and Hayes, E R (2011) *Language and Learning in the Digital Age*. Abingdon: Routledge.

Genette, G (Trans. Morgan, T E) (1995) *Mimologics*. Lincoln and London: University of Nebraska Press.

Giegerich, H J (1992) *English Phonology: An Introduction*. Cambridge: Cambridge University Press.

Giles, H and Bourhis, R Y (1976) “Methodological Issues in Dialect Perception: Some Social Psychological Perspectives”. *Anthropological Linguistics* 18, 7: 294-304.

Giles, H and Powesland, P F (1975) *Speech Style and Social Evaluation*. London: Academic Press.

Gill, P (2011) *Colouring Meaning: Collocation and Connotation in Figurative Language*. Amsterdam: John Benjamins.

Glastonbury, B and MacKean, J (1991) “Survey Methods”. In Allan, G and Skinner, C (eds.)

*Handbook for Research Students in the Social Sciences*. Abingdon: RoutledgeFalmer.

Goldsmith, J A (1979) "The Aims of Autosegmental Phonology". In Dinnsen, D A (ed.) *Current Approaches to Phonological Theory*. Bloomington, IA and London, UK: Indiana University Press.

Gorrell, R (1994) *Watch Your Language: Mother Tongue and her Wayward Children*. Reno: University of Nevada Press.

Grace, G W (2005) "Reflections on the evolution of human language…A Role for Phonaesthesia?". Available online at http://www2.hawaii.edu/~grace/phonaesthesia.html , accessed 20/03/2012.

Grammont, M (1901) "Onomatopées et mots expressifs". *Tretenaire de la Societé pour l'Etude des Langues Romanes*. p.261-322.

Greene, R, Cushman, S, Cavanagh, C, Ramazani, J and Rouzer, P (eds.) (2012) *The Princeton Encyclopedia of Poetry and Poetics*. Princeton, NJ and Woodstock, Oxfordshire: Princeton University Press.

Halle, M and Mohanan, K P (1985) "Segmental Phonology of Modern English". *Linguistic Enquiry*, 16: 57-116.

Halliday, M A K (1970) *A Course in Spoken English: Intonation*. Oxford: Oxford University Press.

Halliday, M A K (2003) "Introduction: On the "Architecture" of Human Language". In Halliday, M A K and Webster, J J (ed.) *On Language and Linguistics*. London and New York: Continuum.

Halliday, M A K and Yallop, C (2007) *Lexicology: A Short Introduction*. London and New York: Continuum.

Hancock, T (2013) "Licensed for Embracement: Insouciance and Exigency in John Betjeman's Love Poems". *The Cambridge Quarterly*, 42, 4: 342-256.

Hanks, P (2004) "Corpus Pattern Analysis". *Euralex 2004 Proceedings*: 87-97.

Hassler, G (2007) "'Analogy': The History of a Concept and a Term from the 17th to the 19th Century". *History of Linguistics 2005: Selected Papers from the Tenth International Conference on the History of the Language Sciences*. Amsterdam: John Benjamins.

Hayes, B (2009) *Introductory Phonology*. Chichester, West Sussex: Wiley-Blackwell.

Hinton, L and Bolinger, D (2003) "Sound Symbolism". In Frowley, W J (ed.) *International Encyclopedia of Linguistics: Volumes 1-4*. Oxford: Oxford University Press. p.111-114.

Hinton, L, Nichols, J and Ohala, J J (eds.) (2006) *Sound Symbolism*. Cambridge:

Cambridge University Press.

Hinton, P R (2004) *Statistics Explained*. Abingdon: Routledge.

Hockett, C F (1960) "The Origin of Speech". *Scientific American* 203, p.88-111.

Hoffmann, T R (1971) "On Assimilation at a Distance". *Journal of Speech and Hearing Research* 14, 4: 880-881.

Hunston, S (1995) "A Corpus Study of Some English Verbs of Attribution". *Functions of Language*, 2 (2): 133-158.

Hurford, J (2012) *The Origins of Grammar: Language in the Light of Evolution*. Oxford: Oxford University Press.

Hutson, R A and Kolbe, A R (2010) "Survey Studies". In Thyer, B A (ed.) *The Handbook of Social Work Research Methods*. Thousand Oaks, London, New Delhi and Singapore: Sage Publications, Inc.

