Corpus onomasiology: A study in World Englishes

SETH MEHL

UCL

PhD

I, Seth Mehl, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

ABSTRACT

This interdisciplinary work bridges corpus linguistics, lexical semantics, and World Englishes. Three methodological arguments are forwarded. First, only via a careful attention to the nature and type of corpus frequency measurements can we derive meaningful information from corpora. This is particularly relevant when comparing corpus frequencies to quantitative data derived via other means, such as lab-based cognitive salience tests. Second, an onomasiological approach is an extremely effective method in corpus linguistics, particularly in studies of lexical semantics. Third, semantic research based on corpora is essential to the study of language in use, including the study of World Englishes. Building on these methodological arguments, this work analyses lexical semantic variation in three highly polysemic verbs – *make*, *take*, and *give* – in the International Corpus of English, representing Singapore, Hong Kong, and Great Britain. This microlinguistic analysis demonstrates the value of the three methodological arguments. The study uncovers previously undiscussed onomasiological regional variation in written and spoken linguistic norms, and concludes that the three verbs exhibit degrees of delexicality which are consistent across regions. These findings challenge some established theoretical frameworks for World Englishes, and the impact for World Englishes as a field is discussed. The study establishes a link between: onomasiological corpus frequencies of each of the words' senses in speech; and reported cognitive salience measurements for those senses.

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

ACKNOWLEDGEMENTS

1. INTRODUCTION

The present work bridges corpus linguistics, lexical semantics, and World Englishes. The study has multiple aims, both methodological and experimental. Methdologically, the study aims to forward three arguments. First, only via a careful attention to the nature and type of corpus frequency measurements can we derive meaningful information from corpora, and in turn compare that corpus information to other types of information. Second, an onomasiological approach is an extremely effective method in corpus linguistics, particularly in studies of lexical semantics. Third, semantic research based on corpora is essential to the study of language in use, including the study of World Englishes. Building on these methodological arguments, the present work analyses lexical semantic variation in three highly polysemic verbs – *make*, *take*, and *give* – in the International Corpus of English, representing Singapore, Hong Kong, and Great Britain. The corpus analysis can be seen as a demonstration of the methodological arguments, and a test of their efficacy. Experimentally, this microlinguistic analysis aims to expand linguistic knowledge of these particular verbs, their semantics, and the three varieties of English. Furthermore, it aims to apply those findings in order to test, complement, and expand existing frameworks of World Englishes. Finally, it aims to investigate the relationship between corpus frequencies of each sense of the polysemic lexical items and the cognitive salience of those senses as measured in Cognitive Linguistic studies.

This work is, therefore, quite interdisciplinary. That said, I would like to situate the present research first and foremost as a corpus study. I approach the study of lexical semantics via corpus linguistics; I identify as a corpus linguist first, and specifically as a corpus linguist studying lexical variation. The nature of my research questions and the stepwise methodologies through which I address those questions reflect this identification. Corpus linguistics tends to be a quantitatively-oriented field, and as such, the nature of *frequencies* in corpora is an extremely important methodological subject in the present work. As a methodological exploration of corpus linguistics, the present work seeks to

precisely define various approaches to *frequency*, and to evaluate the efficacy of each of those approaches. I argue that *frequency* is generally under-defined and under-examined in the field. Using corpora, I extrapolate quantitative semantic usage patterns and selection preferences amongst sampled populations of language users in three different geographical regions. In particular, I analyse semasiological exposure rates – *i.e.* the rate at which a listener or reader can expect to encounter a given sense of a word – and onomasiological selection processes – *i.e.* the observed probability at which a speaker or writer selects one semantic alternate over another given the need to express a given concept. Most fundamentally, I argue that an onomasiological approach allows us to observe, in a logically sound and mathematically verifiable way, linguistic norms in use.

Arising from this corpus work in lexical semantics, naturally, are inquiries about the mechanisms of semantic similarity and variation between datasets like the ICE corpora. Thus, both theories of World Englishes and theories of Cognitive Linguistics have informed the experimental design in the present study, and the findings relate to such theories as well. Nevertheless, the approach here, to reiterate, is primarily corpus linguistic and lexical, and only secondarily sociolinguistic and cognitive. The different perspectives offered by corpus studies, lexical semantic research, World Englishes research, and Cognitive Linguistics research each shape the types of research questions scholars ask and the directions taken in answering them. The present work critiques the limited role of lexical semantics in existing frameworks of World Englishes, and argues that research in World Englishes includes as much socio-political work as linguistic work. Ultimately, language, society, and politics are not absolutely separable, but the present study is primarily linguistic, and the present research questions and experimental design are founded on linguistic ideas more than social or political ones. In light of that approach, the present research examines how lexical semantic findings in the ICE corpora, *i.e.* naturally occurring data representing varieties of English worldwide, may inform frameworks of World Englishes. I argue that semantic study should be an essential component of linguistic work in World Englishes, and that

increased semantic study in World Englishes can inform and enhance the existing body of World Englishes research. Similarly, the present work offers an opportunity to explore and discuss the importance of corpus frequency data in Cognitive Linguistics. The present work seeks to apply insights and findings from corpus work to Cognitive Linguistic ideas about lexical semantics, in order to enhance and expand discourse in Cognitive Linguistics. The present study, with its methodological focus on the nature of corpus frequencies, aims to inform an ongoing debate in Cognitive Linguistics on the relationship between corpus frequencies and cognitive salience.

Specifically, I ask the following research questions:

a. Do the high-frequency, polysemic verbs *make*, *take*, and *give* vary in their semantic categories; their semasiological frequencies; or their onomasiological frequencies between ICE-Singapore (ICE-SIN), ICE-Hong Kong (ICE-HK), and ICE-Great Britain (ICE-GB)?

b. Do existing findings on the cognitive salience of the multiple senses of *make*, *take*, and *give* correlate to the corpus frequencies of the multiple senses of these verbs, given multiple means of measuring corpus frequencies?

c. Do the similarities and differences in the lexical semantic behaviour of *make*, *take*, and *give* in ICE-SIN, ICE-HK, and ICE-GB corroborate established categories for World Englishes varieties?

The first question can be seen as micro-linguistic, a basic question about the precise details of the semantics of three words in use, and their potential variation in three regions. The three verbs have been selected because of their high polysemy and high frequnecy in use, factors which should allow for a great deal of nuance in semantic analysis. The three varieties have been selected to represent different categories of World Englishes according to established theoretical frameworks of World Englishes (see 5.1). Expanding linguistic knowledge of the three verbs and the three varieties is an important research

aim. The second research question offers significant impact potential in Cognitive Linguistics: if corpus frequencies can in fact be seen to correlate with cognitive salience, then the present study moves one step closer towards resolving a major issue in Cognitive Linguistics literature, *i.e.* the apparent disparity between corpus frequencies and cognitive salience. Finally, the third research question contributes to the broader theoretical discussion of World Englishes: to corroborate or refute existing theories is an important step towards the establishment of a theoretical framework with ever greater integrity. These research questions are investigated in detail in Part 2.

Part 1 forwards crucial methodological and theoretical arguments, as summarized above. Those arguments do not constitute a sort of preliminary scene-setting for the corpus study that follows, but are rather essential arguments in their own right. Chapter 2 argues that the nature of corpus frequency measurements must be carefully examined in corpus research, and onomasiological frequency data is presented as a sound choice, particularly for lexical semantic research. Chapter 3 presents and supports Geeraerts's (2010) hypothesis of onomasiological salience as a means of linking onomasiological corpus frequencies with findings on cognitive salience, and also argues that traditional polysemy tests can be adapted to produce a useful heuristic tool for identifying polysemy in corpus data. Chapter 4 argues that semantic research is essential to World Englishes study, just as research in regional variation is essential to semantic study.

The corpus study in Section 2 is, therefore, not only an investigation into the three lexical items in the three corpora, but equally importantly a testing ground for the methodological and theoretical assertions in Section 1. Chapter 5 summarizes the methods employed in the present study, building from the methodological arguments in Section 1. I distinguish *methodology* in Section 1 from *method* in Chapter 5 in the traditional sense: *method* is a description of the actual practices and approaches employed; *methodology* is the study of methods, a discourse about practices and approaches, including an evaluation of the pros and cons of various methods, and a justification for the

methods employed here. Methodology can also be seen as relating methods to the field of epistemology, asking how it is that a particular approach allows us to know a particular thing. This is especially important in the discussions of various measures of frequency in Chapter 2 and in discussions of polysemy tests as applied to corpus data in Chapter 3. Findings for *make*, *take*, and *give* are presented in Chapters 6, 7, and 8, respectively, and concluding arguments follow in Chapter 9.

PART 1: THEORETICAL AND METHODOLOGICAL ARGUMENTS

2. CORPUS LINGUISTICS

2.1. Why corpora?

To begin, it is useful to review the types of data used in linguistics. Meyer and Nelson (2006) propose a useful framework of three types of data collection: introspection, experimentation, and corpus building. Arppe *et al*. (2010), not dissimilarly, discuss two types of data collection, referring to 'found data' rather than 'corpus building' and 'elicitation' rather than 'experimentation'. In Meyer and Nelson's (2006) first type of data collection, introspection, the researcher invents linguistic examples and then evaluates their acceptability, grammaticality, meaning, and so on via reflection. In this situation, it is a reflective process that serves as the basis not only of data collection but also of data analysis. It might be possible to distinguish *introspection* as a reflective process involving critical thinking skills or some particular method or methods of reflective thought, from *intuition*, which might be a more emotive, immediate evaluation of the acceptability or meaning of a particular example, but Meyer and Nelson seem to use the two terms interchangeably, and such interchangeability is common in the field, as I show below. Employing some kind of reflective process in data analysis is likely inevitable, as I discuss further in this section – some steps in analysis must be executed via clearly defined, methodically sound reflection and critical thinking – but relying on reflection for data collection raises problems. In the words of Sinclair (1991: 39), such reflective methods are useful 'in evaluating evidence rather than creating it'. I generally employ the term *subjective reflection* to refer to introspective or intuitive methods that are not clearly defined or explicitly justified. As Meyer and Nelson (ibid: 97) point out, 'linguists who rely on their own intuitions for data often produce theories of language that are reflective of their own idiolects'. Butterfield and Krishnamurthy (2000; quoted in Krishnamurthy 2000: 147) explain:

> ...each of us has only a partial knowledge of the language, we have prejudices and preferences, our memory is weak, our imagination is powerful (so we can conceive of possible contexts for the most implausible utterances), and we tend to notice unusual words or structures but often overlook ordinary ones.

Indeed, due to the obvious limits of such subjective reflective processes for data collection in particular, it is difficult to imagine any other contemporary scientific field that would accept personal reflection as a foundation for both data collection and data analysis.[1]

In response to the problems raised by subjective reflection as a data collection method, linguists have in many cases turned to experimentation. Experimentation can take multiple forms: a good deal of experimentation is designed to collect linguistic examples and judgments about acceptability, grammaticality, or meaning from a large set of subjects, or a small, representative set of subjects, rather than from the linguist alone. In many effective experiments, subjects are unaware of exactly what element of linguistics is being tested. Regardless of the subjects' awareness of the research question, experimentation as a data collection tool is hindered by the 'Observer's Paradox' (Labov 1972: 209). That is, experimental subjects' production and evaluation of language does not occur under natural circumstances, but under observed conditions in an experimental setting. Well-designed experiments aim to minimize this problem, but it cannot be eliminated completely in an experimental context (within contemporary ethical constraints).

---

[1] The reliance on reflective processes in generative grammar is based upon a unique view of the object of study. In generative grammar, the object of study is explicitly and emphatically the mental phenomenon of language rather than language in use. That said, contemporary psychological sciences, including psycholinguistics, also study mental phenomena, and do not rely on reflection for data collection. A full discussion of linguistic methodology in light of generative grammar is, however, beyond the scope of the present work.

Corpus linguists collect examples of language in natural use, usually from large numbers of language users. A corpus can be concisely defined as a sample of language in use, purposely collected to be representative of that language in a particular context or contexts (*cf*. McEnery *et al.* 2006: 4-5; McEnery and Wilson 2001: 29-32). Corpus linguistics eliminates the problems generated by collecting data via reflection and minimizes or eliminates the observers' paradox (*cf*. Meyer and Nelson 2006): corpora contain examples of written and/or spoken language in use, most often in entirely natural settings, and the examples are generally compiled after the language has already been produced for an original purpose separate from linguistic research. For example, the ICE corpora used in the present study include transcripts of courtroom proceedings as well as excerpts from newspapers (see 5.2 for a full discussion of the ICE corpus design). Hanks (2000: 211) describes corpora as 'traces and patterns of linguistic behaviour' for linguistic study. In that sense, corpus linguistics is an *observational science* as opposed to an *experimental science* (*cf*. Wallis 2014a): that is, like astronomers, geologists, or, often, epidemiologists, corpus linguists observe naturally occurring phenomena as a body of data but they do not *ex post facto* re-elicit the data or re-experiment with the source of the data. Corpus linguistic methods allow us to measure and quantify linguistic behaviours - or traces, patterns and linguistic features - which no language user is consciously aware of during either production or reception, features that occur outside of language users' conscious attention, including features that may be so common as to go unnoticed. Corpora allow us to move beyond marked features and apparent language trends, and therefore beyond subjective reflection, partial knowledge, prejudices, preferences, weak memories, powerful imagination and unusual words or structures, to otherwise unnoticed features of actual language use. In the study of World Englishes, this type of approach is under-utilized; as I argue in Chapter 4, the sociolinguistic approaches that tend to dominate World Englishes research generally depend on

someone (the linguistic researcher or an informant) being acutely aware of the social and cultural weight or markedness of a linguistic feature.[2] The study of corpora allows a researcher to investigate features that no one could be aware of otherwise. This unique advantage of corpus linguistics is particularly crucial in the present study, which moves beyond immediately obvious 'lexical exotica' to subtle variation in the semantics and usage preferences of high-frequency, polysemic lexis.

Aarts (1991) presents an insightful contrast between 'intuition', or subjective reflection, and corpora as sources of linguistic evidence. Subjective reflection can provide examples like the following:

> My grandfather's grandfather's father's grandfather's grandfather's father's grandfather's father's grandfather's grandfather's father's father was an Indian.
>
> (Ziff 1974: 528, quoted in Aarts 1991: 48)

This sentence is grammatically error-free, but we can reasonably predict that no one will ever use it (or understand it). Aarts (ibid: 48) contrasts that example with an example from the Nijmegen corpus of spoken English:

> And what a performance by the man who some of us thought that maybe the pressure of being the favourite of Wimbledon might not let him win.

Constructions such as the above may occur regularly, but seem to be difficult to invent via subjective reflection, and are by general standards seen to break the rules of English. Aarts's example only reinforces Krishnamurty's (2000; quoted above) argument on the limits of subjective reflection.

---

[2] Leimbgruber's (2013) recent study of Singapore English is a good example of the possibilities and the limitations of studying features that are consciously accessible in World Englishes.

Thus, corpora can clearly and directly answer research questions related to what people actually do with language, consciously or not, and sort out *imaginable* linguistic events from actually realized linguistic events. For the present research questions, this distinction is crucial: the present study investigates very subtle differences and similarities in actual use of English in three corpora representing three regions of the world.

As is often the case with corpus linguistics, collection and analysis of features in the present study is largely quantitative. In fact, corpora are unique in that they can provide evidence for frequencies of occurrence of particular linguistic features in particular contexts – information that no other data type can reliably provide (*cf*. Arppe *et al*. 2010; Gilquin 2006). As McEnery and Wilson (2001: 15) point out, data on frequency in use is not reliably available via subjective reflection. Frequencies of linguistic features are crucial to cognitive accounts of language, and therefore the degree to which corpora provide evidence for cognition and cognitive functioning, or the degree to which corpora are useful tools in cognitive linguistics, are important and contested questions (*cf*. Taylor 2012; Gries 2012; Arppe *et al*. 2010; Geeraerts 2010; Gilquin 2008; Heylen *et al*. 2008; Hundt 2007; Gilquin 2006; Gries 2006a, 2006b; Geeraerts 1997; Geeraerts 1988; see also 3.4 and 9.3). As Taylor (2012: 148) summarizes, 'frequency influences performance on all manner of experimental tasks' related to the psycholinguistics of language production and reception. It is therefore extremely worthwhile to examine the nature of frequencies in corpora – what is a corpus frequency and how do we measure it? – a question which is, in fact, underexamined. A thorough analysis of the nature of frequency measurements in corpora is conducted in 2.2.

If corpora are valuable, in part, because they allow researchers to avoid reliance on their own subjective reflection for data collection, then the question arises: Can reflection of any kind be eliminated completely from linguistic analysis? As with any scientific endeavour undertaken by human beings, the answer must be 'no'. As noted above, reflective processes and critical thinking are necessary in data analysis. Marianne Hundt (2007: 4-5) proposes a model of

corpus linguistic research in which the linguist's 'intuition' or subjective reflection is, explicitly, 'another sort of data' that complements corpus data. In Hundt's framework, reflective processes constitute a constant and inevitable input to the linguist's analyses. McEnery and Wilson (2001: 19) concur, stating that 'corpus linguistics is, and should be, a synthesis of introspective and observational procedures'. Nonetheless, in Hundt's (2007) study of mediopassives (ibid.), she minimizes dependence on reflection by designing her research questions and her experiments to rely on corpus data more than on subjective reflection for data collection. The present study does the same. It is also important to distinguish subjective reflection for data collection from rigorously defined reflective, critical thinking used for data analysis. That is, we can reduce subjective reflection throughout the process of linguistic research, particularly in data collection, while explicitly employing rigorous reflection in the process of applying logical and critical thinking skills in analysing data. Moreover, I would assert that findings and conclusions based on corpus research can be usefully corroborated or refuted via experimental observation of subjects. Experimentation and corpora complement each other extremely well.

Corpus linguistics has been linked to the study of English varieties worldwide from a very early stage, and the present study continues that tradition. The Brown Corpus of American English, compiled in 1968 (*cf*. Francis and Kučera 1982), is widely considered the first contemporary corpus for linguistic research (*cf*. McEnery *et al*. 2006: 3). By 1978, the Kolhapur Corpus of written Indian English was completed by researchers 'consciously attempting to construct… a corpus of distinctively Indian English' (Nelson 2006:735). By 1982, Hofland and Johansson had compared the Brown Corpus of American English to the LOB corpus of British English, drawing conclusions about differences and similarities between the two varieties. In 1991, the International Corpus of English project was initiated (Greenbaum 1996), and this provides the data for the present research. To date, the ICE project has produced parallel corpora for fourteen varieties of English (The ICE Project 2014, http://ice-corpora.net/ice/). The ICE corpora have led to a huge

increase in studies into World Englishes, as data on varieties is now readily available to linguists worldwide. (For a detailed discussion of the value and limitations of the ICE corpora, see 5.2).

2.2. Experimental design and statistics: The nature of frequencies

Corpus linguistic research is predominantly quantitative, insofar as it measures frequencies of various linguistic features in natural use. Frequency data is not an end in itself, but it is one sort of data that can uniquely be derived from corpora, and corpus linguists tend to rely on frequency data to draw broader conclusions about language. Wallis (2014a) identifies three classes of evidence that can be derived from corpora: factual evidence, frequency evidence, and interaction evidence. Factual evidence is evidence that a linguistic feature *can* occur, or has occurred at least once. When researchers use corpora as sources for occasional examples (*cf.* Aarts 2011), they are presenting factual evidence. Frequency evidence is evidence of the frequency of an event in a corpus. Interaction evidence is evidence for the observed probability of a feature, given the presence of some other feature. The present study, like much of the work in corpus linguistics, can be seen as gathering frequency evidence and interaction evidence.

As Arppe *et al*. (2010: 7) note, corpus linguistics poses 'questions that can be answered through the observation of (relative) frequencies of occurrence. Such data can then yield generalizations about questions of natural language use'. It is therefore necessary to define the nature of *frequencies*. Arppe *et al*. (ibid: 7) are careful to note that corpora can be used to derive '(relative) frequencies', but this term requires a bit of analysis – *relative* to what? The nature of frequencies is rarely discussed in the field. This section begins with a discussion of probability theory and statistics, much of which is derived from Wackerly (2008) and Sheskin (2003).[3] That discussion establishes

---

[3] My approach to statistical analysis in corpus linguistics has also developed thanks to numerous discussions with Sean Wallis, senior research fellow at the Survey of English Usage, University College London.

two important types of frequency measurements which are crucial to understanding the methods employed in the present study. It might have been assumed that a discussion of the basics of probability theory, statistics, and experimental design does not deserve a place in a work like this one. However, these concepts tend to be under-examined, which raises issues in the field of corpus linguistics, as evidenced in the studies discussed below. Moreover, the basics of experimental design are not necessarily simple to execute in any given experiment: the nature of frequency measurements and experimental design is unique to individual research questions, and even very skilled statisticians must wrestle with these basic (or fundamental) questions each time they perform statistical analyses.

As stated in 2.1, a corpus is a sample of language in use, purposely collected to be representative of that language in a particular context or contexts. As *samples*, corpora contain language that represents a *population* (*cf.* Wackerly 2008, Chapter 2, on samples and populations); that is, corpora contain a small sub-set of the full range of language produced in a particular context. ICE-Singapore (henceforth ICE-SIN), for example, contains 1 million words of language produced in Singapore, in the 1990s, by speakers and writers who matched specific criteria (see 5.2 for details on those criteria). That is, in the case of corpus linguistics, the sample is the corpus, drawn from a population of actual language in use. As an alternative example that serves as a useful parallel in the discussion below, a sample might be a series of 10 or 100 coin tosses, drawn from a suggested population of infinite coin-tosses. Researchers employ samples to draw conclusions about populations, often counting instances of particular features in a given sample in order to draw conclusions about the population from which the sample was drawn.