Jackson, H and Zé Amvela, E (2000) *Words, Meaning and Vocabulary: An Introduction to Modern English Lexicology*. London and New York: Continuum.

Jespersen, O (1922a) *Language: Its Nature, Development and Origin*. London: George Allen & Unwin Ltd.

Jespersen, O (1922b) "Symbolic Value of the Vowel I". In Jespersen, O, *Selected Writings of Otto Jespersen*. Abingdon: Routledge.

Johns, R (2010) "Likert Items and Scales". Available online at http://ukdataservice.ac.uk/media/262829/discover_likertfactsheet.pdf , accessed 26.06.2014.

Jones, D (2003) "Our Spoken Language". In Collins, B and Mees, I M (eds.) *Daniel Jones: Selected Works Volume 8*. London and New York: Routledge.

Jurafsky, D (2003) "Probabilistic modelling in psycholinguistics". In Bod, R, Hay, J and Jannedy, S (eds.) *Probabilistic Linguistics*. Cambridge, MA: MIT Press.

Kays, K M, Keith, T L and Broughal, M T (2013) "Best Practice in Online Survey Research with Sensitive Topics". In Sappleton, N (ed.) *Advancing Research Methods with New Technologies*. Hershey, PA: Information Science Reference.

Kenstowicz, M and Kisseberth, C (1979) *Generative Phonology*. San Diego: Academic.

Kenstowicz, M (1994) *Phonology in Generative Grammar*. Cambridge, MA and Oxford, UK: Blackwell.

Kim, K O (1977) "Sound Symbolism in Korean". *Journal of Linguistics* 13:67-75.

Knight, R A (2012) *Phonetics: A Course Book*. Cambridge: Cambridge University Press.

Knowles, G (1987) *Patterns of Spoken English: An Introduction to English Phonetics. London: Longman.*

Kreidler, C W (2004) *The Pronunciation of English*. Oxford, Malden and Carlton: Blackwell.

Krosnick, J (1991) "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys". *Applied Cognitive Psychology*, 5: 213-236.

Krosnick, J (1999a) "Maximising Questionnaire Quality". In Robinson, J P, Shaver, P R and Wrightsman, L S (eds.) *Measures of Political Attitudes*. San Diego, CA: Academic Press.

Krosnick, J (1999b) "Survey Research". *Annual Review of Psychology*, 50: 537-567.

Labov, W (1966) *The Social Stratification of English in New York City*. Washington DC: Center for Applied Linguistics.

Ladefoged, P (1999) "Illustrations of the IPA: American English". In *The Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.

Ladefoged, P and Johnson, K (2010) *A Course in Phonetics*. Boston, MA: Wadsworth.

Lardiere, D (2006) "Words and their parts". In Fasold, R W and Connor-Linton, J (eds.) *An Introduction to Language and Linguistics*. Cambridge, UK and New York: Cambridge University Press.

Lechte, J (2008) *Fifty Key Contemporary Thinkers: From Structuralism to Post-Humanism*. London: Taylor & Francis Ltd.

Leech, G (1974) *Semantics*. Harmondsworth: Penguin.

Liberman, A (1990) "Etymological Studies III: Some Germanic Words beginning with FL-. Language at Play". *General Linguistics*, 30:2.

Liberman, A (2009) *Word Origins and How We Know Them: Etymology for Everyone*. New York: Oxford University Press.

Litosseliti, L (ed.) (2010) *Research Methods in Linguistics*. London and New York: Continuum.

Lum, C (2013) *Scientific Thinking in Speech and Language Therapy*. New York and Hove: Psychology Press.

Lyons, J (1977) *Semantics*. Cambridge: Cambridge University Press.

Machin, D (2007) *Introduction to Multimodal Analysis*. London: Hodder Education.

Magnus, M (1999) *Gods of the Word: Archetypes in the Consonants*. LaVergne, TN: Createspace.

Magnus, M (2001) "What's in a Word? Studies in Phonosemantics". University of Trondheim: PhD thesis.

Mansfield, P (2004) "Accepting what we can learn from advertising's mirror of desire". *British Medical Journal*, 329: 1487-1488.

Marchand, H (1966) *The Categories and Types of Present-Day English Word-Formation*. Alabama: University of Alabama Press.

McClendon, M J and Alwin, D F (1993) "No opinion filters and attitude measurement reliability". *Sociological Methods and Research*, 21: 438-464.

McGregor, W (1996) "Sound symbolism in Gooniyandi, a language of Western Australia". *Word* 47: 339-364.

McGregor, W (2001) "Ideophones as the source of verbs in Northern Australian Languages". In Voeltz, F K E and Kilian-Hatz, C (eds.) *Ideophones*. Amsterdam and Philadelphia: John Benjamins.

McIntyre, S H and Ryans, A B (1977) "Time and accuracy measures for alternative multidimensional scaling data collection methods: Some additional results". *Journal of Marketing Research*, 14: 607-610.

McKillup, S (2012) *Statistics Explained: An Introductory Guide for Life Scientists*. Cambridge: Cambridge University Press.

Meester, R (2008) *A Natural Introduction to Probability Theory*. Basel, Boston and Berlin: Birkhäuser Verlag AG.

Meier, H H (1999) "Imagination by Ideophones". In Fischer, O and Nänny, M (eds.) *Form Miming Meaning*: *Iconicity in Language and Literature.* Amsterdam: John Benjamins.

Meithe, T D (1985) "The validity and reliability of value measurements". *Journal of Personality*, 119: 441-453.

Meyer, P G (2005) *Synchronic English Linguistics: An Introduction*. Tübingen: Narr Francke Attempto Verlag GmBH + Co.

Meyer, C F and Nelson, G (2006) "Data Collection". In Aarts, B and McMahon, A (eds.) *The Handbook of English Linguistics*. Oxford, Malden and Victoria: Blackwell.

Mill, J S (1843) *A System of Logic: Ratiocinative and Inductive*. London: John W Parker.

Morrow, J R Jr., Jackson, A W, Disch, J G and Mood, D P (2011) *Measurement and Evaluation in Human Performance*. Champaign, IL; Windsor, ON; Lower Mitcham, Torrens Park and Leeds: Human Kinetics.

Mukařovský, J. (1970) “Standard and Poetic Language” in Freeman, D (ed.) *Linguistics and Literary Style*. New York: Holt, Rinehard and Winston.

Munson, J M and McIntyre, S H (1979) “Developing practical procedures for the measurement of personal values in cross-cultural marketing”. *Journal of Marketing Research,* 16: 48-52.

Murphy, M L and Koskela, A (2010) *Key Terms in Semantics*. London and New York: Continuum.

Myers-Shaffer, C (2000) *The Principles of Literature*. New York: Barron’s.

Neidell, L A (1972) “Procedures for obtaining similarities data”. *Journal of Marketing Research,* 9: 335-337.

Nöth, W (1995) *Handbook of Semiotics*. Bloomington, Indiana: Indiana University Press.

Nöth, W (2001) “Semiotic foundations of iconicity in language and literature”. In Fischer, O and Nänny, M (eds.) *The Motivated Sign: Iconicity in Language and Literature 2.* Amsterdam: John Benjamins.

O’Grady, G N (2013) *Key Concepts in Phonetics and Phonology*. Basingstoke: Palgrave Macmillan.

Oppenheim, A N (1966) *Questionnaire Design and Attitude Measurement*. London: Heinemann.

Osterlind, S J (2002) *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance, and Other Formats*. Dordrecht: Kluwer Academic Publishers.

Partington, A (1998) *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. Amsterdam: John Benjamins.

Pearce, M (2007) *The Routledge Dictionary of English Language Studies*. Abingdon: Routledge.

Peirce, C S (1873) “On the Nature of Signs”. In Peirce, C S and Hoopes, J (ed.) (1991) *Peirce on Signs: Writing on Semiotic*. Chapel Hill: University of North Carolina Press.

Pierrehumbert, J B (2001) “Stochastic Phonology”. *Glot International* 5, 6: 195-207.

Pierrehumbert, J B (2003) “Probabilistic Phonology: Discrimination and Robustness”. In Bod, R, Hay, J and Jannedy, S (eds.) *Probabilistic Linguistics*. Cambridge (MA) and London: The MIT Press.

Pinker, S (1995) *The Language Instinct*. London: Penguin.

Pinker, S (1999) *Words and Rules*. New York: Basic Books.

Plato (2009) *Cratylus*. Translated from Greek, by Jowett, B. Rockville, MD: Serenity Publishers, LLC.