If a researcher has counted features in a given sample, *e.g.* linguistic features in a corpus, then what exactly can he or she conclude about the population? What information does a simple count actually convey? In fact, a raw count or a simple number does not convey very much at all. The nature of frequencies in samples and populations is more complex than simple numbers

or raw counts of instances of an event. To illustrate this point, an example situation commonly used in statistics pedagogy is useful here (*cf.* Wackerly 2008: 37). In this common example, a person flips a coin and reports that the coin landed heads-up 4 times. It is very difficult to interpret the import of that number if the person does not report how many times he or she flipped the coin. A coin that lands heads-up 4 times out of 5 flips suggests something very different than a coin that lands heads-up 4 times out of 50 flips or 500 flips. The total number of flips in this case is called a *statistical baseline* (*cf.* Bowie *et al.* 2013). The statistical baseline provides a relative measuring stick by representing the total number of possible instances of an event, against which to interpret a raw number of actual instances. That is, a frequency can be presented as a ratio of actual realizations of an event to possible realizations of the event. The ratio of actual instances to the possible number of instances is a *probability*, and more specifically an observed probability (as opposed to a prediction, for example, about a sampled population, or about some future sample or population). That probability can range from 0% to 100%, but can never be less than 0% or more than 100%. This conception of frequency relations forms the foundation of probability theory, and probability theory in turn forms the foundation of statistical analysis (*cf.* Wackerly 2008). If a baseline is inappropriate or incorrect (or absent), then any statistical conclusions drawn from the quantitative data are likely to be unsound.

When we flip a coin, the baseline is, ultimately, easy to determine, and perfectly obvious: the baseline is the number of flips. When we count instances of linguistic features in a text, in corpus linguistics, we need a baseline as well, but linguistic baselines are more difficult to determine. If I report that I counted instances of *make* with a concrete Direct Object (as I do in section 6.2.2),[4] and that I found 50 instances, it is very difficult to interpret the import of that

---

[4] A concrete Direct Object represents something that can be directly observed by any of the five senses. The finer points of this distinction are discussed at length in 6.1, 6.2.2 (for *make*), 7.2.2 (for *take*), and 8.2.2 (for *give*).

number if I do not report anything about the text(s) from which instances of *make* were drawn. That is, it is very difficult to interpret the import of that number without establishing a baseline drawn from those texts. What is an appropriate baseline for measuring linguistic features in a corpus?

Perhaps the obvious and intuitive choice for a baseline, and certainly the choice which has often been employed, is the total number of words in the text(s) (for recent examples of that baseline in World Englishes lexical research, see Fuchs *et al.* 2013; Fuchs 2012; Yao and Collins 2012; Gilquin 2008; Hundt 2007; Lange 2007; Schneider 1994; a selection of which is discussed below). Interpreting the number of instances of a linguistic feature as a ratio of the number of words can be called a *per word* baseline, or a *per million words* (*pmw*) baseline,[5] and this baseline is the standard in multiple linguistics textbooks (*cf.* Lindquist 2009: 41-42; McEnery *et al.* 2006: 52-3; McEnery and Wilson 2001: 83), as well as in the reference resource *The Routledge handbook of corpus linguistics* (Evison 2010: 126). As a result, many corpus linguists learn this approach to analysing frequencies: the relative measuring stick, or baseline, against which to interpret a raw number of instances is, for many students as well as academic corpus linguists, the total number of words in the text or corpus. The total number of words in a text constitutes an obvious baseline for many reasons. Words are basic and obvious elements of language (*cf.* Hanks 2009). They are highly cohesive, or minimally divisible, in comparison to larger units such as phrases or clauses (Huddleston 1984: 37). They are also highly independent and mobile in relation to smaller units such as morphemes (*ibid*.: 38). They seem to be extremely salient units of language to the average language user (*cf.* Libben *et al.* 2012). But the sheer obviousness of

---

[5] In fact, a *per million words* baseline creates additional problems beyond that of a *per word* baseline in quantitative analysis insofar as it skews confidence intervals (*i.e.* margins of error; see footnote 28). For example, a count of 100 instances of a feature in a corpus of 100,000 words might be seen as equivalent to 1,000 instances *per million words*. This is problematic because the ratio 100/100,000 results in a very different confidence interval (or margin of error) from the ratio 1,000/1,000,000.

words as a baseline begs the question: are there other possible baselines? In fact, there are many other possible baselines, most of which are not discussed as potential options in corpus linguistic textbooks or in published studies. For example, linguistic features might be measured per morpheme, per verb phrase or noun phrase, or per clause, among other options. For the corpus linguist, it is necessary to ensure that the baseline is justified by the research question and the object of study. Different baselines will be more or less appropriate in different research contexts, for different research questions.

The *pmw* baseline can be seen as indicative of exposure: an *exposure rate* (p.c. Wallis). The *pmw* measurement relates to a very particular research question that might be stated roughly as follows: given that a reader or hearer is exposed to 1 million words of a language, how many times will the listener or hearer likely encounter the given feature? This is comparable to an exposure rate to *heads* or *tales* when flipping a coin: measuring features *pmw* is, in a very rough way, comparable to measuring how many times a coin lands heads up *per hour*, answering a research question like this: given that a viewer watches a coin being flipped for one hour, how many times will that viewer likely see the coin land heads up? With the coin, it is clear to see that this kind of exposure rate is an epiphenomenon of many different contributing factors: how quickly the coin can be flipped and therefore how many times the coin is flipped, as well as whether the coin is weighted, and so on. In linguistics, too, an exposure rate with a *pmw* baseline is an epiphenomenon of many different variables: the length of the verb phrases or clauses in a sample, and therefore the number of possible verb phrases or clauses in the sample, among other things. The *pmw* baseline can, however, be useful. If a lexicographer is interested in including only the most commonly encountered words in a particular dictionary – and the corpus represents the contexts and situations that a language user is likely to encounter – then a *pmw* measurement can give a general indication of those words. Additionally, if a grammarian is interested in including the most commonly encountered grammatical constructions in an introductory grammar, and the corpus is representative of the contexts and situations that a language

user is likely to encounter, then a *pmw* baseline could be useful as well. For corpus studies with precise research questions about exposure rates, a *pmw* baseline can occasionally be a justifiable choice – but there are alternatives which might be preferable for various reasons as well. A researcher could certainly measure exposure to a certain feature *per phrase* or *per clause*. Alternatively, other lexical baselines are available: two will be discussed below as particularly significant for lexical semantic research. Ultimately, the baseline should fit the research question, and should be explicit and justified.

The *pmw* measurement is useful in certain types of circumstances, but it is important to recognize that the *pmw* baseline is not an observed probability – it is instead a rate, another type of frequency measurement. Because it is not an observed probability, it is not useful for statistical analysis. To explain this, let us return to the coin-toss experiment. The logically pertinent point in the coin-toss experiment is that the frequency of actually realized instances is presented as a ratio over the number of possible instances. Mathematically, this means that as many as 100% of the instances may land heads up, or as few as 0% of the instances may land heads up. This fact is a defining factor of the probabilistic baseline itself: any statistical baseline must represent the possible number of occurrences of an event, and the actual number of occurrences may range from 0% to 100%. In a corpus, a baseline of the total number of words, or *pmw*, implicitly assumes that as few as zero words (0% of one million words) or as many as all words (100% of one million words) in the corpus could hypothetically be instances of the feature in question. This possibility is only very rarely, if ever, the case. For example, when counting instances of *make*, it is certainly possible that no words in the corpus will be instances of *make*. It is never, however, possible for all words in the corpus to be instances of *make*. Such a phenomenon is outside the bounds of language use. Each of the million words in a corpus cannot represent a possible instance of the word *make*. Rather than a probability, the *pmw* measurement can be seen as a rate, and I have argued that it represents an exposure rate. This fact is often not explicated fully in published corpus studies.

There are two important lexical baselines that allow for two very different kinds of probabilistic measures. The first considers a particular sense of a word in relation to all instances of that word (this baseline can be seen as *semasiological*, a term whose history and recent use are discussed further below). A semasiological baseline can be seen to answer the following research question: given that a listener or reader encounters lexical item *a*, what is the probability that it represents sense *x*? This question relates to exposure rates, but it is slightly different from the research question on exposure rates stated above for the *pmw* baseline.[6] A research question like this might be useful for a lexicographer deciding how to arrange senses for a lexical entry in a dictionary. This type of exposure rate might also be useful for a cognitive psychologist investigating the relationship between the frequencies of a word's meanings and, for example, recall speeds for various senses of the word. Perhaps most importantly for quantitative analysis, this baseline is probabilistic and therefore suitable for a statistical analysis. That is, it is entirely possible that a listener or reader is exposed to *x* instances of *make*, and that 0% of those instances represent a particular sense, or that 100% of those instances represent a particular sense.

In addition to the semasiological baseline, it is also possible to consider a particular sense of a word in relation to instances of other words that convey the same or nearly the same meaning (this baseline can be seen as *onomasiological*, a term whose history and recent use are discussed further below). This onomasiological baseline can be seen to answer a different type of research question: given that a speaker or writer is expressing concept *x*, what is the probability that the speaker or writer selects lexical item *a* to express it? This is also a probabilistic baseline: it is hypothetically possible, for example,

[6] Semasiological studies do not necessarily employ semasiological baselines. For example, a semasiological study might investigate the different meanings of a given word, and measure the frequencies of those meanings *pmw* rather than measuring those frequencies in relation to all instances of all senses of the given word. *Semasiology*, its meaning and history, are discussed further below.

that given the opportunity to describe 'creating a concrete thing', a speaker or writer might select *make* 100% of the time, or 0% or the time. This research question is a production-oriented question, reflecting a selection process, whereas the *pmw* ratio and the semasiological ratio above represent a reception-oriented question, or an exposure rate. In fact, this production-oriented approach has a unique advantage – it definitively and explicitly counts actual opportunities for the lexical item in question to be produced, and employs that count as a probabilistic baseline. That fact underlines a subtle shortcoming of the semasiological and *pmw* baselines: although a semasiological baseline is probabilistic, with a range of possibilities from 0% to 100%, the baseline does not represent opportunities for a lexical item to appear. This is because an instance of a word with one sense is not commonly an opportunity for an instance of that word with a different sense. For example, an instance of *crane* with the sense 'construction tool' is not an opportunity for an instance of *crane* with the sense 'bird'. Similarly, *take* in Example 1 is used to express the transfer of a concrete object.

1. We **took** only two cases of whiskey and twenty cartons of Japanese cigarettes. [ICE-HK W2B-002 #39]

Take in this example is not an opportunity for *take* with some other sense – *i.e.* this is not an opportunity for *take* with the sense 'assume/adopt' (see Chapter 5 for more on the semantics of *take*). It could, however, be an opportunity for some other word with a similar sense, such as *carry* or *transport*. This production-oriented onomasiological baseline is logically sound because it enumerates actual opportunities for a word (or other linguistic feature) to occur, whereas a reception-oriented semasiological baseline measurement does not.

Both reception and production are important parts of linguistic investigation, but the difference between the two types of frequency measurement is rarely discussed in corpus linguistics. The production-oriented onomasiological baseline of selection processes can be seen as primary:

exposure rates are a direct epiphenomenon of a selection process. That is, language must be produced by a speaker or writer in order for a listener or reader to be exposed to it. Put differently, a listener or reader is exposed directly to the produced language of others, *i.e.* to the result of a selection process. That said, both reception and production are essential to linguistic investigation; corpus researchers should consider which type of frequency information they are pursuing and why. In the following paragraphs, I briefly survey a selection of studies that employ a non-probabilistic *pmw* measurement. In general, I would like to show that although a *pmw* baseline can be useful for some purposes, there is a body of existing research that could be improved by a closer consideration of the nature of various frequency measurements and what those measurements represent linguistically. Because the number of corpus studies employing a *pmw* baseline is very large, I limit the present survey to recent studies of lexis in World Englishes. The example studies here are therefore particularly relevant and comparable to the present study. In general, it is apparent that the *pmw* ratio is often employed without an explanation of what exactly it represents in terms of receptive exposure rates.

Fuchs *et al*. (2013) aim to describe 'the genesis, nature, development and spread' of the innovative uses of focus particles in Nigerian English (*ibid*.: 125). Because this is not a precisely formulated research question, an appropriate probabilistic baseline for statistical analysis is difficult to determine; as is common in the field, *baselines* are not discussed explicitly in any way in the study. To achieve their aim, Fuchs *et al.* (*ibid*.: 125) analyse 'the usage of two focus particles, *even* and *still*' in ICE-Nigera and ICE-GB, and compare those usages to 'the usage of rough equivalents in Yoruba and Igbo' (*ibid*.: 125). The study then reports that '*still* and *even* are used with different frequencies in NigE and BrE… calculated as frequency per million words' (*ibid*.: 130). In turn, Fuchs *et al*. (*ibid*.: 130) report that '*even* is significantly

more frequent in NigE', determined by a log-likelihood test.[7] As I have explained above, statistical analyses (including the log-likelihood test) are only valid insofar as they are derived from probabilistic baselines. Fuchs *et al*.'s (ibid.) statistical analysis is flawed because it accepts a non-probabilistic baseline of *pmw*. That is, while the *pmw* rate can certainly provide information on exposure, it is not a probability, and therefore does not constitute a valid probabilistic baseline for a statistical test. Moreover, Fuchs *et al.* (*ibid*.) do not frame their research in terms of exposure rates, but only in terms of these *pmw* numbers. Interestingly, Fuchs *et al*. (*ibid*.) go on to describe usages of each sense of *still* and *even*, respectively, as a percentage of total use of *still* and *even*. Measuring actual usage as a percentage of total use can be seen as a probabilistic baseline rather than a rate like *pmw*, because it is limited to a probability range of 0% to 100%, and is therefore a sound mathematical basis for a statistical analysis. This analysis would answer the question: given that a language user encounters the words *still* or *even*, respectively, what is the probability that the word conveys sense *a*, *b*, *etc*., and does that probability vary significantly between ICE-Nigeria and ICE-GB? However, Fuchs *et* al. (ibid: 137) draw conclusions about *preferences* in the regions represented by the corpora. The term *preferences* is used vaguely, but would seem to relate to selection processes in production, such that one word might be preferred over another word to express a given concept. Indeed, it would seem that a speaker or writer does not have a preference between multiple senses of *even* and *still*, because multiple senses of *even* and *still* cannot be selected in place of each other. That is, multiple senses of *even*, for example, do not alternate with each other. *Preferences* would seem to relate to a choice between two roughly equal

---

[7] In the present work, I do not offer a thorough critique of the variety of statistical tests used by corpus linguists. Fuchs *et al*. (2013: 130) assert that the log-likelihood test 'has been shown to be more suitable than the chi-square test for the comparison of corpus counts', though that claim is not fully explained or cited. Wallis (2012) has shown that both the log-likelihood and the chi-square test can result in errors insofar as both tests can result in confidence intervals reaching to less than 0% and greater than 100% when comparing two data sets.

alternates. Production preferences are not what is examined in the study, and the possibility of measuring preferences in terms of exposure rates *pmw* is not explained in any way. This example illustrates, then, not only the importance of formulating a precise research question that can then be addressed using an appropriate analytical approach, but also some problems in statistical analysis.

Fuchs (2012) aims to 'take a closer look at *also*' and to corroborate or refute hypotheses previously presented by Balasubramanian (2009) – namely, that *also* is more frequent than *too* in Indian English, and that *also* occurs medially in Indian English less often than in BrE and AmE. Fuchs (*ibid*.) investigates ICE-GB and ICE-India. Like Fuchs *et al*. (2013), Fuchs (2012) does not outline a specific method for approaching frequencies in relation to a specific research question, simply reporting 'relative [per thousand words] and absolute frequencies of the senses of *also* in ICE-India and ICE-GB' (*ibid*.: 33). Fuchs (*ibid*.), however, does not apply statistical tests using *per thousand words* as a baseline (as Fuchs *et al*., 2013, did, above), and therefore does not commit a statistical error in relation to a non-probabilistic baseline. Fuchs (2012) can thus be seen to address two research questions like the following:

a. Given that a hearer or reader in ICE-GB or ICE-India is exposed to 1,000 words of text, at what rate is the hearer/reader exposed to *also* and *too*, and to sense *a*, *b*, *etc*. of each?

b. Given that the same hearer or reader is exposed to 1,000 words of text, at what rate is the hearer or reader exposed to *also* and *too* in medial or non-medial positions?

These are essentially interesting questions, but the precise usefulness of the *pmw* baseline is left unstated by Fuchs (*ibid*.), and it is difficult to derive why exactly the *pmw* baseline would be useful. It would likely be even more useful to ask the following:

a. Given that a hearer or reader in ICE-GB or ICE-India is exposed to *also* and *too*, at what rate is the hearer/reader exposed to each in relation to the other, and to sense *a*, *b*, *etc*. of each?

b. Given that the same hearer or reader is exposed to *also* and *too*, what is the probability that the hearer or reader is exposed to *also* or *too* in medial or non-medial positions?

These revised research questions offer the added benefit of being probabilistic, and therefore suitable for testing statistical significance.

Werner and Mukherjee (2012) is a study that is particularly relevant to the present research because it investigates 'highly polysemous verbs in New Englishes': specifically, *give* and *take* (two of the three verbs studied in the present work) in 230,000-word subsections of the written portions of ICE-Sri Lanka, ICE-India, and ICE-GB.[8] Frequencies of use for each sense of each polysemic verb are therefore given *per 230,000 words*. Again, this measurement can be seen as an exposure rate; it would be useful if that fact were stated clearly in the study. The stated aim of the study is to describe 'quantitative differences between the postcolonial varieties of Indian and Sri Lankan English and their historical input variety of British English with regard to the frequencies of the individual meanings of *give* and *take*' (*ibid*.: 253). The research addresses a question like the following: given that a listener or reader in ICE-Sri Lanka, ICE-India, or ICE-GB encounters one million words of text, at what rate does the listener or reader encounter each sense of *give* and *take*? Werner and Mukherjee (*ibid*.: 254) claim to have determined statistically significant differences between the varieties, but they do not explain their statistical analysis or the tests employed, and they have not presented any kind of probabilistc baseline. A more thorough description of their analysis here would be useful. In addition, Werner and Mukherjee (*ibid*.: 250) aim to

---

[8] Werner and Mukherjee (2012) is modelled after Gilquin (2008), which is discussed in section 2.2.3. Both studies are discussed at greater length in 7.3.

'establish a link between a prototype-theoretical categorization of the various meanings of high-frequency verbs and the corpus-based description of the frequencies of the individual meanings in actual use'. It is, as I have argued, extremely important to consider the nature of frequency measurements and to justify the type of measurement taken in such a discussion; their study would certainly be strengthened by explaining their choice of baseline and its relevance to prototype-theoretical frameworks.

Hundt (2009) is a study of lexico-grammar, investigating usage of *get* in passive constructions (*i.e.* the *get*-passive) in ICE-SIN, ICE-Philippines, and ICE-GB. The study tests the hypothesis that ICE-SIN and ICE-Philippines will display limited variation in usage preferences between written and spoken registers in comparison with ICE-GB (a hypothesis I address further in section 9.2), and also asks whether the *get*-passive is more common in ICE-SIN, ICE-Philippines, or ICE-GB. Hundt (*ibid*.: 125-6) presents her data in an effectively straightforward way, first counting *get*-passives as 'raw frequencies' and then using a baseline of *per 100,000 words*. To justify this decision, Hundt (*ibid*: 126) asserts that 'raw frequencies… do not allow for a straightforward comparison of spoken and written data, as the spoken and written parts of the corpora are not of equal size'. This explanation resembles that given in many corpus linguistics textbooks (described above); it is an attempt to recognize that frequencies must be measured in relation to some kind of baseline, but, like the textbooks already mentioned, there is no discussion of the array of possible baselines that might be employed, nor a justification for the *pmw* baseline. Hundt does propose, wisely, that future studies of the *get*-passive in the ICE corpora should be conducted onomasiologically, *i.e.* that *get*-passives should be counted as a proportion of total passives, for more meaningful results (see 2.3). Counting *get*-passives as a proportion of total passives would constitute a probabilistic baseline and would allow for statistical analysis. Moreover, that probabilistic baseline effectively addresses Hundt's (*ibid*.) precise research goal of determining usage preferences, a production-oriented question, while a *pmw* ratio does not.

Lange (2007) investigates usage of *itself* and *only* in ICE-India and the Kolhapur Corpus, specifically attempting to identify five elements of usage: preferred position and syntactic environment of *itself* and *only*; innovative uses and their usage restrictions; agreement patterns for *itself*; overlap in meaning and function between *itself* and *only*; and standard-ness of innovative usage for each word. Of particular relevance here is her report on how many times traditional and innovative senses of *itself* occur in each corpus, alongside a report of those frequencies *per 1,000 words* (*ibid.*: 100). She also comments that her analysis *per 1,000 words* was especially difficult because the corpus analysis software Wordsmith reported different word counts for the corpus from the word counts reported by the ICE Project (*ibid.*: 99). This problem, however, only arises if the study depends on knowing the total number of words in the corpus; the problem would have been prevented if Lange had avoided a *pmw* baseline and only used a probabilistic baseline such as the total number of instances of *itself* and *only*, or such as semantic alternates for each sense of *itself* and *only*. In fact, she goes on to present the former type of probabilistic baseline, in the form of a semasiological report on each word (*ibid.*: 101, 107). All of these numbers and ratios represent various bits of useful information, particularly in relation to exposure rates of various kinds, but the research question here is never related explicitly to the nature of these various relative measurements and ratios.

Lee and Ziegeler (2006) investigate usage of *get* in 'causative constructions' in ICE-GB, ICE-SIN, and ICE-New Zealand (ICE-NZ), asking specifically whether causative *get* is more commonly used in Singapore and New Zealand than in Great Britain. Elements of their methodological outlook are discussed, and their data is then presented in the following way: total instances of causative *get*; instances of causative *get per 100,000 words*; and instances of causative *get* measured *per get causative in ICE-GB* (*ibid.*: 129). In that third type of measurement, they use the number of instances of *give* and *take* with various senses in British English as a sort of quasi-baseline for measuring frequencies of those senses in Indian and Sri Lankan English. In this

case, an exposure rate in one variety is presented as a proportion of an exposure rate in another variety. This is an innovative tool that succinctly conveys an intuitively interpretable difference between the varieties. Without a probabilistic baseline, however, it is impossible to interpret these numbers as significant or not significant; the situation is comparable to knowing that a coin landed heads-up twice as often in Great Britain as in New Zealand, but not knowing how many times the coin was flipped in either location. There is a tremendous amount of information here, which might address several different research questions, none of which are stated. Nonetheless, Lee and Ziegeler (2006) is effectively onomasiological in a way: they do compare causative *get* to other means of expressing the same concept, which constitutes a probabilistic baseline. That onomasiological approach is reasonable, and a future researcher could derive from that data as presented a strong statistical analysis of the frequencies in terms of selection processes and production preferences. Lee and Ziegeler (2006) communicate their findings thoroughly and clearly enough for it to be a simple task to re-analyse their figures in order to compare exposure rates to production preferences.

In this section, I have argued that a *per word* or *pmw* rate measurement can be a rough tool for conveying a particular kind of exposure rate, but that probabilistic semasiological and onomasiological baselines facilitate more precise analyses that can tested statistically: the semasiological baseline relates to a particular kind of exposure rate (a receptive process), and the onomasiological baseline to a selection process (a productive process). Equally importantly, I have shown that a precise understanding of the nature of *relative frequencies* is essential to corpus linguistics, and greater attention to the various ways of measuring frequencies would help to improve an array of studies in the field.

2.3. Semasiology and onomasiology

In order to further explain the two key approaches to corpus frequencies discussed in 2.2, it is necessary to fully explicate the terms *semasiology* and

onomasiology. The semantic concepts of semasiology and onomasiology emerged among philological linguists in the nineteenth and early twentieth centuries (Geeraerts 2010: 23). *Semasiology* is traditionally an approach that begins with a word and examines the meaning or meanings of that word; the term here is applied to corpus studies that examine the different senses of polysemic lexical items. The complementary concept of *onomasiology* begins with a concept and examines the various ways of expressing that concept; the term here is used for corpus studies that examine various linguistic means of expressing a single concept. For example, a semasiological study might examine the many meanings of the polysemic English verb *get*. In quantitative corpus research, that might include the corpus frequencies of *get* representing one sense or another. An onomasiological study, in contrast, might identify the concrete meaning 'acquire' and identify means of expressing 'acquire', including not only *get*, but also *acquire*, *obtain*, *procure*, *come by*, *get one's hands on*,[9] and others. As a fairly rough description, *semasiology* might be represented by a dictionary, *onomasiology* y a thesaurus. Both approaches allow for a probabilistic lexical baseline in corpus linguistics, but it is necessary to bear in mind that a semasiological study may not necessarily employ a semasiological baseline, and an onomasiological study may not necessarily employ an onomasiological baseline. Either a semasiological or an onomasiological study, for example, may measure the various words that express a given sense and report examples *pmw*, rather than as a proportion of all instances of the given sense.

The term *onomasiology* has been common in continental lexicological research since the early 20th century (*cf.* Zauner 1903, quoted in Geeraerts 2010: 23), and its principal proponent in English has been Dirk Geeraerts, with

---

[9] An onomasiological approach may be strictly interpreted to include only single-word lexical alternates for expressing a particular meaning, in which case *come by* and *get one's hands on* would be excluded as onomasiological alternates for *acquire*. In that case, *come by* and *get one's hands on* might be seen as pragmatic alternates rather than onomasiological alternates of *acquire*. I discuss this subject at greater length in section 5.3.

the Quantitative Lexicology and Variational Linguistics research group at the University of Leuven (*cf*. Geeraerts *et al*. 1994; Geeraerts 1997, 2010). Early onomasiological research tended to focus on various terms for concrete objects, and an implicit onomasiological approach is evident in dialect atlases that map alternate terms for common objects such as *buckets* and *pails* (*cf*. Hempl 1902; Kurath *et al*. 1939, 1949). Structuralist semanticists such as Lyons (1963) emphasized an onomasiological approach as a method with linguistic integrity, and various approaches to onomasiology have been recognized throughout the twentieth century (*cf*. Geeraerts 2010: 49-50, Geeraerts 2006, Geeraerts 2002). Cognitive psychology and Cognitive Linguistics have largely emphasized semasiology, particularly vis-à-vis cognitive processes underlying polysemy (*cf*. Lakoff and Brugman 1988). In contemporary research on lexical semantics in the Cognitive Linguistics tradition, polysemy, and therefore semasiology, have been the focus of considerable attention (*cf*. Rosch 1973, 1975a, 1975b; Geeraerts 1997; Taylor 2003). An onomasiological approach has not been strongly supported by a large number of contemporary researchers. The onomasiological emphasis in the present work arises largely from statistical (including logical and mathematical) concerns, but also from an attention to cognitive selection processes between alternates, which can be seen as probabilistic (see 2.2; *cf*. Hanks 2013). Specifically, the onomasiological approach employed in the present study investigates lexical semantics in corpora by identifying a concept to be expressed and then comparing the frequencies of each of the words that might express it. That is, given that a language user needs to express a particular concept, what is the probability that he or she selects each of a set of lexical items to express that concept? Examples of onomasiological studies in World Englishes are relatively rare (*cf*. Balasubramanian 2009; Haase 1994; Schneider 1994). Geeraerts (1997) conducted an important study of this kind on Dutch lexical semantics, and has called for more onomasiological work to be done, particularly related to regional variation within a given language.

Why should an onomasiological approach be so rare in corpus studies of lexical semantics? One reason is that onomasiology and 'alternation studies' were an important part of generative linguistics, and corpus linguistics has often been at odds with generative linguistics philosophically. Another reason is the emphasis on polysemy, and therefore semasiology, in early cognitive psychology (*cf.* Rosch 1973, 1975a, 1975b) and subsequently in Cognitive Linguistics (*cf.* Geeraerts 1997, 2010; Gilquin 2006, 2008; Taylor 2003). One final reason may be the tremendous cost in time and analysis required for thorough onomasiological studies. Onomasiological studies generally require extensive manual analysis of individual instances of linguistic elements in corpora. Lexical searches are still the easiest and most straightforward entry point for most corpora. Because most corpora are not tagged or parsed, including most of the ICE corpora (the exception being ICE-GB, which is both tagged and parsed), researchers often search corpora for instances of individual words, rather than for word classes, grammatical constructions, or semantic or pragmatic categories. For semantic researchers, this means that the simplest approach is to collect data on a single word, such as *make*, *take* or *give* (the words examined in the present study), rather than identifying alternates of given words and their various senses, and then identifying instances where alternation is in fact plausible in the corpus. The simplest semantic studies are semasiological, analysing individual words which can easily be found via simple lexical searches of a corpus. Data collection for a semasiological study is generally more straightforward than data collection for an onomasiological study.

The production-oriented onomasiological baseline can be complemented by a precise, reception-oriented semasiological baseline, but semasiological baselines give rise to the unique problem, discussed in 2.2, that every instance of a polysemic word is not strictly a potential instance of any sense of that word. Although a semasiological baseline produces a valid probability that can range from 0% to 100%, it is clear that in actual use, *make* with a given sense is not generally a potential use of *make* with any other sense.

To take this point further, semasiological studies can be seen as pointing to contexts and situations of use at least as much as they point to the nature of the words in question. For example, a 2,000 word text excerpted from a fictional piece about a baker 'making' bread and sweets (such as ICE-HK W2F-005) will likely include far more uses of concrete *make* than a 2,000 word text about philosophy of mind (and, indeed, ICE-HK W2F-005 does). The high frequency of concrete *make* in the text, measured against a semasiological baseline, is not properly seen as a linguistic norm – it is not in itself a fact about language use in a region like Hong Kong, but a fact about the topic of a particular text. For semantic purposes, the crucial question is not simply how often a language user selects concrete *make* in any given text, but how frequently a language user selects it in relation to the number of opportunities to select concrete *make*. To study selection in relation to opportunity, we can consider, for example, instances of concrete *make* against instances of *bake*, *cook*, or *prepare*, which evinces a language user's selection process between those alternate verbs. This selection process, and the contexts which define the final selection, is a fact about linguistic norms across populations, rather than about text topic or about the nature of baking, for example. In the language of Wallis (*forthcoming* b: 11), the onomasiological perspective is most viable because it 'minimises invariant Type C terms'. Type C terms are terms that are not substitutable for the sense in question, and that should therefore ideally be eliminated from the baseline of a statistical model. Wallis (*forthcoming* b) effectively reiterates Geeraerts *et al.*'s (1994) argument that onomasiology reflects psycholinguistic processes, but Wallis (*forthcoming* b) frames the argument in terms of statistical analysis.

As an example of this issue in practice, using a specific case of lexical studies of World Englishes, Schmied (2004) has observed that the ICE-East Africa corpus contains far more lexis *pmw* related to superstition and folk religion than the other ICE corpora, and has proposed this difference in content as indicative of cultural norms in East Africa. The problem with Schmied's argument is that the corpus is not controlled for topic or content in any way –

the corpus is only controlled for speaker background and 'text type', which indicates modes and social contexts of production rather than topic or content specifically. To be sure, samples of courtroom proceedings in the ICE corpora are to some extent controlled for topic insofar as their topic and content will relate to legal fields. Nevertheless, one ICE corpus might, for example, contain an overwhelming number of drugs-related courtroom dialogues (as, in fact, ICE-HK does; see 7.2.17), which may affect various linguistic features in use in ways that will not be repeated in other corpora. Resampling personal conversations for ICE-East Africa might result in very different topics. Similarly, resampling other ICE corpora during a local folk religious holiday might very well result in increases quantities of lexis related to folk religion in those corpora. The linguist's work in identifying linguistic norms, rather than anomalous features of topic and content, is to tease out these effects in the experimental design. A probabilistic baseline, via an onomasiological approach as I forward in 2.2, goes some way towards isolating the appropriate variables by considering instances of a linguistic feature in terms of opportunities to use that linguistic feature.

In fact, if a semasiological baseline and a *pmw* baseline convey more about topic and context than about linguistic norms, an onomasiological baseline reveals even more about topic and context, in addition to linguistic norms. To continue with the example of CE-HK text W2F-005, that text includes 13 instances of *make* with a concrete Direct Object, far more than the average per text in the corpus. It is even more useful to know that the same text contains 5 instances of alternate verbs with the sense 'Produce (Concrete)', and that all 5 instances more specifically relate to food (e.g. *bake*, *prepare*). In addition to providing valid data on usage preferences between alternate verbs, and therefore linguistic norms, this second fact, representing an onomasiological approach, also provides more complete and precise data on the topic of the text ('producing food') and how frequently the particular act of producing food is actually described in the text. The act is actually described not just 13 times (with *make*) but 18 times (including alternates for *make*). It is

clear from an onomasiological approach not only that producing food is an important part of the text's topic and content, but that the author of the text prefers the verb *make* to semantic alternates.

Onomasiological studies are sometimes referred to as alternation studies. Dylan Glynn (in Arppe *et al*. 2010) presents 'binary alternation studies' as a starting point toward broader studies of a full range of onomasiological alternation. Glynn (*ibid*.) proposes that studying simple alternations between two similar grammatical constructions in corpora, for example, is a useful first step in corpus linguistics. In fact, binary alternation studies are extremely common outside of lexical semantic research (*cf.* among many others: Aarts *et al*. 2013 on *shall* vs. *will*; Xu 2013 on the dative alternation; Hilpert 2008 on ADJ-*er* vs. *more* ADJ; Collins 2005 comparing sets of epistemic or deontic modals), and they establish probabilistic baselines by representing each of the two features in question (*e.g. shall* or *will*) as a proportion of the total (*e.g.* all aggregated instances of *shall* and *will*). These are naturally production-oriented studies, though they are often not explicit about the production/reception distinction. Glynn (in Arppe *et al*. 2010) argues that such binary alternation studies should be viewed only as a starting point, and that the simplicity of binary studies reflects the 'theoretical heritage' of generative syntax (*ibid*.: 12). Moving further, Glynn (*ibid*.: 12) asserts that sophisticated corpus studies must examine more than just two alternates and also investigate a wide array of variables that might influence a language user's selection of a particular alternate. Wallis (2014b) agrees with Glynn (in Arppe *et al*. 2010) that binary alternation studies are the rightful building blocks of corpus research, but unlike Glynn (*ibid.*), Wallis (2014b) emphasizes that such binary studies can in fact be extremely powerful and sophisticated in themselves. Gilquin (in Arppe *et al*. 2010: 14) further claims that 'even if a given alternation does not exhaust the logical space of what is possible… the study of alternations can nonetheless give us important clues about general organizational principles of language'. The present study investigates a wide array of alternates, but not, for example, additional grammatical variables or social variables such as speaker gender, that

might influence the selection of alternates; in agreement with Wallis (2014b) and Gilquin (in Arppe *et al*. 2010), the isolation of the semantic variables here is seen as a powerful tool.

Heylen *et al*. (2008) assert that the complex array of variables in spontaneously produced language in corpora, particularly in speech, prevents researchers from isolating and controlling individual linguistic variables as they would in controlled experimental studies (*ibid*.: 101). Heylen *et al*. (*ibid*.: 112) note that social variables such as region and social stratum influence language in corpora, and that such factors must be considered when deriving conclusions from corpus data. Along similar lines, Smith and Leech (2013: 75) have noted 'the difficulty in determining "true alternants"', and they argue in favour of the *pmw* baseline for its ease of use and its consistency with most corpus design, in that corpora are generally sampled to contain a set number of words. Smith and Leech (*ibid*.) ask where true alternation can actually occur, and imply that it is rare. They are undoubtedly correct. However, onomasiological research like the present study does not assume true alternation like that discussed by Smith and Leech, but instead serves as a testing ground for alternation. When we identify potential alternates to test, by observing that words appear in similar contexts or do similar things, we can then ask what variables contribute to the selection of one alternate over another (as Glynn, in Arppe *et al*. 2010, and Wallis 2014b explain as well; see above), and whether the selection of one option over another is consistent enough to qualify as a linguistic norm. That is the approach in the present study.

As noted above, an onomasiological approach is implicit in early dialect atlases (*cf*. Hempl 1902; Kurath *et al*. 1939). In contemporary sociolinguistic theory, ecological models such as Mufwene's (2001) model also depend implicitly on onomasiology (*cf*. Schneider 2007). Mufwene (2001: 6) proposes an ecological theory of language environments in which a full range of linguistic features (*e.g*. words, morphemes, grammatical constructions, meaning correspondences) constitute an ecological 'feature pool'. Language users in a region or environment select elements of the feature pool at various times and

in various situations, for various purposes. This production-oriented description of a selection process depends on the idea that one feature can 'replace' another. That is, linguistic alternation occurs (*ibid.*: 5). Mufwene's ecological model is a fundamental element in Schneider's (2007; see 4.1) model of Postcolonial Englishes as well.

Exemplifying an onomasiological approach in research in lexical semantics in World Englishes, Balasubramanian (2009) 'determines the proportions of *also* and *too* with respect to each other' in ICE-INDIA and ICE-GB, and includes variation between genres and in various syntactic locations. The study posits the premise that, given a choice between *also* and *too* in various positions within a sentence, a language user in India or Great Britain can choose one or the other. Hypothetically, users might always select *also*, or they might never select *also*, allowing for probabilities from 0% to 100%. The study then asks if there are differences between the two corpora in the probability of selecting one or the other. Findings show that ICE-India and ICE-GB differ significantly from each other vis-à-vis where in a sentence language users tend to select *also* or *too*; and that ICE-India shows a great deal of internal variation between genres in this respect. The study presents a strong methodological approach, and also explicitly raises and explores significant questions as to the actual potential for *also* and *too* to alternate (echoing Smith and Leech 2013). Any onomasiological study raises unique issues in identifying actual potential for alternation. Similar issues are explored throughout the present study, particularly in chapters 6, 7, and 8.

Haase (1994) compared usage of English motion verbs and their multi-word alternates in ICE-East Africa. That study compared *ascend*, *go up*, and *ascend up*; *descend*, *go down*, and *descend down*; *exit*, *go out of*, and *exit out of*; *enter*, *go into*, and *enter into*; and *pass*, *go along/through*, and *pass along/through*. The study asks: given the choice between *ascend*, *go up*, and *ascend up*, etc., which option is a speaker of East African English most likely to select? The study found that language users in the corpus display a preference for expressing *path* (e.g. *up*, *down*) even if that leads to redundancy (e.g. *enter*

into). Again, the approach to usage preferences is clearly supported by the onomasiological method employed, and the nature of the frequency measurements accords with the research question as stated.

Schneider (1994) investigates the ICE corpora for Great Britain, Singapore, Philippines, India, Kenya and Tanzania, and reports frequencies of common particle verbs in comparison to common synonyms of those particle verbs. Schneider (*ibid*.: 236) describes this as a 'paradigmatic approach', and it can also be seen as onomasiological. Schneider acknowledges the limitations in his work, as he does not rigorously distinguish all polysemous senses of each item in question, given the high frequency of the items. This issue evokes Smith and Leech's (2013) concern regarding the difficulty of identifying true alternates. Schneider (1994) notes that reporting such paradigmatic ratios (*i.e.* instances of particle verbs in relation to instances of semantic alternates) is more precise than reporting frequencies *pmw*. The research question is: given that a language user wants to express a given concept, represented by a specific semantic field, what is the probability that the language user chooses a particular particle verb? His findings in this regard are intriguing and worthy of further study, showing for example a significantly higher preference for *assist* over *help out* in East African data, in comparison to a significantly high preference for *help out* over *assist* in Singapore data, among many other findings. These findings would seem to point to *regional norms* for English usage (see 4.1).

While semasiological studies, particularly in cognitive psychology, have made invaluable contributions to our understanding of language (*cf*. Rosch 1973; Taylor 2003), onomasiological studies are indispensable in corpus linguistics in order to accurately measure and model semantic trends, and hence linguistic norms. Geeraerts (1997) exemplifies an approach that effectively combines onomasiology and semasiology. That work is a thorough onomasiological study of garment terms in Netherlands and Belgian Dutch, alongside a semasiological study of one term in particular. The study's first goal is to determine whether preference for a particular term, the neologism

leggings, borrowed from English, changes over time in relation to its alternates. The study also aims to ascertain whether, semasiologically, *leggings* develops more diverse meanings over time, analysing uses of these terms in women's magazines and catalogues. Relying on a photograph of a concrete object as a referent allows Geeraerts (*ibid.*) to categorize images according to concrete attributes, and then to check the various terms that are used to refer to similar or identical images. Likewise, semasiologically, he analyses the attributes of all images connected to the particular term *leggings*. Geeraerts (1997) concludes that *leggings* became more common in relation to its alternates in just a few years, and that preference for alternates varies between varieties of Dutch. He also concludes that, semasiologically, the range of referents for *leggings* increased dramatically in just a few years as well. Geeraerts's (1997) study is enriched by a consideration of both semasiological and onomasiological findings.

The onomasiology question may also implicitly underlie some of the criticism directed at corpus linguistics as a field by researchers in other fields. One of the best-known criticisms comes from Chomsky:

> …some sentences won't occur because they are obvious, others because they are false, still others because they are impolite. The corpus, if natural, will be so wildly skewed that the description would be no more than a mere list. (Chomsky 1962: 159, quoted in McEnery and Wilson 2001: 10)

Chomsky asserts that such a list of frequencies is not interesting or useful. This is an important criticism, because some of the early, important research on corpora was in fact composed of reported lists of simple numbers for each lexical item in a given corpus or corpora (*cf.* Hofland and Johansson 1982; Francis and Kučera 1982), and those studies continue to be cited. As discussed above, however, simple counts of features, *sans* baseline, do not convey much information in terms of linguistic norms. Chomsky apparently, famously, went

on to present a specific, oft-cited argument about what corpora could tell linguists about linguistic norms (*cf.* McEnery and Wilson 2001: 10; though the source of Chomsky's original statement is apparently lost), arguing that *I live in New York* is more frequent and therefore more likely to occur than *I live in Dayton, Ohio* because more people live in New York than in Dayton. Chomsky is right about this example, which relates to exposure rates, and he presumably raised it as an illustrative criticism of the field because of the number of corpus studies that compare unrelated linguistic features *pmw*. Such studies can be useful in measuring exposure rates: in a psycholinguistic study, for example, exposure rates to *New York* and *Dayton* as evidenced by corpora might be seen to correlate with recall times for those two terms. That is a very precise research question relevant to a very precise study. Is there a reasonable onomasiological research question, grounded in probabilistic frequencies and a psycholinguistic production process, to be asked about *I live in New York* and its frequency in a corpus? We might ask an onomasiological question such as the following: given that a speaker or writer in New York City (representing New York City English) wishes to express the idea 'I live in New York', what is the probability that the speaker or writer says or writes *New York* and what is the probability that the speaker or writer says or writes *Nueva York*? This question, while obviously very limited in scope, might be meaningful in relation to significant sociolinguistic questions in New York City. The final conclusion to be drawn here is that corpus linguistics without a strong understanding of the nature of frequencies and baselines, and without precise research questions, may actually be as useless as Chomsky believed it to be.