Podesva, R J and Sharma, D (eds.) (2013) *Research Methods in Linguistics*. Cambridge: Cambridge University Press.

Radden, G and Panther, K-U (eds.) (2004) *Studies in Linguistic Motivation*. Berlin: Mouton de Gruyter.

Rambaud, M G (2012) *Basic Semantics*. Madrid: Universidad Nacional de Educación a Distancia.

Rankin, W L and Grube, J W (1980) "A comparison of ranking and rating procedures for value system measurement". *European Journal of Social Psychology*, 10: 233-246.

Reay, I E (2006) "Sound Symbolism". In Allan, K (ed.) *The Concise Encyclopedia of Semantics*. Oxford: Elsevier.

Reynolds, T J and Jolly, J P (1980) "Measuring personal values: An evaluation of alternative methods". *Journal of Marketing Research*, 17: 531-536.

Roach, P (2001) *Phonetics*. Oxford and New York: Oxford University Press.

Roget, P M (2004) *Roget's Thesaurus*. London: Penguin.

Rosch, E (1975) "Cognitive Representations of Semantic Categories". *Journal of Experimental Psychology: General* 104: 192-234.

Rosch, E and Mervis, C B (1975) "Family Resemblances: Studies in the Internal Structure of Categories". *Cognitive Psychology* 7: 563-605.

Rovai, A P, Baker, J D and Ponton, M K (2014) *Social Science Research Design and Statistics*. Chesapeake, VA: Watertree Press LLC.

Russell-Bernard, H and Ryan, G W (2010) *Analysing Qualitative Data: Systematic Approaches*. Thousand Oaks, CA: Sage Publications, Inc.

Sadowski, P (1999) "Sound-Symbolic Vocabulary of Light in *Pearl*". In Lucas, P J and Lucas, A M (eds.) *Middle English from Tongue to Text*. Frankfurt, Berlin, Bern, Brussels, New York, Oxford and Wien: Peter Lang.

Sadowski, P (2001) "The sound as an echo to the sense: The iconicity of English *gl-* words". In Fischer, O and Nänny, M (eds.) *The Motivated Sign: Iconicity in Language and Literature 2.* Amsterdam: John Benjamins.

Sahu, C (2013) "History of Technology in Research". In Sappleton, N (ed.) *Advancing Research Methods with New Technologies*. Hershey, PA: Information Science Reference.

Salter, M and Bousfield, J (2002) *Poland*. London: Penguin.

Sanders, C (2004) *The Cambridge Companion to Saussure*. Cambridge: Cambridge University Press.

Sapir, E (1929) "A Study in Phonetic Symbolism". In Sapir, E and Mandelbaum, D G (1985) *Edward Sapir: Selected Writings in Language, Culture and Personality*. Berkeley, Los Angeles and London: University of California Press, Ltd.

Saussure, F de (1966) *Course in General Linguistics*. New York, Toronto and London: McGraw-Hill Book Company.

Schilling, N (2013) "Surveys and Interviews". In In Podesva, R J and Sharma, D (eds.) (2013) *Research Methods in Linguistics*. Cambridge: Cambridge University Press.

Schmidt, H-J (2005) "Spreading Contextualisation: English Lexicology in Germany between 1970 and 2000". In Kohl, S (ed.) *Anglistik: Research Paradigms and Institutional Policies 1930-2000*. Trier, Germany: WVT Wissenschaftlicher Verlag Trier.

Seager, W (1983) "Probabilistic Semantics, Identity and Belief". *Canadian Journal of Philosophy* 13, 3: 353-364.

Sereno, J A (2006) "Phonosyntactics". In Hinton, L, Nichols, J and Ohala, J J (eds.) *Sound Symbolism*. Cambridge: Cambridge University Press. p.263-275.

Shakespeare, W (2000) *Romeo and Juliet*. Ware, Hertfordshire: Wordsworth Classics.

Shanker, S (2002) "The Broader Implications of Borderline Areas of Language Research". In Fox, R G and King, B J (eds.) *Anthropology beyond Culture.* Oxford and New York: Berg.

Shih, T H and Fan, X (2008) "Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis". *Field Methods*, 20: 249-271.