3. LEXICAL SEMANTICS

3.1. What constitutes lexical semantics?

Leech (1974) offers an interesting meta-discourse on the field of semantics, beginning his discussion with some reflections on *words* and *meaning*. He describes his preliminary discussion as 'prescientific', not constitutive of a scientific theory, and not testable due to the imprecise, fuzzy nature of the reflections and understandings presented (*ibid*.: 69). Leech (*ibid*.: 69) goes on to explain that 'it is useful to have rough-and-ready ways of charting an imperfectly explored terrain – which is what semantics is. We need tentative ways of looking at and rendering orderly a range of phenomena so vast and perplexing'. That is, we need a prescientific approach, at least at first, and we need to make that prescientific approach explicit. Leech (*ibid*.: 70), however, then argues that in studying semantics, the necessary ultimate goal must be:

> …to devise strictly and explicitly formulated theories, so that anyone can see what is and what is not being claimed… to render such theories accountable to objective data… to account for *all* the data at one's disposal, and to account for it in the *simplest* way possible.

In sum, Leech acknowledges the necessarily prescientific nature of many discussions in semantics, but ultimately advocates a move toward scientific and intellectual rigor in semantic research. I concur with Leech completely.

Despite the need for rigour in semantic analysis, the qualitative and reflective nature of much lexical semantic research – its prescientific nature – cannot be overlooked. It is the qualitative and reflective nature of the object of study that pushes some grammarians to avoid semantic analysis as much as possible (*cf*. Aarts 2011; Huddleston and Pullum 2002). Avoiding an entire sphere of inquiry within linguistics, however, hardly seems a reasonable answer (and, in any case, neither Aarts, 2011, nor Huddleston and Pullum, 2002, actually avoids semantics completely). I would argue that the qualitative and

reflective nature of the object of study ultimately demands not that it be avoided or relegated to secondary status, but the opposite: semantics demands an explicitly defined approach that employs the utmost rigour and systematicity. If we are to work in a scientific way, and that is certainly the goal, then our experiments must be empirically verified and reproducible, and our accounts must be as simple as possible. Just as Hundt (2007: 4-5; see 2.1) acknowledges the impossibility of eliminating reflection from corpus linguistics, we must acknowledge that reflection and subjective judgements about meaning will inevitably play a role in semantic analysis as well. As with corpus linguistics, it is the linguist's work to design experiments with such intellectual and scientific rigour and systematicity that the role of reflection is explicitly acknowledged and defined, and subjectivity therefore minimized.

Lexical semantics is most directly and simply understood as 'word meaning'. I would like to begin by identifying a few aspects of word meaning that are not the subject of inquiry here. The present study does not ask how meaning can exist. Moreover, as the present research questions convey, this study does not address the full range of possible or imaginable meanings of a word, however improbable or rare those meanings may be. Reflective semantic work on the range of imaginable meanings of a word has occupied no small part of linguistic study (*cf.* Wierzbicka 1982, Brugman and Lakoff 1988, as two important and representative studies among many). However, the possible range of meanings of a word may indeed be infinite: Cruse (2004: 106) has commented on the infinite possible meanings a word might express and Derrida (1977) has pointed out, in a parallel argument related to speech acts, the infinite potential meanings of any utterance. Hanks (2013: 4-5) has argued that an emphasis on the range of imaginable meanings for words has skewed the field of lexical semantics away from the study of linguistic norms:

> The tradition of speculative approaches to the study of language through the invention of data is particularly regrettable because the focus on boundary cases supported by invented examples left linguistics

drowning in a welter of imagined possibilities, while the central and
typical norms of languages as systems were neglected...

Very much in line with Hanks (*ibid.*), the present study addresses meaning in
use –the meanings of words as they are actually used in natural language, and
the choices that speakers make between words in context – and, specifically,
subtle fluctuations in semantic usage. These semantics norms are quite different
from the range of imaginable meanings of a word, just as they are different
from 'lexical exotica' (such as *stranger* being used to mean 'guest' in Nigerian
English; see 4.3). As such, the present study describes the meaning, as well as
the semantic alternates, of English words in use.

I am also interested in lexical semantics as the intellectual inquiry into
the dynamic relationship, and often blurred lines, between constant and
contingent meaning. From the 19th century philological tradition to present day
research in lexical semantics, the distinction between constant and contingent
meaning has been a foundational question. The distinction is emphasized in
contemporary lexical semantics texts by Leech (1974: 71-81), Palmer (1981:
44-66), Saeed (2003 :59-62), and Cruse (2004: 101-22). In essence, the
distinction between constant and contingent meaning is the distinction between
that meaning which relates to a word itself (*i.e. lexical semantics*), and that
meaning which does not relate to the word itself. The second category of
meaning might include meaning related to the surrounding words, the medium
of expression, intonation, gestures, setting, context, or anything from the
broader sphere of discourse and pragmatics. The dynamic between constant and
contingent meaning is the subject of a great deal of contemporary research,
across multiple sub-disciplines of linguistics (*cf.* Traugott and Dasher 2002;
Evans and Wilkins 2000; Carston 2008; Falkum 2011). The terms of the
discourse are varied: it can also be seen as the dynamic between semantics and
pragmatics (Enfield 2002).

The distinction between constant and contingent meaning is not always
clear. Indeed, the line between the two seems to be fuzzy. That is not to say that

the terms or the categories they represent are entirely fuzzy. One important question for lexical semantics is the nature of the boundary between these two categories, and how we can explore and understand the categories and the boundary between them. The categories might be best described as one continuous cline, or they might represent two largely distinct categories with just a bit of overlap. We might investigate the categories reflectively or from corpora, and we might investigate all of a word's possible constant and contingent meanings, or just those meanings represented in natural use in corpora. In investigating this unclear distinction, I use the terms *relatively* constant and *relatively* contingent meaning, and I employ traditional polysemy tests to explore relatively constant and relatively contingent meaning (see 3.3). Because the present research is usage-based, relying on corpora, I examine consistent meanings and unusual meanings in use, and instances in natural language where semantic categories seem to be blurred. Based on a linguist's own reflective processes, any word might indeed be used to convey an infinite array of meanings, but in actual use, there are meanings that are more constant – *i.e.* more commonly related to the word in question than to any contextual factors – than others. There are also unique extensions and exploitations (Enfield 2002; Hanks 2013), in which words are used creatively in unique ways for expressing meanings they would not usually express, and such extensions can be distinguished quantitatively in corpora from norms of use.

Cruse (2004: 106) notes that 'it is almost certainly the case that all words are potentially usable with meanings other than their default readings (*i.e.* the meanings which would come to mind in the absence of any contextual information)'. The issue of which meanings are 'default' is not clarified significantly by the reference to meanings that 'come to mind'. Indeed, the question of what comes to mind is quite separate from polysemy as such (I address the question of meanings that 'come to mind' in 3.4 and in 9.3). Cruse (*ibid.*: 107) goes on to note that from the variety of potential meanings, most would not belong in a dictionary, and most are not 'permanently stored in the mental lexicon'. It is beyond the aim of the present work to answer questions of

lexicographical integrity and the psycholinguistic nature of the mental lexicon. Nonetheless, the core point is crucial: a word can potentially convey an infinite array of meanings. That is, semantic possibility is infinite; it is only by studying words in use that we can begin to describe semantic norms. Because the present study aims to investigate semantic norms across World Englishes, the question of semantics in use must be central, rather than the question of the array of possible meanings associated with a word.

Meaning is associated with a word by convention, and convention arises from use. For example, the meaning associated with *dog* is not inherent to its phonetic form. Nonetheless, by convention, linguists recognize that some meaning tends to be associated with *dog* that does not tend to be associated with other words. We can illustrate this by replacing an individual word in an example sentence with a different word and observing that the resulting sentence tends to carry a new meaning:

1. There is a cat on the roof.
2. There is a dog on the roof.

These two examples show quite clearly that there is a relatively constant meaning related to *cat* which differs from the relatively constant meaning related to *dog*. A creative linguist might, exploring the full range of imaginable meanings for a lexical item, might point out that, in fact, *cat* in the first sentence could refer to a dog in a cat costume, or a dog that looks like a cat – this is not an entirely unreasonable objection if the aim is to explore the range of imaginable meanings for *dog*. Moreover, Austin (1962: 22) might point out that the proposition could be 'parasitic', consisting of a joke, a lie, a quotation, or perhaps a recitation of a line in a play, such that in the real context of the

utterance, there is actually neither a cat nor a roof at all.[10] These objections, too, are not entirely unreasonable, depending on the research aims. The point here, in a corpus study, is that the researcher can identify these problems when they arise in natural use (which they very rarely do), and can explore and discuss them. It then becomes reasonable to ask, in various instances of use, what meaning tends to be associated with an individual word (its constant meaning, or perhaps more precisely, its *relatively* constant meaning), and what meaning is associated with the surrounding context (the contingent meaning), including each of the surrounding words, their grammatical relationships, their discourse functions, and further information such as extended discursive context and physical environment. That is, in the present corpus study, this issue is addressed by asking not how much meaning a word might be able to convey, or what cognitive processes might allow a word's meaning to be extended or re-interpreted across a potentially infinite array of imaginable scenarios, but by instead asking about actual use and usage preferences, and about evident alternations in the corpus.

In the following section, I present the methodological preliminaries for identifying semantic alternates, the starting point of the onomasiological approach. Then, in 3.3, I show that in order to approach polysemic lexical items onomasiologically, the distinct senses of each lexical item must be determined as well, and I present systematic ways of determining those distinct senses, including application of polysemy tests as heuristics and exploratory tools in lexical semantics. Finally, in 3.4, I relate corpus studies to psycholinguistics once again to explore the relationship between word meanings and word frequencies, particularly vis-à-vis concrete senses and light senses of polysemic verbs from the perspective of Cognitive Linguistic theoretical frameworks.

---

[10] Derrida (1972: 16) has focused on deceit and quotation as essential to language rather than peripheral or parasitic, and to Davidson (1982), the nature of quotation is not peripheral or parasitic, but essential to understanding meaning.

3.2. Corpus onomasiology: Identifying semantic alternates

An onomasiological approach requires an effective identification of semantic alternates. In this section, I first discuss two traditions of identifying semantic alternates: the lexicographical and the computational. Importantly, those two traditions are converging. I summarize the traditional lexicographical approach, and I then discuss the computational collocational approach, and the convergence between the two. Finally, I present an alternative to those approaches which minimizes subjectivity and also offers greater precision.

The traditional lexicographical approach applies to creating thesauri. Thesauri generally derive directly from dictionaries: a list of vocabulary items is drawn from a dictionary, and those vocabulary items are categorized intuitively according to taxonomies (*cf.* Kay and Chase 1990). The recently completed *Historical Thesaurus of the Oxford English Dictionary* (HTOED), for example, contains three super-categories: *world*, *mind*, and *society*. In creating the HTOED, each word from the OED was first categorized into one of the three super-categories; each super-category contains several immediate sub-categories, which branch into lower and lower levels of a taxonomy (Kay and Chase 1990). Words in the same lower levels of the taxonomy can be expected to be – roughly and with exceptions and nuances – semantic alternates or synonyms. Like much lexicographical work, this process can be considered an art as much as a science, dependent upon the tastes of the lexicographer, and subject to cultural and social influence from the lexicographer as well (*cf.* Kay and Chase 1990). This system is practical, and the product is very useful, but it does not allow for the sort of scientific rigour that the present study aims for in terms of verifiability and reproducibility, due to the subjectivity and social and cultural bias inherent in the method. Many lexicographers are not deliberately addressing norms beyond those of British or American English, and to assume that the intuitied norms of British or American English were the same for Singapore and Hong Kong English would be to undermine the very purpose of the present study.

The computational, automated approach can identify semantic alternates in data sets that are too large to be analysed manually. The approach is generally based on identifying collocations, the words that tend to occur near a given word, based on Firth's (1957) principle that a word's meaning can be understood by looking at the other words around it. For example, the automatic 'Thesaurus' feature in the Sketch Engine corpus analysis software automatically identifies words that share most collocates with a given word in a given corpus (Kilgariff and Kosem 2012). If manual semantic analysis is not an option, collocation analysis is generally a practical alternative. However, it must be borne in mind that collocation analysis provides a very general picture of a word's field of meaning. By way of illustration, *farmer* collocates strongly with *part-time* in the British National Corpus, but it is not at all clear whether that is because farming is an act that tends to be performed part-time or, perhaps, because farming is decidedly *not* an act that tends to be performed part-time, not to mention an extremely wide array of other possibilities, such as farmers employing part-time help. Collocational analysis is therefore useful for research questions inquiring into broad pictures of words that co-occur, but not for precise questions regarding specific nuances of meaning distinctions in polysemic lexical items, nor for investigating usage preferences between semantic alternates. Moreover, collocational analysis also shows similar shortcomings, though in a different manner, to the *pmw* baseline (see 2.2): findings reflect text topic and content, and in corpora that are not controlled for topic and content, findings will thus reflect random, uncontrolled variables.[11]

---

[11] In fact, collocation analysis in a corpus that is not controlled for topic or content can actually help to identify the nature of the content. For example, collocational analysis can reveal what topics tend to co-occur within a given text type. If we analyse the words that co-occur with *terrorism* in a newspaper corpus, we are asking about the details of the content of the discussion of the broader topic of *terrorism* within that newspaper, and also in the group of language users represented by that newspaper. Such data can be seen as reflecting real-world context and topics first, and linguistic norms only insofar as they intersect with real-world context and topics. This is a valuable use of collocation analysis, but it is different from the aims of the present study.

For example, the most frequent collocate for the high-frequency, highly polysemic verb *thought* in the Corpus of Contemporary American English is, surprisingly, *Christina*. This is due to the overwhelming influence of a single text about the thoughts of a woman named Christina.

The lexicographic tradition and the computational tradition are converging. For example, the automated semantic tagger WMatrix, developed at Lancaster University, combines a syntactic tagger with conceptual categories from the *Longman Lexicon of Contemporary English* (McArthur 1981; *cf.* http://ucrel.lancs.ac.uk/usas/). WMatrix identifies a word form and its word class and then places it in one of the *Longman Lexicon*'s more traditional lexicographic taxonomies. Likewise, contemporary thesaurus creation often refers to collocational data as well as older taxonomic systems (*cf.* Kilgariff and Kosem 2012). Collocation analysis offers a verifiable and reproducible complement to the often subjective (and difficult-to-verify) tradition of lexicographic taxonomies.

For the present study, neither the cultural bias of the traditional lexicographical approach nor the broad strokes of collocation analysis are adequate. I am not interested in every word that co-occurs with *make*, *take*, and *give*. My research question is not on the nature of the topic or content that tends to surround uses of *make*. Moreover, collocation analysis lists all co-occurrences, including both modifiers and complements, and I am not interested in the modifiers that co-occur with the words in question: I am not asking, for example, 'If *make* is modified by *x*, what other words are also modified by *x*?' That is, I am not inquiring what other acts tend to be described as being performed *well*, *poorly*, *quickly*, *slowly* and so on. Similarly, I am not asking what other verbs or other words tend to co-occur with the word in question: that is, I am not asking 'Given that a language user speaks or writes the word *make*, what is the probability that he or she also speaks or writes the word *x* within close proximity?' I can justifiably leave behind the broad strokes of collocation analysis because I manually analyse each instance of each word, attaining a high degree of consistency in each analysis. Instead of automatically

identifying all co-occurring words, I first identify grammatical complements of the word in question. For example, I catalogue Direct Objects that occur with *make*. I do not perform that task because I am interested in what actions, other than *make*, can be done to a *cake*, which might include *eat*, *ruin*, *deliver*, or *catapult*. Rather, I am working from Geeraerts's (1994) description of onomasiology: given that a language user is expressing a concept *x*, what is the probability that the language user employs word *a* or *b*? In this case, what semantic alternates can replace *make* in expressing the act 'making a cake'? That is, I am interested in investigating the corpora to answer the question how else we can express the meaning of *make* in *make a cake* (for example, *cook*, *bake*, and so on, according to the corpus), and what preferences do language users have in each region?

Alternates, therefore, are words that occur with the same grammatical complements, and that seem to be substitutable. That statement - *seem to be substitutable* – indicates the role of subjective reflection here. Subjective reflection is minimized insofar as each decision regarding the substitutability of two words is not based solely on subjective reflection but also on corpus data. Ultimately, it must be borne in mind that the present research is usage-based, built on corpora as evidence of language in use, so these questions are not purely abstract ones regarding words that could potentially alternate with the given ones. Instead, these questions relate to which words actually do alternate in use with the given ones. This distinction is equally important in identifying polysemy, which is discussed in the next section.

Each of the words in question, *make*, *take*, and *give*, has multiple meanings as well; each word is polysemic. It is, therefore, also necessary to identify the distinct senses for each word. The present approach is methodologically cyclical: identifying alternates aids in identifying discrete senses, and identifying discrete senses aids in identifying alternates. This cycle is in line with the research-theory cycle presented in 4.2. A working hypothesis on a word's polysemy can be corroborated by looking at a word's onomasiological alternates in use; those findings on alternates in use serve to

revise the hypothesis on polysemy. This cycle forms the foundation of deductive scientific inquiry. This research-theory cycle is not a kind of logical circularity; because it is an evidence-based cycle which includes empirical observation, the cycle does not represent circular reasoning.

Having discussed the process of identifying onomasiological alternates in corpora, I now proceed, in the following section, to discuss the process of identifying polysemy in corpora, which constitutes a methodological first step when dealing with the onomasiology of polysemic words.

3.3. Corpus polysemy: Identifying discrete meanings

Polysemy is an extremely complex and highly debated topic. It is perhaps most effective to begin with a traditional and relatively general discussion of polysemy before moving on to the intricacies. Traditionally, *polysemy* describes words with discrete meanings that are nonetheless related in some way (*cf.* Kempson 1977: 40; Cruse 1986, 2004; Geeraerts 2006 [1993]). For example (*cf.* Mehl 2013), *crane* is polysemic, with the two senses 'a kind of long-legged bird' and 'a tool for hoisting objects', the latter meaning having developed from the former metaphorically, due to the physical similarity of the two referents. In the present study, recognition of polysemy is a necessary part of an onomasiological approach: it is apparent that a single word may have multiple senses, and therefore multiple semantic alternates. In this section, I discuss the nature of polysemy and means for identifying polysemy. I continue to argue that scientific integrity requires semantic analysis to be systematic, and minimally reflective, even if semantic analysis is by nature not mechanistic. Below, as in 3.2, I begin with a discussion of polysemy as approached lexicographically and computationally, and then move on to polysemy tests.

Lexicographically, identifying and distinguishing discrete lexical meanings for a single word is relatively subjective. As Hanks (2009: 224) has stated:

A question that engages much lexicographical energy is, 'How many senses are there of this or that word?' To which the riposte is, 'How long is a piece of string?' That is, there is no reliable way of deciding how many senses a word has: deciding this is, in each case, a matter of lexicographical art and judgment.

Atkins and Rundell (2008: 264) similarly note the fuzziness of the lexicographical approach, in that 'there is little agreement about what word senses are (or even whether they exist). Lexicographers are therefore in the position of having to describe something whose nature is not at all clear.' Van der Meer (2006: 604) concurs: 'After centuries of practical lexicography, there is still hardly any consensus on how to divide the semantic space of a lexical item.' Identification of senses in dictionaries can be crafted based on tradition or reflective processes and can be influenced by social and cultural factors as well. Methods tend to be implicit rather than explicit. Of course, dictionaries are tools with particular aims – they are not strict semantic studies – so these facts about lexicographical methods are justifiable and not unexpected. That said, lexicographers do employ the types of polysemy tests I discuss below (Atkins and Rundell 2008: 139), even if in their employment of those tests, scientific rigour is less important than the practicality of the final product. Lexicographical approaches aside, computational approaches have not been terribly successful at automatic identification of polysemy (via computers and corpora), though such approaches are improving. Kilgariff and Kosem (2012: 29) state that 'computationalists are rapidly closing in on the word sense disambiguation task, even if aspects of it (such as mapping the evanescent senses of highly polysemous verbs) remain challenging'. That statement bears particularly heavily on the present study, which investigates highly polysemic verbs.

In traditional philosophy and linguistics, clearly defined criteria for polysemy take the form of polysemy tests. Kempson (1977: 128) asserts that systematic and rigorous polysemy tests are a necessity because such tests are

the only means of ensuring reproducibility and verifiability: given a disagreement between linguists on the polysemy of a word, a polysemy test is the systematic means of working through the disagreement. To be sure, polysemy tests engage reflection, but such tests systematize the approach to reduce subjectivity and allow for greater reproducibility; without them, we are left with haphazard intuitions alone. Although Kempson's argument is decades old, polysemy tests remain the standard for analysing and identifying polysemy. Cruse's (2004) recent textbook on semantics upholds such tests as fundamental, reliable criteria for polysemy.[12] The polysemy tests presented here are not mechanistic processes for outputting a correct answer, nor are they algorithms to process data and report truths. In fact, semantic tests are, in that way, very similar to statistical tests: statistical tests are not algorithms to process raw data and spit out the truth, but must be designed and built up from rigorous, systematic experimental designs unique to each analytical situation. As I have shown in 2.1, a sound statistical analysis depends on a specific, clearly defined research question, on a baseline appropriate to the research question, and on appropriate data collection to match the research question: and even with those factors in place, the choice of an appropriate statistical approach can be debated. Similarly, polysemy tests depend upon clearly defined criteria; a standard against which to compare given examples in given circumstances; and a dataset appropriate to the research question. The polysemy tests described here are defined clearly, systematically, and rigourously enough to be reproducible. Combined with corpus data, these polysemy tests can be particularly powerful for testing polysemy in use (as opposed to polysemy in imagined scenarios, tested via reflective processes). Because results of the tests will sometimes be open to debate, the tests can also be seen as exploratory tools and heuristics for negotiating the dynamic relationships between the meanings

---

[12] Cruse (2004: 104-5) prefers to call these tests 'discreteness tests', and his defense of that preference is compelling. Nonetheless, I do not aim to wrestle with such terminological distinctions in the present work, and I accept the standard use of 'polysemy' in this case.

conveyed by a word in use, and for describing the dynamics and degrees of polysemy for a given word in a given instance of use in a corpus. *Degrees of polysemy* is a very useful concept here; in Cruse's (2004: 112) words, 'there are many degrees of distinctness which fall short of full sensehood, but which are nonetheless to be distinguished from contextual modulation'. That is, a word may be used to convey meaning that is not contingent, but also nonetheless not entirely constant.

The most firmly established linguistic tests for identifying polysemy are summed up in Cruse's (2004, 1986) textbooks, and they include: the 'independent truth conditions test' (Cruse 2004: 105); the identity test (*ibid*.: 104), and the antagonism test (ibid: 106). In addition, Quine's (1960) test is a useful variation on the others. Numerous linguistic studies over the years have employed various polysemy tests in various ways (*cf*. Tuggy 1993; Zwicky and Arnold 1975; Lakoff 1970), though most of those studies have depended upon the linguist's subjective reflection, rather than on evidence of language in use. The present research is novel in its explicit, systematic application of polysemy tests to corpora. Polysemy tests indicate the presence or absence of distinct, discrete senses for a single word. Put differently, the tests indicate the dynamic between constant and contingent meaning – the tests help to distinguish meaning that fluctuates markedly under the influence of contextual information, and meaning that is much more consistently related to a given word. Alternatively, we can say that these tests gauge polysemy as opposed to vagueness: *polysemy* indicates that a word can convey, specifically, distinct or discrete senses, while *vagueness* indicates that a word does not specify for particular sense attributes.

The Quinian test was designed for identifying *ambiguity* (Quine 1960), but *polysemy* is an adequate term as well. The test is based on the notion that if a word can be simultaneously asserted and negated, it is polysemic.

3. He saw a crane, but he didn't see a crane.

As I have explained elsewhere (Mehl 2013), Example 3 can express that he saw a bird, but not a construction tool. The meaning of *crane* can vary between the first proposition and the second (negated) proposition, and if we accept consistent meanings for the other words in the example, then the variation in sense must be at the level of the word *crane*. The word itself can therefore be related to two distinct meanings, and is therefore polysemic. Example 3 cannot express that he saw a male bird but not a female bird; that information is not constant to the word *crane*; *crane* is therefore not polysemic with the two senses 'male bird' and 'female bird'. The Quinian test is the preferred fundamental polysemy test for Enfield's (2002) areal study of polysemy in Southeast Asian languages. It is also worth noting recent arguments by Kemp (2012) that, considering his broadly naturalist philosophy, Quine himself would not have insisted on the absolute scientific reliability of this test, but would instead have viewed it as a heuristic or exploratory tool. For the present study, it is viewed in the same way.

Geeraerts (2006 [1993]: 110) rightly notes that if the Quinian test (or as Geeraerts says, the 'p and not p' test) is accepted, then a 'p and p' test should also be acceptable, in which an assertion can be restated 'without redundancy or awkwardness'. Applied to Example 4, the result would be as follows:

4. He saw a crane and a crane.

Example 4 can be interpreted to mean 'He saw a bird and a construction tool', and may also be accepted as evidence for the polysemy of *crane*.

The independent truth-conditions test (Cruse 2004: 105) is conceptually very similar to Quine's test, but is executed in a different way. If 'a context can be imagined in which a *Yes*/*No* question containing the relevant word can be answered truthfully with both *Yes* and *No*' then the word in question is polysemic (Cruse 2004: 105). This test can clearly be mapped onto Quine's test (above).

5.   A. Did he see a crane?

B1. Yes, he saw a bird.

B2. No, he didn't see a construction tool.

The independent truth-conditions test also represents the sort of feasible interaction that could easily occur due to one interlocutor interpreting a word with the wrong distinct meaning.

The identity test (Zwicky and Arnold 1975; Kempson 1977: 130; Palmer 1981: 106; Cruse 1986: 62; Cruse 2004: 104) can be performed in multiple ways. One use of the test employs an anaphoric expression to refer to the word being tested, so that identity of meaning between the anaphoric expression and its antecedent is necessary. The word is polysemic if there are multiple senses, but those multiple senses must be attached to both the antecedent and the anaphoric expression. Example 6, below, illustrates this:

6.   Mary is wearing a light coat; so is Jane. (Cruse 2004: 104)

In example 6, *light* can refer to weight or colour in both instances, but not weight in one instance and colour in the other instance. *Light* is therefore polysemic with the meanings 'light in colour' and 'light in weight'. On the other hand, *light* is not polysemic for other features, such as age. It is perfectly acceptable for the above sentence to refer to an old coat that is light in colour in the first instance and a new coat that is light in colour in the second instance. Any meaning related to age can be attached to either instance of the word individually with no difficulty.

The identity test can also be performed using coordination, with a simple transformation of the above example into the following:

7.   Mary and Jane are wearing light coats.

Like Example 6, Example 7 can refer to two coats that are light in colour or two coats that are light in weight, but not one coat that is light in colour and another coat that is light in weight. Moreover, either coat individually may possess any number of other attributes not related to the word *light*, such as age.

According to the antagonism test (Cruse 2004: 106) 'a sentence which calls for two discrete and antagonistic readings to be activated at the same time will give rise to the phenomenon of zeugma'. *Zeugma* here indicates a sort of intuitive semantic dissonance, awkwardness, or humour arising from this simultaneous activation of multiple senses within a single word.[13] The antagonism test is therefore far more subjective and reflective than the other tests. If such a phenomenon arises, then the word is polysemic, with two discrete senses, as in Examples 8 and 9.

8. John and his driving licence expired last Thursday. (Cruse 2004: 106)
9. Conservationists encourage great care to be taken with cranes, and so do building contractors. (Mehl 2013)

The antagonism test relies on the fact that humourous, awkward, zeugmatic effects can only arise if two constant senses of a word are at stake. This effect indicates that the word is polysemic. The antagonism test sometimes employs anaphoric expressions similarly to the identity test: Example 8 is similar structurally to Example 7; alternatively, consider 'John expired last Thursday, and so did his driving licence', as structurally similar to Example 6, both of which rely on anaphora. Both the antagonism test and the identity test tend to rely on coordination or anaphora. The antagonism test requires more creativity,

---

[13] *Zeugma* traditionally, and perhaps strictly, refers to the actual activation of simultaneous senses, rather than the intuitive response (*cf.* OED, *zeugma*, n.). *Zeugma* can also refer to cases of coordination 'in which one element could not form a construction on its own', as well as to cases of syllepsis (Matthews 2007: 442). Cruse's (2004: 106) usage of the term is in line with Geeraerts's (2006 [1993]) usage, and is accepted in the present study.

and, it seems, more subjective reflection as well, than the other tests. In particular, a sense of the humour of a zeugmatic construction is highly subjective, and dependent upon the social and cultural background of the analyst.

Geeraerts (2006 [1993]: 111) also offers a conjunction/disjunction test for vagueness and polysemy: 'When a vague predicate is negated, the specifications of the predicate are both negated as well… On the other hand, in the case of polysemy, the negation need not extend to both readings at the same time'. Examples 10 and 11 illustrate this test.

> 10. If it is not a crane, then it is not a male crane *and* it is not a
>   female crane.
> 11. If it is not a crane, it is not a bird *or* it is not a construction tool.

The conjunction in Example 10 cannot be made a disjunct (that is, *and* cannot be replaced with *or*). In Example 11, in fact, either a conjunct or a disjunct is acceptable. The former demonstrates that *crane* is vague for gender, and the latter demonstrates that *crane* is polysemic with the senses 'bird' and 'construction tool'.

Geeraerts's (2006 [1993]) critical examination of polysemy tests is probably the most valuable and rigourous critique of these tests. Geeraerts (*ibid*.: 99) argues that 'the distinction between vagueness and polysemy… is unstable: what appear to be distinct meanings from one point of view turn out to be instances of vagueness from another'. Geeraerts shows that polysemy tests do not point to a clear, categorical, binary dichotomy between polysemy and vagueness. More specifically, he demonstrates that multiple polysemy tests can result in contradictory assessments of a term as polysemic or not. Geeraerts (*ibid*.: 115) provides the following examples.

> 12. Lady is a dog alright, but she is not a dog.
> 13. Fido is a dog and so is Lady.

Example 12 employs the Quinian test and can be interpreted to mean that Lady is a canine (insofar as dog_1 means 'canine' as opposed to some other animal), but she is not male (insofar as dog_2 means 'male canine' as opposed to *bitch*). If Example 12 is acceptable, and Geeraerts intuits that it is, then it evinces the polysemy of *dog*. Example 13, on the other hand, is designed to employ the antagonism test: if *dog* is polysemic, then Example 13 should be zeugmatic. Geeraerts intuits that it is not, in which case the Quinian test and the antagonism test can be seen to produce differing results: the former suggests polysemy for *dog* while the latter does not. Geeraerts (*ibid.*: 122) also cites Cruse's (1982) example for the verb *like*.

14. John likes blondes and marshmallows.

According to Cruse (*ibid.*), Example 14 is awkward because two senses of *like* are at stake: the first related to sexual attraction and the latter related to culinary preference, and this would generally be considered evidence for polysemy.

15. John likes blondes and racehorses.
16. John likes racehorses and fast cars.
17. John likes fast cars and elegant clothes.
18. John likes elegant clothes and expensive after-shave.
19. John likes expensive after-shave and vintage port.
20. John likes vintage port and marshmallows.

Cruse (1982) considers Examples 15 through 20 evidence for a continuum of meanings between 'sexual attraction' and 'culinary preference', such that *like* represents a spectrum of meanings from one to the other, rather than the two distinct meanings. Geeraerts (2006 [1993]: 101) concurs. In sum, Geeraerts (*ibid.*) advocates 'a reorientation of our conception of lexical meaning' and questions 'the objectivity of the methodology of lexical semantics', concluding

that the vagueness/polysemy distinction is not as distinct as it was once believed to be: words might not be polysemic or vague in a black-or-white way. Geeraert's argument is convincing: the distinction between vagueness and polysemy is certainly unstable, and these tests offer little in the way of objectivity.[14] I would further propose that recognising and then evaluating evidence of that instability in actual use, via corpora, can be an effective tool in semantic analysis: more specifically, an approach to language in corpora which seeks natural evidence of the types of juxtapositions shown in these polysemy tests might effectively evaluate the dynamic instability between vagueness and polysemy in actual use. As Geeraerts (*ibid*.) asserts, none of these tests is airtight in an absolute way, and he therefore challenges the use of the tests. I would assert, however, that dismissing the tests on those grounds demands too much of the tests. Traditionally, the tests were used to analyse the possible polysemy of a lexical item based on the linguist's own reflection. I have already discussed the problems with relying on reflection alone; and Geeraerts (*ibid*.) has convincingly argued for the instability of these polysemy tests when used strictly in relation to the linguist's own reflective processes. However, the application of these polysemy tests as heuristics in searching a corpus has not been fully explored. Indeed, Geeraerts (*ibid*: 148) acknowledges that his focus on problematic and unstable cases may exaggerate the instability of the tests that he is challenging. In a way, then, the present research responds to that acknowledgment, and aims to identify just how often such problematic and unstable cases occur in actual use. To reiterate, these tests should be seen not as programmes run in order to report a binary outcome but instead as tools that give results for linguists to interpret, like statistical tests. To return to Kempson's (1977) point, these tests can be employed in a systematic and consistent way; we can add to her point that the tests can also be exploratory

---

[14] *Objectivity* is a philosophically difficult term. It may be that *rigour* is the goal, rather than objectivity; I have already argued that introspective linguistics generally lacks rigour, and polysemy tests used introspectively are no exception.

and heuristic tools, with the understanding that many words will not simply appear either polysemic or not. For the present study, the aim is to tease out multiple senses as the first step towards an onomasiological study of each sense. By combining examples in use (the empirical data from the corpus) with reproducible tests, I aim to show that an effective method emerges. Corpora can provide evidence for words in natural use, evidence that resembles the polysemy tests above. Examples 21 and 22 appear in the ICE-GB corpus, and provide interesting evidence for the semantics of *give* in actual use (both examples are discussed further in 8.2.4).

21. **Give** him a long run and lots of steak from now too. [ICE-GB W2F-001 #142]
22. He **gave** the young couple his blessing and a rather elegant house to live in. [ICE-GB W2F-001 #052]

Neither Example 21 nor Example 22, in context in the corpus, are apparently intended as zeugmatic. Both examples occur in an unmarked way in a context that is not otherwise humourous or awkward, suggesting that the coordinated Direct Objects (*run* and *steak* in Example 21; *blessing* and *house* in Example 22) are not intended to be zeugmatic but deemed acceptable. This evidence would suggest that *giving a steak* and *giving a run* evoke a single meaning for *give*. This type of corpus evidence can be an extremely useful tool for the semanticist in identifying polysemy, or lack thereof, based on actual language and natural use. Geeraerts (2006 [1993]: 100) points out that future corpus work will be crucial in developing our understanding of polysemy; indeed, Geeraerts's own corpus work on polysemy (*cf*. Geeraerts 1997) has already been crucial in that regard. The present corpus study aims to make another contribution to our developing understanding of polysemy.

The present study investigates words in use, in corpora, and therefore asks what senses are in fact conveyed in practice, rather than what senses can possibly be conveyed as evidenced by a linguist's own reflective processes.

Unique meaning extensions such as poetic metaphors are catalogued in the present study, but the linguistic creativity they represent does not bear particularly on the research questions here, which relate to norms of use in each region. As discussed in 3.1, *what can be conveyed* by a given word might be infinite. In identifying discrete senses of a polysemic lexical item, therefore, Palmer (1981) warns against a careless over-proliferation of proposed meanings. Discussing the polysemic verb *eat*, Palmer (*ibid*: 100-101) offers the following warning:

> [We can] distinguish between eating meat and eating soup, the former with a knife and fork and the latter with a spoon. Moreover, we can talk about drinking soup as well as eating it. In one of its senses, then, *eat* corresponds to *drink*... If we decide, however, that there are two meanings of *eat*, we may then ask whether eating jelly is the same thing as eating toffee (which involves chewing) or eating sweets (which involves sucking)... if we are not careful, we shall decide that the verb *eat* has a different meaning with every type of food that we eat.

Palmer (*ibid*.) concludes that a monosemic bias is the rational approach – but implies that, in any case, he would not expect linguists to commit the error that he describes for *eat*. I discuss a monosemy hypothesis for *make* in particular in 6.1.

It is interesting that Palmer (1981) dismisses this overproliferation problem as unlikely to arise because this very issue has come to the fore in Cognitive Linguistic analyses of polysemy. Enfield (2002: 33) has criticized Cognitive Linguistics researchers for assuming that polysemy is pervasive in lexical semantics, when our default assumption should not be an extremely large number of senses for any given word. The 'polysemy everywhere' attitude may well be a reaction against the monosemic bias of generative

linguistics.[15] A monosemic bias, as Palmer (1981: 101) encourages, seems logically defensible, since it is also a bias toward ontological parsimony, or Occam's Razor, a principle that holds up well in the history of science (*cf.* Baker 2010). That is not to suggest that all words should or must be monosemic – only that ontological parsimony is a reasonable goal in describing a word's semantics.

Brugman and Lakoff (1988) argue for the polysemy of *over*. They employ eleven examples and argue for eleven polysemous meanings. I select four of their examples for discussion here, which adequately illustrate the issue:

> 23. The plane is flying over the hill.
> 24. Sam walked over the hill.
> 25. The bird flew over the yard.
> 26. The painting is over the mantle.

Brugman and Lakoff define the various potential meanings of *over* in terms of paths, regions, and orientations, and provide visual images for each propositions using *over*. Because those visual images illustrate entire propositions, it is impossible to clearly identify what meaning relates distinctly to *over* and what meaning relates to the other words in each example. Nonetheless, Brugman and Lakoff interpret differences between these general illustrations of entire propositions as differences in meaning between senses of *over*. Thus, in Example 23, Brugman and Lakoff argue that *over* entails a path (flight), a trajector (the plane), and a landmark (the hill). Example 24 is argued to entail a trajector (Sam) and a landmark (the hill), as well as 'contact' between the trajector and the landmark (ibid.). In example 25, Brugman and Lakoff argue that the landmark is now an extended area, and that this change therefore constitutes yet another sense of *over*. Finally, the authors argue that example 26 relates to the sense 'above' rather than a trajectory, and that this is

---

[15] On monosemic bias, see also Ruhl (1989).

yet another sense of *over*. The similarities with Palmer's (1981) argument on *eat* are apparent. Enfield (2002: 32-3) has directly criticized the approach to polysemy exemplified in the examples above (specifically in reference to Lakoff, 1987: 416-61) insofar as it does not rigorously address *how* distinct these meanings might be.[16] Enfield's (2002: 32-3) emphasis on the distinctness of a word's meanings is very much in line with the emphasis of the present study. Enfield (2002: 32-3) relies primarily on the Quinian test to respond to Lakoff (1987); I reconstruct his challenge and extend it to Brugman and Lakoff's (1988) work, which is parallel to Lakoff (1987), but not identical. If *over* in Examples 23 and 24 were polysemic according the criteria forwarded in the present study, then we would expect to be able to apply the Quinian test to those examples as follows:

27. The plane is flying over the hill but not over the hill.
28. Sam walked over the hill but not over the hill.

If *over* relates to distinct senses like the ones described by Lakoff and Brugman (1988), Example 27 should be comprehensible as 'The plane flew over the hill (not in contact with the hill) but not over the hill (in contact with the hill)'; and Example 28 as 'Sam walked over the hill (in contact with the hill) but not over the hill (not in contact with the hill)'. That distinction is very likely not

---

[16] For Brugman and Lakoff (1988), the distinctness of the various meanings of *over* is not assumed at all, and thus the range of meanings might be seen as either vague or polysemic in a traditional sense. Brugman and Lakoff employ the term *polysemic* to describe a word with 'related' meanings, but they do not explicitly define *polysemy* as involving discrete or distinct meanings: indeed, they are working within a paradigm that accepts a wide array of meanings for a word, which may not be discrete at all, but may be vague instead (*cf.* Lakoff 1987). As Geeraerts (1997: 19) has explained, Cognitive Linguistics has allowed for thorough explorations of lexical vagueness as a valid field of linguistic investigation. Brugman and Lakoff (1988), nonetheless, do not claim to be investigating vagueness – they instead use the term *polysemy*, but only in a roughly defined way.

discernible in Examples 27 and 28. Instead, the verb *flying* conveys the lack of contact in Example 27 and the verb *walked* conveys the contact in Example 28. The preposition *over* accommodates those meanings, but does not itself distinguish them. The Quinian test suggests that the meaning 'contact/no contact' is not related to the preposition *over* but to the main verbs of each sentence. In addition, the identity test might encourage an example like the following.

29. Did the dog go over the fence or the hill?

In Example 29, the question refers to two possible paths for the dog's travel. Presumably, a dog going can go over a fence by jumping over the fence (without contact) or by climbing over the fence (with contact), and a dog can only go over a hill by climbing (with contact). In either case, the presence or absence of contact is determined by an understanding of dogs, fences, and hills, and does not seem to be distinct within the preposition *over*; *over* accommodates those meanings but not distinguish them. In addition, *over* would seem to be able to express both *contact* and *no contact* simultaneously for the coordinated Complements *the fence* and *the hill*. According to the criteria forwarded for the present study, therefore, *over* does not seem to be polysemic for those meanings. That is, *over* is vague for those meanings. Even more importantly for the present method, however, is corpus evidence. A corpus could provide evidence for polysemy in use between the senses 'in contact' and 'not in contact' for the preposition *over*. Regardless of whether we accept the Quinian test as evidence of polysemy, or Lakoff and Brugman's (1988) evidence of polysemy, an additional corpus investigation would be an interesting complement to Lakoff and Brugman's (*ibid*.) study.