Shinohara, K and Kawahara, S (2010) "A Cross-Linguistic Study of Sound-Symbolism: The Images of Size". *Proceedings of the Thirty-Sixth Annual Meeting of the Berkeley Linguistics Society.*

Shipley, J T (1977) *In Praise of English: The Growth and Use of Language*. New York: Net Work Times Book Co.

Short, M (1996) *Exploring the Language of Poems, Plays and Prose*. Harlow: Pearson Education Limited.

Simon, H A (1957) *Models of Man – Social and Rational*. New York: John Wiley and Sons.

Sinclair, J (2004) *Trust the Text: Language, corpus and discourse*. London and New York: Routledge.

Smith, D V L and Fletcher, J H (2004) *The Art and Science of Interpreting Market Research Evidence*. Chichester: John Wiley and Sons Ltd.

Stewart, D W and Shamdasani, P M (2015) *Focus Groups: Theory and Practice*. Los Angeles, London, New Delhi, Singapore and Washington, DC: Sage Publications, Inc.

Stoop, I, Billiet, J, Kock, A and Fitzgerald, R (2010) *Improving Survey Response: Lessons Learned from the European Social Survey*. Chichester: John Wiley & Sons, Ltd.

Strinati, D (2004) *An Introduction to Theories of Popular Culture*. London and New York: Routledge.

Sturtevant, E H (1947) *An Introduction to Linguistic Science*. New Haven, CT: Yale University Press.

Swan, O E (2002) "The Polish Alphabet and Sounds". Available online at http://polish.slavic.pitt.edu/firstyear/sounds.pdf, accessed 17.03.2014.

Tabakowska, E (1999) "Linguistic Expression of Perceptual Relationships: Iconicity as a Principle of Text Organisation (A Case Study)". In Fischer, O and Nänny, M *Form Miming Meaning: Iconicity in Language and Literature*. Amsterdam: John Benjamins.

Tavakoli, H (2012) *A Dictionary of Research Methodology and Statistics in Applied Linguistics*. Tehran: Rahnama Press.

Taylor, D G (2015) *The Mathematics of Games: An Introduction to Probability*. Boca Raton, FL: CRC Press.

Taylor, J R and Kinnear, T C (1971) "Empirical comparison of alternative methods for collecting proximity judgments". *Proceedings of the Fall Conference of the American Marketing Association*. p.547-550.

Thompson, G and Ye, Y Y (1991) "Evaluation in reporting verbs used in academic papers". *Applied Linguistics*, 12: 365-382.

Tourangeau, R, Conrad, F G and Couper, M P (2013) *The Science of Web Surveys*. Oxford: Oxford University Press.

Tranel, B (1987) *The Sounds of French: An Introduction*. Cambridge: Cambridge University Press.

van Peer, W, Hakemulder, F and Zyngier S (2012) *Scientific Methods for the Humanities*. Amsterdam and Philadelphia: John Benjamins.

Voeltz, F K E and Kilian-Hatz, C (eds.) (2001) *Ideophones*. Amsterdam and Philadelphia: John Benjamins.

Wagner, E (2010) "Survey Research". In Paltridge, B and Phakiti, A (eds.) *Continuum Companion to Research Methods in Applied Linguistics*. London and New York: Continuum.

Wales, K (1990) "Phonotactics and Phonaesthesia: The Power of Folk Lexicology". In Ramsaran, S (ed.) *Studies in the Pronunciation of English: A Commemorative Volume in Honour of A.C.Gimson*. London: Routledge.

Wallis, J (1653) *Grammatica linguae anglicanae*. Oxford.

Weathington, B L, Cunningham, C J L and Pittenger, D J (2010) *Research Methods for the Behavioural and Social Sciences*. Hoboken, NJ: John Wiley and Sons, Inc.

Wijk, A (1966) *Rules of Pronunciation for the English Language*. London: Oxford University Press.

Williams, G (2014) "Glossing Over the Lamb: Phonaesthetic GL- in Middle English and Aural Scepticism in *Pearl*". *The Review of English Studies* 65, 271: 596-618.

Wilson, A N (2006) *Betjeman*. London: Arrow Books.

Wintle, J (2002) *Makers of Modern Culture*. London, New York and Canada: Routledge.

Wray, A and Bloomer, A (2013) *Projects in Linguistics and Language Studies*. Abingdon and New York: Routledge.

Zuraw, K R (2000) *Patterned Exceptions in Phonology*. PhD thesis: University of California.