There is another type of variation in meaning that has been forwarded as polysemy, particularly in World Englishes research, and the polysemy tests described here can be useful tools for engaging with it. Adejare and Afolayan (1982; quoted in Platt *et al*. 1984) present an intriguing proposal on the precise

semantics of *dog* in Nigerian English (see 4.3 on the hypothesis that common English words will vary semantically from one region of the world to another). They suggest that *dog* presents shades of meaning in Nigeria that do not exist in the UK, because *dog* in Nigeria refers not to a pet but to a source of meat and a tool for hunting. In this case, language users' knowledge related to *dog* varies between the two locales, because the real-world role of dogs varies between the two locales. Whether it follows that *dog* has a discrete meaning or meanings in Nigeria that it does not have in Great Britain, however, must be investigated. The lexical item *dog* in English is vague as to the attributes 'meat' and 'hunting tool': a speaker in the UK can assert that he or she has eaten dog and the assertion can be readily understood precisely because *dog* does not specifically refer to an animal that is or is not eaten (corpus evidence could be sought for corroboration of that claim). If Nigerian English had developed a sense for *dog* such that the assertion 'I have a dog that sleeps at the foot of my bed and plays with my children in the evening' was nonsensical in the way that 'I have pork that sleeps at the foot of my bed and plays with my children in the evening' is nonsensical, then Nigerian English *dog* would be a definite instance of semantic variation. Adejare and Afolayan (*ibid*.), do not, apparently, make such a systematic claim, and *dog* does not seem to be polysemic in that way.[17]

3.4. Polysemy and corpus frequency: The case of light verbs

The present study relies on corpora for evidence of polysemy and semantic alternations in natural use. Observations of the corpus frequencies and meanings in use of polysemic lexical items have contributed to a valuable body of academic literature. Within that body of literature, the study of verbs with *delexical* or *light* senses as well as concrete senses is particularly noteworthy, and relevant to the present study. In fact, the verbs investigated here (*make*,

---

[17] If there is no apparent semantic variation in this case, there may still be a pertinent question regarding usage preferences: how often in each locale do people refer to 'dog meat' as dog and how often do they refer to it as meat or some other word? This information can be quantified, and such an analysis could be meaningful in terms of linguistic norms.

take, and *give*) were selected in part because they have already been studied as examples of polysemic verbs with concrete and light senses. In this section, I first review some key literature on light verbs, then discuss the nature of light verbs as a semantic question, and finally discuss the existing literature on frequencies of concrete and light senses of polysemic verbs in use. The goal is to provide a background from which to understand both the selection of *make*, *take*, and *give* in the present study, and the findings and discussion on the frequencies of the concrete and light senses of those verbs.

Algeo (1995: 213) refers to light verb constructions as 'expanded predicates', noting that they are 'at the boundary between grammar and lexis, partaking of some of the characteristics of each' (*ibid*: 213). It is perhaps no surprise, then, that grammarians in particular have examined these constructions for nearly a century. In 1926, Poutsma (1926: 394-400) identified 'group-verbs' as transitive verbs in which the entire construction is semantically equivalent to a verb that is related to the Direct Object. His examples include *give an answer* and *make an answer*, both equivalent to *answer* (v.), and *give a reply* and *make a reply*, both equivalent to *reply* (v.), among many others (*ibid*: 394). For Poutsma (*ibid*: 394) the verbs *make* and *give* in these constructions are a sort of 'connective', while the Direct Object is the word that actually conveys the action. Hence, in *give a reply*, *give* is seen as semantically less significant than *reply* (and thus *light*). Poutsma (*ibid*: 394) reflects on the semantics of these constructions:

> It is but natural that the vagueness of the verb entering in these group-verbs is not equally pronounced in all of them. Nor is it possible to tell to what degree a verb should have weakened semantic significance to justify its being called a mere connective.

The first portion of Poutsma's claim here is ambiguous: it is not clear whether he is arguing that various instances of a single light verb may be more or less vague depending on the rest of the construction (and in particular the Direct

Object); or whether he is arguing that a given light verb may be more or less vague than other light verbs; or, perhaps, both. Soon after Poutsma's work, Jespersen (1954) identified these constructions as composed of a Direct Object which is a 'substantive' – a noun derived from a verb – and an 'insignificant' or 'light' verb such as *have* or *take* (1954: 117). The term *light verb* has become relatively standard since Jespersen's time. Jespersen's (*ibid*.: 117) primary examples include 'have a bath', 'have a smoke', and 'have a cry', and Jespersen's description is comparable to Poutsma's (1926). In contrast to Jespersen's (1954) argument but echoing Poutsma's (1926) interest in the degrees of vagueness of light verbs, Wierzbicka (1982) argues that *have* conveys discernible meanings even in light verb constructions. Similarly, Brugman (2001) has argued that a given light verb may be more or less vague depending on the Direct Object it takes, but that all light verbs carry some meaning. As I discuss in Chapters 6 through 8 in relation to *make*, *take*, and *give*, these verbs do in fact differ from each other insofar as they do not appear equally light. Huddleston and Pullum (2002: 290-6) identify these constructions as 'light uses of verbs'. Their semantic analysis is not systematic, but their main justification for the 'semantic lightness' of these verbs lies in the assertion that the verbs themselves contribute little to the meaning of the construction compared to the contribution of the Direct Object (ibid: 290). That is, the meaning of the construction is interpreted as equivalent to the related verb of the Direct Object, which they call the 'associated verb'. They identify the 'main' light verbs (without providing evidence for their 'main'-ness) as *give*, *make*, *have*, *take*, and *do*, in that order.

In the literature cited above, light verbs are identified as transitive constructions, but primarily characterized by the meaning conveyed by the verb and the Direct Object in relation to each other. This meaning is not, strictly, lexical semantic meaning, as it involves at least two lexical items, and in particular, at least two lexical items that can be arranged in various ways in relation to each other (e.g. in active or passive constructions, or with modifiers between them, as I discuss in 6.2.4, 7.2.6, and 8.2.4). Thus, while some key

lexical semantic issues already presented in this chapter do apply to the study of light verb constructions, some of the issues already presented do not apply to light verb constructions in the same way that they apply to the lexis, or types of lexis, discussed already. I elaborate on the similarities and differences below.

First, it is perfectly reasonable to explore either the meaning in use of light verbs, or to reflectively derive possible or imaginable meanings of light verbs, just as for other lexis. The present study examines meaning in use, as stated in 3.1. Second, it seems that it is possible to explore relatively constant and relatively contingent meaning of light verb constructions, with caveats. The simplest instance of a light verb construction can be considered a grammatically canonical one: in active voice with no modifiers. Because such a construction is already defined primarily in terms of its meaning relationships (i.e. the Direct Object's meaning is more salient than the verb's meaning for interpretation), it seems perfectly reasonable to explore the relatively constant and relatively contingent meaning in use of, for example, *make a decision* just as for individual lexical items such as *decide*, or for *cat*, *dog*, or *love*. There is no reason that we cannot explore more consistent and more unusual meanings in actual use of light verb constructions, and there is no reason that we cannot identify instances in natural language where semantic categories seem to be blurred (e.g. where two putative senses of *make a decision* might both come into play). One of the typical instantiations of light verbs in which semantic categories might be blurred is via passivisation or grammatical modification; that issue is discussed further in chapters 6, 7, and 8, in relation to each verb in question.

The approach to corpus onomasiology in 3.2 raises distinct difficulties in relation to light verbs. Generally, in the present study, semantic alternates of *make*, *take*, and *give* are identified by first cataloguing grammatical dependents of those verbs (i.e. their Direct Objects) and then identifying comparable verbs that occur with the same Direct Objects. With light verb constructions, this approach is limited in scope, and I therefore employ two approaches to light verb constructions in the corpora. The first is the study of alternation in the

verbs themselves. In very few cases (e.g. *make a decision* and *take a decision*; *give support* and *provide support*; *give information* and *provide information*), the verb in a light verb construction alternates with some other comparable verb. Although limited in scope, this approach is not limited in value: those few alternations reveal some important insights into the nature of light verb constructions, the differences between them, and the discreteness of their semantics, in addition to providing evidence on alternation in use (see 6.2.10, 7.2.21, 8.2.10, and 9.1). Beyond those few cases, the present study operationalizes the onomasiological investigation of light verbs by relying on the definitional equivalency between the light verb constructions and the verb related to the Direct Object (e.g. the equivalency between *make a decision* and *decide*) as a basis for studying alternation, observed probabilities, and preferences between the three sampled corpora. This second approach, however, does not address degrees of discreteness and the nature of these constructions as polysemic, but only addresses alternation in use. These two approaches therefore address separately these two key questions: discreteness of meaning, and alternation in use. In the second approach, the data can be viewed as providing evidence for or against the hypothetical equivalency between the two alternates. Such issues are weighed in the discussions in Section 2 (see, for example, 7.2.13). This prevents the blind circularity of defining light verbs in terms of an equivalency between the light verb construction and verb related to the Direct Object, and then affirming that equivalency unquestioningly by cataloguing instances of each in the data. The hypothesis of that equivalency is forwarded in the present study in order to engage fully with existing literature on light verbs and to address directly the second of the three research questions forwarded in chapter 1: Do findings on cognitive salience of light verbs correlate with findings on corpus frequencies of light verbs, given multiple means of measuring corpus frequencies? This question is crucial to the present study, and can only be investigated by building on established definitions of light verbs.

Finally, polysemy tests are possible for light verb constructions as for individual lexical items. A light verb construction can readily replace any of the individual lexical items in the polysemy tests described in 3.2. That said, corpus evidence resembling that of polysemy tests (as proposed in 3.3) does not appear for light verbs in the corpora.

Like the present study, some recent research has approached polysemic verbs with light senses quantitatively, by measuring corpus frequencies of the verbs' various senses (*cf*. Gilquin 2008; Nordquist 2009; Werner and Mukherjee 2012). Gilquin (2008) uses the *Switchboard* corpus of spoken English to measure frequencies of each sense of *take* and *give*. Gilquin's study is semasiological, and it relies on a *pmw* baseline, so it relates (implicitly rather than explicitly) to exposure rates rather than to selection processes. She then compares those frequencies to data on the cognitive salience of each sense as derived from elicitation tests. To do that, native speakers were asked to generate the first sentence that came to mind using the target verbs *give* and *take*, as well as several other polysemic verbs. Effectively, then, Gilquin's research question can be seen as something like the following: Do exposure rates to each sense of *give* and *take*, relative to a semasiological baseline in English speech, correlate with readiness of elicitation of each sense of *give* and *take*? (Her stated research question is less precise, insofar as it does not deal with the nature of frequencies or types of frequency measurements.) It is worth noting that both the corpus study and the elicitation tests were semasiological, though not explicitly so: both began with a word in question, and then moved towards the various meanings that word can express. Gilquin (*ibid*.: 248) concludes that 'while language [in corpora] shows a strong preference for abstract, grammaticalized senses such as the delexical [light] use, the senses most often elicited are more concrete'. That is, light senses of verbs in corpora are far more common than concrete or other senses of those verbs in corpora, but when asked for example sentences using verbs like *take* and *give*, respondents are most likely to produce concrete senses of those verbs. This is a remarkable fact that raises a question: Given that language users are exposed to light senses of verbs far more often

than concrete senses, why should concrete senses come to mind first? This question connects a particular type of corpus frequency – an exposure rate – to the nature of cognition. Gilquin's (2008) findings serve as a foundation for my decision to investigate *take* and *give* as well as *make*, another verb with both concrete and light senses. The corpus portion of her study was replicated by Werner and Mukherjee (2012), who concluded that in written samples of English from India, Sri Lanka, and Great Britain, the light senses of *take* and *give* were not consistently most frequent. That replication study itself reflects the importance attributed to Gilquin's (2008) study. The present study includes an analysis with a semasiological baseline that resembles Gilquin (2008) and Werner and Mukherjee (2012), but also, most importantly, an analysis against an onomasiological baseline as well.

Gilquin (2006; 2008; in Arppe *et al*. 2010) has been a significant contributor to a discussion on semantic frequencies in use and cognitive salience, and has pointed out that both frequency and salience are often used as measures of *prototypicality*, or as tools for identifying a prototypical sense. That is, Gilquin (2008) has underlined that prototypicality is often employed in Cognitive Linguistics literature either to refer to cognitive salience as evidenced by elicitation tests, or to frequency in use as evidenced by corpora, but that those two references may not coincide, as in the case of verbs with light and concrete senses. As Gries (2006: 9) notes, identifying a prototypical sense is 'a problematic question since many different subjective and conflicting criteria can be brought to bear on this issue,' including both salience tests and frequency analyses. Indeed, *prototype* has been a contested term over the forty years since its introduction. Prior to being adopted by Cognitive Linguistics, prototypicality was first forwarded in psychological work by Rosch (1973), and it referred to the 'clearest cases [of category members], best examples… [which] serve as reference points against which other category members are judged' (Rosch 1975: 544-5). For Rosch (1973), prototypes were not individual senses of polysemic words but various concrete referents of a given category. For example, Rosch (1973) showed that a specific shade could be seen as

prototypical red for a given language community; likewise, specific types of
birds could be seen as prototypical birds, and so on. Rosch (1973; 1975a;
1975b) used various methods for identifying prototypes. These include
subjects' response speeds in elicitation tests, where the quickest response is the
prototype; subjects' intuitive categorizations, where the example that subjects
identify as prototypical is the prototype; and subjects' reference to prototypical
category members in describing non-prototypical category members. The key
point is that some members of a category, such as *bird* or *red*, may not share
any features with each other, so that the category itself is not strictly logically
definable and bounded – there is no 'single set of criterial (necessary sufficient)
attributes' (Geeraerts 1997: 11). Prototypicality can thus be seen as the
cognitive organizational principle by which dissimilar referents can be seen as
members of a single category. From the beginning (*cf*. Rosch 1973),
prototypicality was identified by psychological tests using a variety of criteria,
and a recognition of prototypicality depended initially upon a recognition of
prototypes as reference points by which categories could be understood, and by
which non-prototypical members of a category could be understood and
categorized. Cognitive Linguistic researchers then began to describe the various
senses of a polysemic lexical item in terms of prototypicality: in that adaptation
of the theory, the lexical item is viewed as a cognitive category, and the
different related senses of that lexical item are seen as the members of the
category (*cf*. Geeraerts 2006 [1989]).

Bearing in mind that it has been defined and measured in many different
ways, how can data on frequency in use inform our understanding of
prototypicality? Corpus frequencies are a recently adopted tool for measuring
prototypicality (*cf*. Geeraerts 2006 [1989]; Gilquin 2006, 2008; Gries 2006;
Heylen *et al*. 2008; Arppe *et al*. 2010). According to Geeraerts (1988: 222)
frequency of use can be a 'heuristic tool in the pinpointing of prototypes'.
Taylor (2003: 54) asserts that prototypicality correlates with semasiological
salience, indicated by semasiological frequency, with semasiological baselines,
in use – what I have called exposure rates (see 2.2). That is one possible way of

relating frequency to prototypicality, but Gilquin (2008) has shown that semasiological frequency in use does not correlate with elicitation test results. In determining the role of frequencies of any kind in recognising prototypicality, the term *entrenchment* is useful. *Entrenchment* refers to 'the degree to which the formation and activation of a cognitive unit is routinized and automated... Deeply entrenched cognitive units are more likely to become cognitively salient than less well entrenched ones' (Schmid 2007: 119-20). Routinization of cognitive units, or specific meanings of words, is dependent upon exposure to those meanings: entrenchment emerges from frequency of exposure, and salience (which is one measure of prototypicality) emerges from entrenchment. Thus, frequency of exposure should be expected to correlate with salience; but, again, Gilquin (2008) shows that it does not in the case of verbs with both light uses and concrete senses. As an alternative to exposure rates or semasiological frequencies, Geeraerts (2010: 201) has forwarded the notion of *onomasiological salience* as the preference for a word over its semantic alternates exhibited in corpus data. He has innovatively proposed that onomasiological salience 'can be equated with the notion of "entrenchment".' (Geeraerts 2010: 201). As discussed in 2.1 and in 2.2, however, onomasiological studies of corpus data are relatively rare and extremely demanding. An alternative to Gilquin's (2008) study might have investigated *give* and *take* onomasiologically in a corpus, and perhaps also performed elicitation tests onomasiologically, in which subjects are given a concept to express (possibly via a picture or video) and then asked to describe what is happening, resulting in examples of not just *give* and *take*, but also, potentially, a number of semantic alternates.

The present study does not set out to measure onomasiological salience experimentally, but it engages with Geeraerts's idea of onomasiological salience by investigating the three verbs *make*, *take*, and *give*, which have both light and concrete senses, onomasiologically. In section 9.3, I return to Gilquin's (2008) study in particular, and to its implications for Cognitive Linguistics. In that section, I also employ Geeraerts's (2010) hypothesis of

onomasiological salience to discuss the present data in light of Cognitive Linguistic theoretical frameworks.

4. WORLD ENGLISHES

4.1. Frameworks for World Englishes

In Chapter 1, I stated that the present study is not first and foremost a study of World Englishes – instead, it is a corpus study into lexical semantics, investigating the possibility of semantic variation in use between varieties of English. In this section, I discuss the tradition of World Englishes research. I begin by describing the aim of any framework of World Englishes: to adequately generalize and describe linguistic trends in multiple varieties of English around the world. I then describe some of the most commonly cited frameworks, as well as some lesser-known ones. In my view, frameworks of World Englishes tend to generalize about multiple varieties of English worldwide based on social and political features more than linguistic features such as grammar, semantics, or pragmatics. In fact, social and political features are ultimately tied to linguistic features, but linguists must sometimes choose to approach language study more from the social and political perspective, or more from a linguistic perspective, and World Englishes research tends towards the former. I suggest an alternative to that basis in social and political facts, and advocate for a cyclical approach that moves between observations of linguistic data, and generalized theoretical frameworks for describing World Englishes.

In general, theories of English varieties around the world propose for each variety:

> a distinct rule-governed system of the English language, sustaining and sustained by a community of users spread across the area, who share the norms by which its rules are determined and for whom that system will have some kind of self-identificational value by virtue of the fact that it serves their distinct semiotic and pragmatic needs. (Begum and Kandiah 1997: 191)

Defining norms in the context of World Englishes, Melchers and Shaw (2003: 30) echo Begum and Kandiah (1997: 191) by proposing 'an implicit set of rules speakers appear to use for what it is appropriate to say in what grammatical or social context'. I would forward, therefore, the following important question for researchers of English varieties: what rules (if any) exist for each variety of English, and how are each variety's rules similar or different to each other variety's?

Although British colonial figures occasionally collected glossaries or lists of linguistic curiosities in various colonized regions (*cf*. Yule and Burnell 1886; Wilson 1940 [1885]), rigorous research into rules or norms for English around the world did not emerge until the 1960s in the form of case studies of specific features in individual varieties (*cf*. Kachru 1965, 1969). Kachru (1969), for example, argues for a detailed investigation of South Asian English on multiple linguistic levels, including collocational studies. According to Kachru, regional collocational norms that differ between varieties can hinder intelligibility. *Rules* or *norms* in this case clearly include a variety of linguistic characteristics – grammatical, social, lexical, semantic, and so on – that are observed to be shared across a variety.

It is important to carefully define *linguistic variety* when discussing World Englishes. A linguistic variety, as a whole, is a theoretical abstract not empirically observable in its entirety because it contains far more data than we can realistically monitor (*cf*. Enfield 2003: 4). A variety such as *Singapore English* or *British English* is an idea constructed by the linguist (or casual observer) based on generalized observable linguistic features in use in a region. The idea is constructed by the linguist based on apparent norms. By observing and identifying linguistic norms, features that occur consistently, the linguist creates generalizations that develop into a theory about a language variety. Thus, the linguist describes *British English*, and that description constitutes a theoretical construction abstracted from observations about speakers of English in Great Britain. *British English* is not empirically observable as such – all of the language produced in Great Britain cannot be collected and studied – but

the idea of British English is maintainable as a construct. That construct is based on all the observable evidence about regional norms. Generalization about that evidence allows for the construction of a theoretical framework (generalization is key to theory construction), and theory construction must avoid both over-generalization, which ignores important details and internal variation, and under-generalization, which fails to construct a theory at all and instead just describes the evidence.

In order to meaningfully connect pieces of evidence about regional norms, and to observe the relationships between them, various frameworks have been proposed, which constitute systems for describing English worldwide. In the following pages, I describe an array of frameworks that have been established for describing English varieties worldwide, from the first frameworks of the 1970s through today's frameworks, and from the most-cited frameworks to some lesser-known ones. I would like to show in the present discussion that these theoretical systems have moved a considerable distance away from describing and generalizing about linguistic evidence (grammatical, phonetic, semantic, pragmatic, and so on), and instead depend largely on social and political characteristics of regional varieties.

Early distinctions between varieties of English worldwide were codified by Quirk *et al*. (1972), who identified three categories of English varieties: varieties where English is spoken as a native language; varieties where English is spoken as an additional language; and varieties where English is spoken as a foreign language. Already, it is clear that this theoretical framework for English varieties worldwide is far removed from a rigourous study of linguistic norms in use. Quirk's system is obviously built on generalizations about societies and their socio-political structures, including a region's official language and its language policy and language education system. Generalization is an essential, definitive element of theory construction; in order to establish an overarching framework for describing many variables and features, generalization is necessary. Quirk's (ibid.) model, however, may over-generalize the social and political aspects of a region even while it tends to ignore the linguistic norms of

a region: Schneider (2007: 12) notes that Quirk's (1972) model fails to account adequately for, or even accommodate, internal sociolinguistic variation, *e.g.* speakers of English as a foreign language living within a 'native-language' country like the US or UK. Nonetheless, this system of differentiating English varieties continues to be employed in linguistic studies (*cf.* Görlach 1995 and, interestingly, Schneider 2011).

The earliest diachronic or process-oriented model for the development of English varieties worldwide is Moag's (1982) 'life cycle of non-native Englishes'. Moag proposes a common process for the development of English varieties around the world, which I summarize below.

I. English is transported to a new locale, where there is only limited contact between English speakers and locals.

II. A new indigenous variety begins to emerge as local words are borrowed into English and as English eventually becomes the default code for communication with native English speakers.

III. Use of English expands to multiple registers and social functions, and internal variation in the local variety begins to appear.

IV. The variety is institutionalized via literature, dictionaries or glossaries, and a homegrown English language education system and media.

Moag's (*ibid.*) model is valuable as it proposes the novel idea of a common diachronic process that results in the synchronic variation of English worldwide today. It is noteworthy that the model includes some specific linguistic features such as lexical borrowing and the emergence of internal linguistic variation, but that the stages are predominantly sociolinguistic and socio-political (*e.g.* the establishment of a state-run English language education system). Moag's model shares this socio-political perspective with Quirk's (1972) model. Indeed, attention to the sociolinguistic and socio-political elements of English's

development worldwide is apparent in most theoretical models in the field, as I continue to show below.

The next model to emerge is probably the most cited by researchers today: Kachru's (1985) synchronic 'concentric circles' model of World Englishes. Having researched specific linguistic and sociolinguistic attributes of South Asian English for two decades (*cf.* Kachru 1965, 1969, 1975, 1983), Kachru recognized the need to summarize and generalize 'the sociolinguistic attributes of English in its international context' by establishing a theoretical model for describing the vast number of multilingual settings for English around the world (Kachru 1985: 11). Kachru's model was therefore a deliberate attempt at generalizing a social theory to complement research on English linguistics (including grammar and lexis) in use. To that end, Kachru (*ibid*.: 12) recognized three 'circles' of world Englishes: the inner circle 'refers to the traditional bases of English… where English is used as a primary language'; the outer circle consists of multilingual areas colonized by the inner circle, where English is being or has been 'institutionalized' and is used for a variety of functions with significant internal variation; the expanding circle includes regions with no history of colonization, but with systems for studying English as a foreign language. Kachru (*ibid*.: 16) also identifies the inner circle as 'norm-defining', the outer circle as establishing some unique internal norms that compete with the inner circle's norms, and the expanding circle as relying solely on norms established by the inner circle.

Kachru's (1985) three circles would seem to map nicely onto Quirk's (1972) three categories, though Kachru (1985) does not claim any relationship between the systems. In fact, this parallel between Quirk's and Kachru's models is particularly interesting given the opposing viewpoints they argued for in their famed debate about World Englishes in *English Today* (*cf.* Quirk 1990; Kachru 1991). Although each model presents three levels for differentiating the sociolinguistics of English worldwide, Quirk's model was apparently never intended to legitimize non-standard or non-normative varieties as valid, while

Kachru built his representation with the goal of legitimizing those very varieties.

Kachru's (1985) model is clearly based on social, political, and historical facts, more than linguistic ones, and is explicitly aimed at describing English varieties in social terms, more than linguistic ones. Moreover, although Kachru does not initially define his three circles in terms of language policy (*i.e*. locales where English is or is not an official language), he states that the outer and expanding circles 'cannot be viewed as clearly demarcated' precisely because nations may grant and then revoke official language status to English over time (ibid: 12-13). Kachru thus strongly implies that official status (a fact of politics and language policy) is a defining factor for the outer and expanding circles. In this sense, Kachru's model, like Quirk's (1972) and Moag's (1982), can be seen as largely socio-political. More recently, Kachru (2002: 1) identifies 'the distinctiveness of the sociolinguistic contexts of Englishes, their diffusion and location in world contexts' as his first emphasis in his co-edited volume, *Handbook of World Englishes*. The emphasis on sociolinguistics and socio-political elements therefore continues to be foregrounded.

Schneider (2007: 13) has criticized Kachru's (1985) three circles model, asserting that 'the exact criteria for inclusion in... [Kachru's] categories are not always clear'. I would agree with Schneider, as Kachru's three circles are not defined by precise and explicit linguistic or even socio-political criteria, and his social and political criteria, like Quirk's (1972), are in reality more complex than his model recognizes. Like Schneider (2007), Begum and Kandiah (1997) also criticize Kachru's (1985) model for failing to acknowledge the variation within regional varieties of World English; that is, Kachru's model *overgeneralizes*, and does so based on social criteria rather than linguistic criteria. In Kachru's defense, however, his model serves an explicit purpose related to sociolinguistic attitudes and language policy: the model works to establish the validity and integrity of English varieties around the world, regardless of their differences from the British standard, and to create a foundational sociolinguistic framework for discussing those varieties. The

model thus represents a more forceful political turn from some of Kachru's earlier work, which pioneered the study of precise linguistic attributes (such as verb collocation) of South Asian English (*cf*. Kachru 1965, 1969, 1975, 1983).

Schneider's (2007) model, perhaps the most influential since Kachru's (1985), includes more detailed descriptions of common linguistic developments in English varieties worldwide. It is nonetheless built on notions of 'social identity and its construction and reconstruction by symbolic linguistic means' (*ibid*.: 26), and on post-colonial theory, more than on linguistic attributes such as grammar, phonetics, semantics, and pragmatics. Schneider proposes 'a fundamentally uniform developmental process, shaped by consistent sociolinguistic and language-contact conditions' (*ibid*.: 5), which evokes both Kachru's (1985) definition of post-colonial varieties and Moag's (1982) notion of a variety's life cycle. Elsewhere, Schneider (2011: 136) refers to his theory as 'a comprehensive and coherent perspective on globalization', an essentially sociological and historical process. That said, Schneider (2007: 56) proposes, in more linguistic detail and with greater complexity than previous models, five stages in the development of World English varieties, which I summarize here:

I.   Foundation – An English variety or varieties arrive in a new territory. English speakers and locals meet, a *koine* or *pidgin* may emerge, and limited borrowing begins.

II. Exonormative stabilization – The English-speaking settlers become stable in the locale and develop a local identity. English is a prestige language, but also associated with exploitation, and the norms for English are set by Great Britain or the United

States. Lexical changes occur in both the settler and indigenous population.[18] Syntactic and morphological changes emerge.

III. Nativization – The region becomes increasingly independent from Great Britain or the United States, culturally and politically. Settler population is divided between the conservative, who continue to identify with Great Britain, and the innovative, who identify with the new locale. Identity gap between settler and indigenous community is reduced. Heavy lexical borrowing ensues, and pronunciation, morphology and syntax change significantly.

IV. Endonormative stabilization – A new local cultural identity is defined, followed by a new self-confidence in local linguistic traits. Codification of those traits emerges.

V. Differentiation – Smaller sub-communities in the locale develop their own linguistic traits within the regional variety, related, for example, to age or class.

As the above summary shows, Schneider (*ibid*.) does describe elements of phonetics and grammar, and lexical borrowing processes, alongside post-colonial processes of identity construction that form the foundation for his framework.

Schneider (2007) goes on to describe each stage's attributes more specifically in terms of:

a. history and politics

---

[18] Schneider (2007) refers to an *indigenous population*, in the singular. This would seem to neglect the fact that in many locales where English speakers settled, multiple local populations already lived side by side. In the case of Singapore, lexical changes in Phase 2 would presumably have impacted multiple indigenous populations, including speakers of multiple Chinese varieties, Malay, and the already existing contact variety of Baba Malay (*cf*. Ansaldo 2010).

b. identity construction

c. sociolinguistics of contact

d. linguistic developments and structural effects

The fourth attribute above addresses linguistic norms, and constitutes a relatively small section of his framework. Schneider (2007: 55) notes that his model functions as 'a generalization which abstracts from many complexities and details'. Although he explicitly states that his work is not politically driven (*ibid.*: 20), his model nonetheless depends first and foremost on social and political criteria (*ibid.*: 8).

Schneider's (2007) model is also unique in its foregrounding of *identity construction* in the development of World Englishes. *Identity construction* has come to play a significant role in sociolinguistic theory, and refers to 'the systematic establishment and signification, between individuals, between collectives, and between individuals and collectives, of relationships of similarity and difference' (Schneider 2007: 26; *cf.* Jenkins 1996, Woodward 1997). According to Wong (2007: 759), Schneider does not acknowledge that a convergence in linguistic form need not be accompanied by a convergence in culture or values. Indeed, Scheider's (2007) model posits a causal relationship between shared identity construction, or identity convergence, and convergence of linguistic features. In the present discussion, Schneider's (*ibid.*) claim may represent the overpowering force of social and political factors in theories of World Englishes. The complex relationship between converging and diverging linguistic forms and converging and diverging identities is rich ground for research, but we need not assume that all social convergence causes linguistic convergence, nor that all linguistic convergence is caused by social convergence.

While Kachru's (1985) and Schneider's (2007) models are the most widely embraced systems for categorizing World Englishes, a handful of other researchers have proposed alternatives. Görlach's (1988) model proposes an idealized standard international English used by a huge portion of the world's

population; followed by national and then regional standards used by fewer people. He then proposes a category of regional nonstandard varieties where English is spoken as a second or foreign language with few norms and a great deal of internal variation, and a final category of nonstandard mixed varieties including pidgins and creoles. Like those already discussed, Görlach's (*ibid*.) model also focuses on sociolinguistic or socio-political processes, including socio-political standardizing forces such as formal education.

Brutt-Griffler (2002) argues for the importance of language acquisition in a process-oriented model of World Englishes. Her model is explicitly and deliberately sociolinguistic, but she innovates by addressing not only questions of language spread, language contact, and language change, but also questions of language acquisition. Brutt-Griffler (*ibid*.) employs a notion of 'macroacquisition', or foreign language acquisition by entire speech communities rather than just by individuals. In Brutt-Griffler's (*ibid*.: 110) account of World Englishes, a crucial stage is the 'stabilization of bilingualism through the coexistence of a world language [such as English] with other languages in bilingual/multilingual contexts'. In the beginning of that stablization process, Brutt-Griffler (*ibid*.: 133) actually denies that any target language actually exists – that is, the variety is not at all exo-normative but instead is endo-normative: in the early stages of the emergence of a World English variety 'there is no fixed target language, but the language variety rather develops from the SLA [second language acquisition] process itself… The target language only *develops* as the result of the SLA process, rather than preceding it'. The local English learner has very little sense of an exonormative standard in any real way, and therefore cannot adhere to it. This theory is particularly important for the present study, as I discuss further in 9.2.

Anchimbe (2009) has forwarded an extremely valuable and relatively unique ecological perspective on World Englishes. Anchimbe (*ibid*.) explicitly interprets Schneider's (2007) model as a maturation model, whereby English varieties proceed through stages towards a sort of maturity, and claims that there has been an acceptance of such maturation models both implicitly and

explicitly throughout the field of World Englishes. However, he argues against this kind of 'maturation process'. Anchimbe (2009: 337) asks:

> is there any such thing as *language maturity*? Do languages actually grow in the same way as living organisms? To become mature? Can language evolution be likened in any substantial way to the cycle of life in living organisms of which maturity is a stage? These questions are important in buying back the place of new Englishes from the general conception that they are *infant varieties* that are growing towards the perfection reflected in the parent varieties.

Anchimbe (*ibid.*) is thus addressing questions of language and authority, in addition to the socio-political bias in existing theories of maturation. In place of a maturation model, Anchimbe (2009) adopts the perspective that regional varieties of English evolve around facts of their linguistic ecology. An ecological model (*cf.* Mufwene 2001) posits a pool of linguistic features to which language users are exposed, and from which they can subsequently draw for various communication needs. Feature pools in World Englishes include grammatical, semantic, lexical, and pragmatic features of English and other languages, all of which interact with particular communication needs in an array of colonial and postcolonial environments. Anchimbe's (*ibid.*) argument is compellingly data-driven (based on observable features in a region) and decidedly non-metaphorical, non-teleological, and non-biased towards social and political aspects of World Englishes.

As I have shown, most theories of World Englishes centre primarily around sociolinguistic and socio-political attributes. Why should theories of English varieties worldwide, in abstracting from the linguistic complexities and details that each theory's proponents have forwarded over the years, shift the emphasis from the linguistic to the social and political? Why have we not seen a primarily linguistic theoretical model for 'World Englishes'? A comprehensive model for the development of English varieties worldwide that

is primarily linguistic would be a worthwhile endeavour. Such a model might identify in disparate branches of English worldwide specific trends in phonetics, morpho-syntax, lexis, semantics, pragmatics, and discourse.

One example of the potential of such an approach was in fact conducted by Schneider (2008), who computationally derived categories for English varieties worldwide based on an array of specific phonetic attributes. In this data-driven approach for categorizing World Englishes based on phonetic similarities, the following regional groups emerged:

a. Ghanaian, Ghanaian Pidgin, Nigerian Pidgin, Cameroon Pidgin, Cameroon, East African, Black South African, Indian South African, Pakistan, Singapore, Malaysia, Trinidad and Tobago, Hawaiian Creole, Fiji, Bislama, Tok Pisin

b. Australia, New Zealand, White South African, East Anglia, North of England

c. Jamaican Creole, Suriname Creoles, Australian Creoles, Aboriginal English

d. Gullah, Bahamian English

e. Wales, Orkney and Shetland, Scotland

f. Ireland, Standard American, Newfoundland, African American Vernacular, Chicano

Schneider's (2008) approach was computational, and varying granularity might result in variations in these clusters. That is, varying computational algorithms might demand greater similarity within a group of regions, and therefore result in a larger number of groups with fewer regions in each; or, alternatively, a slightly altered algorithm might require less similarity within a group of regions, and result in a smaller number of groups containing more regions in each. The study raises extremely intriguing questions: for example, why should Ghanaian Pidgin in particular cluster phonetically with Singapore English? To what extent are geography or regional heritage stronger defining factors than

language policy or *nativeness* in World English categorization? What categories might emerge from a comparable large-scale, data-driven study of grammar or even lexical semantics, and how might those linguistic features define World English categories differently from socio-political features? Schneider (2008) is an excellent step towards a phonetic, rather than sociological, model for describing English varieties worldwide.

In summary, the above theoretical models for World Englishes emerged from a need to make generalizations regarding linguistic data from English varieties worldwide, and to compare and contrast those varieties. I have argued that some of those frameworks have in fact *overgeneralized* (an issue their authors would perhaps acknowledge) and that all of those frameworks have emphasized social and political features of English varieties at the expense of linguistic features. Moving forward, I would advocate a methodological framework whereby theory construction is more closely associated with linguistic data collection and analysis. Indeed, a cyclical, scientific approach would move from data to theory and back again such that data collection directly informs theoretical development, and theoretical development directly informs, and is tested by, data collection (*cf*. Wallis *forthcoming*). Such an approach can go some way to preventing both overgeneralization and an unbalanced emphasis on sociolinguistics and socio-political features of a given variety. Applying such an approach, we might begin with data such as that in Schneider's (2008) phonetic study, and draft a theoretical framework from it; then collect more phonetic data to corroborate or refute the theory; and even compare the new theory, based on phonetics, to data on grammar, semantics, or social and political facts for the variety. Alternatively, this cyclical approach would allow us to begin with Kachru's (1985) or Scheider's (2007) framework and ask whether the given categories for World Englishes match new data on, for example, phonetics, morpho-syntax or lexical semantics. If those frameworks do not match the data, then the frameworks can be modified or improved, and then compared to additional data. In fact, that is exactly what I

do in section 9.1 and 9.2 – I compare my findings to the specific predictions of Kachru's (1985) and Scheider's (2007) theoretical frameworks.

4.2. Approaches to World Englishes

Having presented the key generalized frameworks for describing multiple varieties of English worldwide in 4.1, I would now like to discuss the approaches to World English scholarship that employ these frameworks. In the process, I identify the present study as falling within specific types of scholastic approaches.

Bolton (2006: 243) has identified seven 'main approaches to World English study' – or more accurately ten approaches, given the subdivision of one approach into four distinct sub-categories. Given the primacy of social and political factors in existing theoretical frameworks for English varieties worldwide, it is no surprise that approximately half of these approaches highlight social and political questions rather than specifically linguistic attributes:

a. The English Studies approach examines norms and usage, often in a historical context.
b. Sociolinguistic approaches to World Englishes include
    i. sociology of language
    ii. feature-based approaches
    iii. Kachruvian (or, perhaps, *postcolonial*) studies
    iv. pidgin and creole studies.
c. Applied linguistics approaches are concerned with English language pedagogy.
d. The lexicographical approach involves the creation of regional dictionaries.
e. 'The Popularizers' approach is essentially a public engagement or outreach process of publicizing information about World Englishes.

f. The 'Critical Linguists' approach applies critical theory to the study of World Englishes.

g. The 'Futurology' approach attempts to predict what will happen to English and its global varieties in the future.

(Bolton 2006: 243)

I would argue that 'feature-based' approaches (category b.ii) deserve a separate category of their own, as they need not be 'sociolinguistic', but can be strictly phonetic, grammatical, or semantic. With that in mind, Schneider's (2008) data-driven phonetic study (discussed above) is likely either an example of non-sociolinguistic 'feature-based' work or, perhaps, an English Studies approach. The present study could be described in the same way, as it investigates specific linguistic attributes or norms of particular varieties. As a study rooted in lexical semantics, the present study overlaps with category d, the lexicographical approach, but it is not ultimately lexicographical. Indeed, it is difficult to identify an approach from Bolton's list that adequately matches the present work.

Melchers and Shaw (2003: 30) also describe different types of approaches to World Englishes, but they restrict their description to political stance, distinguishing between conservative, liberal or radical. Once again, this delineation is socio-political rather than linguistic. While I concede that all research contains political bias, the present study is not conducted with a grounding in any of Melchers and Shaw's (*ibid*.) stated political stances.

Schneider (2007: 15) has also proposed a system for categorizing World English research and World English theoretical approaches. Schneider (2007) presents a grid with the *x*-axis representing degrees of generality and the *y*-axis representing attention to linguistic structure.
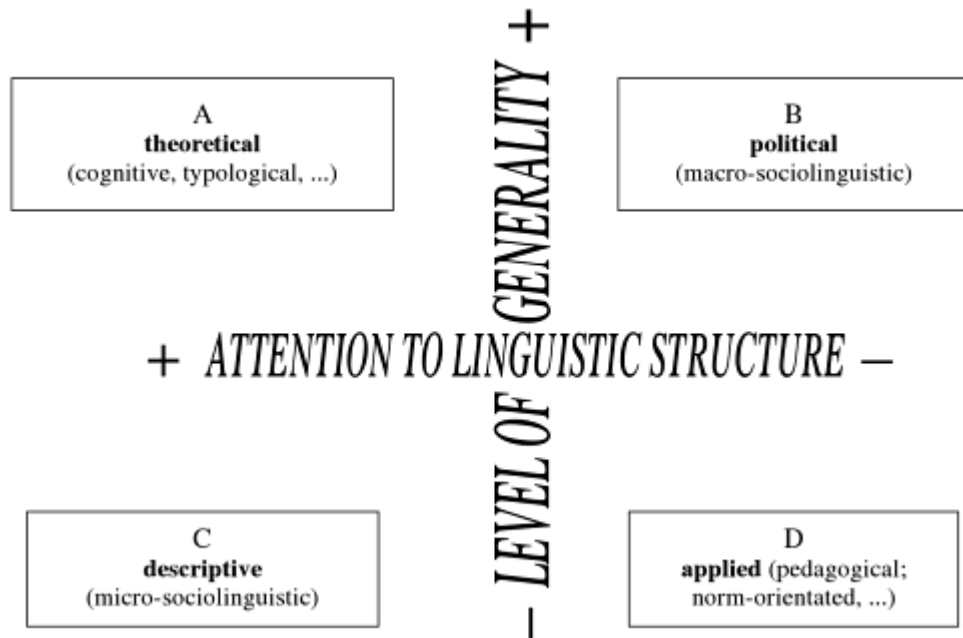
Figure 1: Schneider's (2007: 15) system for categorizing World English research

As represented in Figure 1, a theoretical approach is characterized by high generality and high attention to linguistic structure, and includes cognitive and typological work. Political work is characterized by high generality and low attention to linguistic structure. 'Descriptive, micro-sociolinguistic' research is characterized by low generality and high attention to linguistic structure. Finally, applied or pedagogical research is characterized by low generality and low attention to linguistic structure. In Schneider's terms, his own data-driven phonetic study would likely qualify as descriptive, micro-sociolinguistic. The present study would perhaps also be categorized as 'descriptive, micro-sociolinguistic' – insofar as it seeks a high level of specificity alongside focused attention on linguistic structure – but it must be noted that the present study is not primarily sociolinguistic.

Finally, Anchimbe (2009: 346-7) offers a third system of categorizing World English study:

a. The macro-approach: This approach focuses on social aspects of language, with particular emphasis on language policy, language pedagogy, literature, and the role of English in society.

b. The micro-approach: This approach investigates the details of particular linguistic features in a given region.

Anchimbe (*ibid*.: 347) observes that the bulk of World Englishes research represents the macro-approach; he also observes that the micro-approach tends to discuss features of each variety as deviations from internationally standard varieties, even if it also involves the collection of valuable linguistic data. In Anchimbe's (*ibid*.) model, the present study certainly exemplifies the micro-approach.

By identifying the socio-political slant of most existing approaches to World Englishes research (and much World Englishes research itself), I am not arguing for an avoidance of sociolinguistic questions in the field. However, I do maintain that we must critically reflect on the nature of our approaches and theoretical frameworks. It is crucial that we incorporate linguistic data into our socio-political considerations. A study like the present one (perhaps categorized as 'feature-based' or 'descriptive, micro-sociolinguistic') is part of a relatively small sub-group of World Englishes research. A balanced field of World Englishes research would include studies representing a full spectrum (or full spectra) of possible approaches. The present study therefore helps to complement and even balance the existing body of research in World Englishes.

In the following chapters, I refrain from presenting in extensive detail the histories and social contexts of Singapore, Hong Kong and Great Britain (the regions of the three varieties investigated here). I refer to those social and cultural histories at times (particularly in 5.1 and 9.2), but that is not the starting point of this study, nor a central feature of the research questions here: the primary research questions underlying the present study involve not the political history or language policy issues in a postcolonial country identifying

English as its own, but rather the facts of lexical semantic variation in samples of English in use.

4.3. World Englishes and lexical semantics

In line with existing theoretical frameworks, specific studies into the norms for English in use around the world have very often focused on social and political issues surrounding English in various locales, particularly public, governmental, corporate, or private attitudes towards English. This fact is evinced by the full body of published studies in the two leading journals in the field, *World Englishes* and *English Worldwide*. The non-socio-political research in these journals includes studies in phonetics and local pronunciation, syntax, pragmatics, and discourse. Investigations into lexis are relatively rare. This conspicuous absence is likely due in part to the absence of lexical semantics in existing frameworks for describing English variation worldwide. Because lexis as a linguistic feature has not been much discussed, new investigations into lexis in World Englishes serve a valuable purpose, either to complete existing theories, or to establish new ones.

Moag's (1982) framework and Schneider's (2007) framework both note that lexical borrowing from a local language into English is a significant element in the development of English varieties around the world. A quick glance through the *OED* online reveals thousands of words borrowed in post-colonial contexts. Examples also appear in local dictionaries and glossaries such as Yule and Burnell on Indian English (1886), Rao on Indian English (1954), Ramson on Australian English (1966), Brown on Singapore English (1989), Meyler on Sri Lankan English (2007), Kouega on Cameroon English (2007), and Cummings and Wolf on Hong Kong English (2011), and research on borrowing in English continues (*cf*. Imm 2009; Durkin 2014). This sort of lexical research has often been confined to what Lambert (2012: 305) calls 'lexical exotica', or foreign borrowings easily identified by outsiders.

However, borrowing constitutes just one portion of the lexical data in English varieties worldwide. Brutt-Griffler (2002: 153-4) argues that:

greater importance must be attached not to borrowings but to transformed meanings, a phenomenon that is much more difficult to investigate… It is in the nature of meanings to be subject to change, re-interpretation, recreation… Transformed word meaning, then, is likely to constitute a more general phenomenon than borrowing from a local language, and represents a clear manifestation of shared subjective knowledge as an agent of language change as well as constituting an overlooked question within language change.

Hymes (1996: 9) elaborates further: 'the overt forms may be familiar - the words, the attire, the buildings - but the interpretation given to them is subject to shift, to deepening, to fresh connecting up. It is in the nature of meanings to be subject to change, re-interpretation, re-creation.'[19] Moreover, significantly shifting meanings may include not just emerging novel senses for existing words, but novel preferences in style, register, and usage (*cf*. Durkin 2014). Görlach (1995 [1990]: 127) also notes that semantic distinctness of varieties of English worldwide is often a matter of 'stylistic values'.

Lambert (2012: 307) asserts that 'we should not necessarily equate surface similarity with precise overlap in meaning or usage, or both'. He goes on to ask, specifically focusing on Indian English:

[D]oes the Indian English word *dog* overlap precisely with all the meanings that word has in *standard* English? There must be a high likelihood of connotative differences, but denotative differences may of course also be possible. The case is similar for all common nouns, such as *cat*, *table*, *house*, *home*, and abstract nouns, *love*, *hate*, *affection*,

---

[19] Hymes (1996) presents his argument in the context of ethnographic methodology. Indeed, he defines ethnography as 'a cumulative coming to grips with local meanings and emergent configurations' (*ibid*.: x).

prudence, let alone verbs, adverbs, prepositions, and so on. The assumption that these are identical in meaning and usage between Indian English and Anglo-American English, while tacitly accepted, is untested, and fundamentally unknown. Additionally, there are presumably senses, words, idioms and phrases in British and American English that do not occur in Indian English.

(Lambert 2012: 307)

Lambert's awareness of the potential variation in common vocabulary is insightful. Platt *et al*. (1984: 105) make a similar argument:

Some words which may even appear at first to have 'the same meaning' in two varieties of English, *e.g.* British English and Nigerian English, such as *wife*, *dog*, *rain* may have entirely different shades of meaning for an Englishman than they have for a Nigerian. Naturally, every word may have different shades of meaning for one particular person because of past experiences, pleasant or unpleasant, but we are talking here about those shades of meaning that are based on the social and cultural background of the speakers and that may reflect basic social structures within his or her community or nation.

In a similar vein, Geeraerts *et al.* (2010: 6) have called for more studies of lexical semantic variation within individual languages, following Geeraerts's investigations of Dutch varieties (*cf*. Geeraerts 1997). Such research would include investigations into subtle variation in semantics and usage across World Englishes.

In discussing grammatical variation in World Englishes, Mukherjee and Gries (2009: 28) assert:

structural nativisation not only refers to entirely new and innovative forms and structures in individual varieties, but also covers quantitative

differences between varieties of English in the use of forms and structures that belong to the common core (*cf.* Quirk *et al.* 1985: 16) that is shared by all Englishes. Such quantitative differences in usage are not immediately accessible to intuition and can only be identified by analysing very large amounts of natural data, *i.e.* large and representative computerised corpora such as the International Corpus of English (ICE).

Mukherjee and Gries (*ibid.*) are pointing to exactly the sort of subtle fluctuations that this study investigates (though the present study is not primarily grammatical), and their methodological point on the necessity of language corpora for investigating such subtle fluctuations is essential (see 2.1).

Systematically approaching lexical variation in World Englishes, Melchers and Shaw (2003; *cf.* Gramley 2001: 50) identify three key categories of variants:

a. Localisms – words whose form and meaning are unique to a particular locale
b. Heteronyms – local words for generally available concepts
c. Tautonyms – words with the same form but different meanings in different varieties of World Englishes

The present study examines the possibility that high frequency English verbs may be tautonymous in Singapore English, Hong Kong English and British English. The method used here also allows for the discovery of heteronyms (see Chapter 5 for a description of methods used in the present study). As Brutt-Griffler (2002: 154; quoted above) observes, semantic variation is more difficult to investigate than the existence or non-existence of borrowed vocabulary; this difficulty is likely partly to blame for the shortage of research into lexical semantic variation in World Englishes. It is hoped that the present methods can serve as a model for future work in this area.

Following Melcher and Shaw's categories for lexical semantic variation in World Englishes (2003; discussed above), it is no surprise that localisms (or simply borrowings) are discussed more than heteronyms or tautonyms in recent studies. For example, Kim and Sato (2013) examine the Malay borrowing *kena* in Singapore English, which replaces *be* in standard English passives (*e.g. He kena scolded by his mother* rather than *He was scolded by his mother*). This local borrowing has also been reviewed by others (*cf.* Bao and Wee 1999). Local borrowed discourse particles in Singapore English have been a popular area of study: Smakman and Wagenaar (2013) provide a useful recent review of that research; Leimgruber (2013) adopts an indexicality approach to such discourse particles; Kim and Wee (2009) investigate the discourse particle *hor*; Wong (2005) investigates the discourse particle *one*; and Wong (2004) takes a cultural approach to discourse particles. Grant (2012) has looked at Maori borrowings in New Zealand English. Hashim and Leitner (2011) investigate a large number of borrowings in Malaysian English. Kouega (2006) lists and explains borrowings in Cameroon English. Schmied (2004) more or less automatically retrieves localisms from the East Africa component of the International Corpus of English. In all of these studies, the identification of such localisms, whether manual or automatic, is certainly much simpler than a manual analysis of tautonyms. Localisms qualify as 'lexical exotica' that demand attention, while tautonyms are much more subtle and difficult to identify.

A limited body of literature has reported on new, unique and marked meanings for relatively uncommon lexical items; such research can often be seen as comparable to the instantly recognizable differences from standard English reported by early colonial officials (see above). Some of these differences are explicable by semantic influence from local languages. For example, Platt *et al*. (1984) list numerous instances of extended meanings in English lexis around the world, including *guest* indicating 'stranger' in Nigerian English, ostensibly derived from the meaning of equivalent lexemes in Yoruba and/or Igbo. Chisanga and Alu (1997: 94-5) provide examples of semantic

extension in Zambia and South Africa, in which specific English words have taken on a new meaning due to the influence of local languages, societies and cultures, including *damage* for 'impregnate' and *ripe* describing 'a young woman ready for marriage'.

The OED also describes semantic variation, including contemporary information on English worldwide.[20] Evidence for variation between British and American English pervades the OED, but findings on other regional varieties appear as well, alongside historical developments. For example, the OED offers quotations for *accomplish* with the meaning 'to make complete or perfect; to fit out or equip' from 1524 to 1992, but the editors have noted that the sense is 'now chiefly *Indian English*'. It also presents the construction *long time* with the meaning 'for a long time', with examples going back to EModE, and describes it as 'now exc. *poet.* and in Jamaican English'. Another entry, for *timeous*, shows evidence from EModE and states that the term is only current in English in East and South Africa, and in Ireland. There is a danger in assuming, based on such examples, that some kind of colonial lag is occurring in the given varieties, but innovations are presented in the OED as well. For example, in Indian, Canadian and American English, *academics* in the plural can refer to 'reading, thinking, and study as opposed to practical or technical work', and in Indian English, *tribal* can be used as a noun to refer to a member of a tribal community.

The present study is unlike the studies of localisms or of relatively uncommon words discussed above. Instead, the present study examines very subtle variation in lexical semantics and usage preferences of high-frequency, highly polysemic verbs in Singapore, Hong Kong and Great Britain. Relatively little research has been conducted into this kind of subtle semantic variation in high frequency lexis across World Englishes. Below, I briefly review some of the studies on lexical semantic variation, followed by studies on variation in

---

[20] Though Görlach (1995) criticizes the second edition of the *OED* in its neglect of 'New Englishes', the third edition seems to be much improved.

usage preferences. In addition to the brief summaries here, I critique some of these studies methodologically in section 2.2 and 2.3.

Most studies on lexical semantic variation are initiated based on linguists' conscious observations that particular words are used in unique ways in particular locales. For example, Fuchs *et al.* (2013) examine the pragmatics and semantics of *just* and *even* in Nigerian English and find a range of unattested meanings for each. Such meanings include *even* with an affirmative, focusing sense like British English *really* or *actually*; and a particularizing sense like British English *actually*. Lange (2007) examines the semantics of *itself* in Indian English and finds evidence for *itself* as a focus marker similar to British English *even*. Jeffery and van Rooy (1994) have reported on South African usage of 'emphasizer' *now*, in which *now* functions similarly to the British English emphasizer *really*. These studies are very much in line with the aims of the present study, and they provide a precedent and a grounds for hypothesizing that *make*, *take*, and *give* may vary from one region to another as well.

Perhaps even more similar to the present study is existing research on *got* in Singapore English (Platt and Weber 1980; Brown 1992; Lee *et al*. 2009; Hiramoto and Sato 2012). *Got* is a high-frequency, highly polysemic, transitive verb similar in many ways to the verbs studied here. Lee *et al*. (2009: 295-300) investigate the many unique meanings of *got* in Singapore English, which I summarize in the following examples:[21]

> 30. Possessive – *I got two brothers, one sister.*
>> 'I have/ have got two brothers, one sister.'
> 31. Existential – *Got two pictures on the wall*.
>> 'There are two pictures on the wall'

---

[21] Lee *et al*.'s (2009) 'locative' category in fact seems to be a locative existential, identical to the 'existential' category with the simple addition of a locative word. In addition, it may be that their 'experiential' sense is in fact a grammatical marker of perfect aspect.

32. Locative – *Here got many nice houses.*

    'There are many nice houses here.'

33. Definite Futurity – *I got go Japan next time.*

    'I will definitely go/ have definite plans to go to Japan in the future.'

34. Regular Aspect – *You got play tennis?*

    'Do you play tennis regularly?'

35. Experiential – *I got go Japan.*

    'I have been to Japan.'

36. Emphatic – A: *You never sweep the floor, eh?* B: *I got sweep!*

    A: 'You never swept the floor, eh?' B: I *did* sweep the floor!'

It is clear that *got* carries semantic meaning in Singapore English which differs from other varieties. The catalyst for Lee *et al.*'s (*ibid.*) investigation was linguists' overt, casual recognition of this variation in *got* in everyday discourse; that is, this variation in *got* usage likely qualifies as 'lexical exotica' rather than more subtle, under-the-radar variation. Nonetheless, such semantic variation in a high-frequency, highly polysemic transitive verb provides solid ground for investigating possible semantic variation in other high-frequency, highly polysemic transitive verbs, such as the ones studied here.[22]

      Other studies have examined usage preferences, including for high-frequency verbs, and found varying, subtle preferences in terms of usage, style, and register for common lexis in English varieties. Lee and Ziegeler (2006) found different preferences in causative verbs between Singapore English and

---

[22] Lee *et al.*'s (2009) main aim is to explain the origins of these novel meanings, and they provide a thorough explication of various language contact circumstances that may have given rise to Singapore English usages of *got*. Lee *et al.*'s arguments were more recently revisited and challenged by Hiramoto and Sato (2012). Explaining semantic variation in terms of language contact is beyond the scope of the present work.

British English. In addition, Lee and Ziegeler (*ibid*.) also note that in Singapore English the construction *ask someone to do something* is used with the same meaning as British English *get someone to do something*. Their study, like the present work, therefore encompasses variation in both lexical semantics and usage preferences. De Klerk (2005) found different usage preferences for intensifying adverbs between Xhosa English and New Zealand English, including a narrower range of commonly employed intensifying adverbs in Xhosa English. Haase (1994; see 2.3) briefly surveyed usage of motion verbs and their alternates in East African English. Hundt (2009) investigated usage preferences for *get*-passives and *be*-passives in Singapore English and Philippines English, and concluded that both varieties exhibit strikingly little variation between spoken and written forms; in addition, the spoken and written trends in Singapore and the Philippines do not appear to conform to the norms of British English. Besides the above studies on usage preferences, a particular subset of literature has focused on variation in modal verb usage preferences (Owusu-Ansah 1994; Lee and Collins 2004; Collins 2005, 2009). All of these studies also serve as precedents for studying usage preferences related to *make*, *take*, and *give* in the present study.

One important addition to work on lexical semantics in World Englishes is Stephan Gramley's (2001) textbook *The vocabulary of World English*. While it does not constitute new academic research in the field, it does represent an exciting pedagogical approach which combines the study of lexis and World Englishes. The book addresses lexicology, lexical semantics, and word formation across the breadth of English varieties worldwide, and considers sources of English vocabulary from Old English and Old Norse through colonial contact situations and contemporary *lingua franca* environments. Case studies look in detail at specific vocabulary in particular varieties. The textbook also includes a handful of intriguing and accessible reports on tautonyms, including *bun*, *cookie*, *biscuit* and *pie* in British and American English (ibid: 60-66). The textbook could encourage students to investigate lexis and lexical semantics in World Englishes.

In this section, I have argued that lexical semantic research constitutes only a small portion of research in World Englishes. I have also shown the theoretical potential for previously unobserved variation in lexical semantics in World Englishes, and I have in turn demonstrated the need for more lexical semantic research in World Englishes. The present study addresses that need.

SECTION 2: APPLICATION AND EXPERIMENTATION

5. Methods

5.1. Regions

For the present study, three regions (subcorpora within ICE) were chosen to represent varying World Englishes contexts: British English is a valuable tool for comparison because of its historical exo-normative force (*cf.* Ziegeler and Lee 2006) and its status as a current international standard; Singapore English is considered a highly developed, post-colonial, endo-normative variety (*cf.* Schneider 2007); while Hong Kong English has been labelled a non-native, exo-normative variety (*cf.* Bolt and Bolton 1996; Schneider 2007). Table 1 shows the categories assigned to each of the three varieties studied here according to Kachru's (1985) and Schneider's (2007) widely accepted models (see 4.1). It is evident that the models overlap in some ways, but not in others.

	Kachru's Circles	Kachru's Norms	Schneider's Phases	Schneider's Norms
Singapore English	Outer Circle	Endonormative/ Exonormative[23]	Phase 4	Endonormative
Hong Kong English	Outer Circle	Exonormative	Phase 2/3	Exonormative
British English	Inner Circle	Endonormative	N/A	Endonormative

Table 1: Categorization of English varieties in Singapore, Hong Kong, and Great Britain, according to Kachru (1985) and Schneider (2007)

The present experimental design facilitates comparisons across multiple categories of World Englishes as established in two different theoretical models, in order to test whether these categories correlate with actual variation in lexical semantic usage. For instance, in addition to answering the specific questions of how Singapore, Hong Kong, and Great Britain may be similar to or different from each other in terms of lexical semantic usage, this approach also addresses the broader theoretical question of whether inner circle and outer circle varieties differ consistently from each other in terms of lexical semantics; or, alternatively, whether putatively endonormative and exonormative varieties differ consistently from each other in terms of lexical semantic usage.

Mukherjee and Gries (2009) selected ICE-GB, ICE-India, ICE-Singapore, and ICE-Hong Kong for their study of verb collocation patterns, and their choice sets a useful precedent. It is beneficial to analyse their argument.

[23] In 1985, Kachru (1985) suggested that Singapore is a variety which shows conflict between linguistic norms and linguistic behaviour, and is therefore generating internal norms, but is not yet endonormative. Much has changed since 1985, and it may be that Kachru would now argue for Singapore as an endonormative variety. Indeed, later Kachru (1992) recognized that the role of English in outer-circle varieties changes over time, and Singapore English might be said to have changed from an exonormative to an endonormative variety since then (see 7.3).

Mukherjee and Gries (2009: 27) argue that 'the three varieties [in India, Singapore, and Hong Kong, respectively] represent markedly different stages in the process of the evolution of New Englishes with British English as the historical input variety'. I would point out that British English was the 'historical input variety' for Hong Kong, Singapore, and India only in a very rough sense. *Input variety* must be defined more clearly than by Mukherjee and Gries (*ibid.*): they do not specify whether they are interested in the dominant input, for example, during the first generation of English in these three regions, during the first century, or perhaps for the entire history of each region. Regardless, their conflation of ICE-GB with *British English* and more specifically with *British English as a historical input* is problematic; ICE-GB represents British English in the 1990s, and the reason for identifying that linguistic data as representative of the 'historical input variety' for Hong Kong, Singapore, and India is not explained. It is unclear, for example, how the private conversations that constitute the largest single section of ICE-GB can be seen to establish linguistic norms that Hong Kong speakers would attempt to follow in their own private conversations. Nor do the authors indicate what qualifies as input: whether the language of British English as a whole is the input, or whether face-to-face contact, letters, films, books, television, or electronic media play different roles as input. All of those forms of input can be seen to play a role today, as in the 1990s when the corpora were sampled, and each of those inputs can originate in more than one English-speaking country, rather than just Great Britain. I believe that it is more reasonable to present ICE-GB as representing a unique variety within Schneider's (2007) and Kachru's (1985) models of World Englishes. Mukherjee and Gries (2009) include Figure 2 in their study to illustrate the diachronic progression of each of their four selected varieties according to Schneider's (2007) model.[24]

---

[24] It is worth considering again the discussion of frameworks of World Englishes in 2.1.1, particularly Anchimbe's (2009) suggestion that we forego suggestions of a maturation process
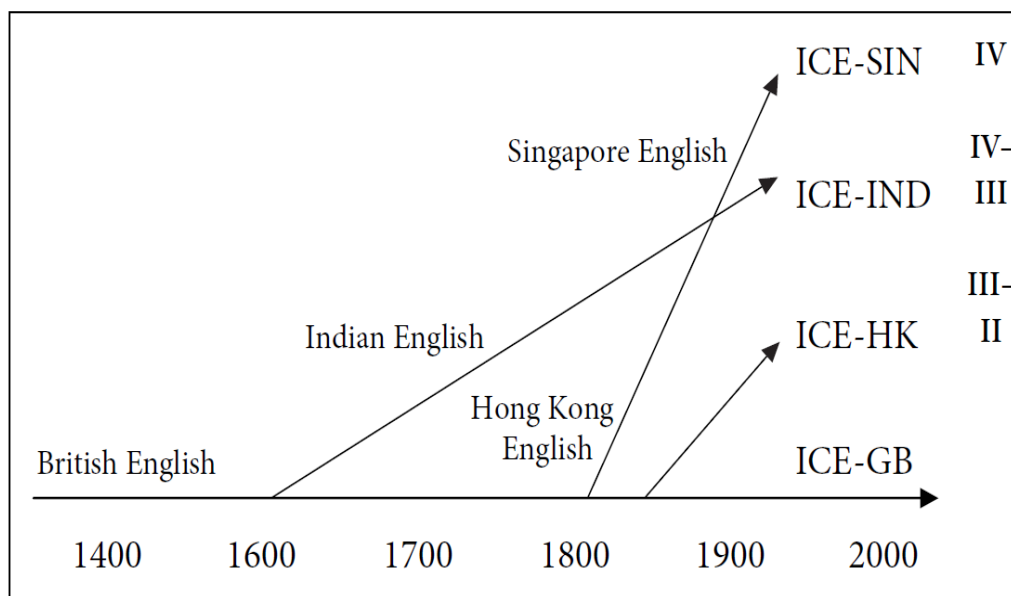
Figure 2: Comparing evolutionary stages of different Asian Englishes with ICE Corpora (Mukherjee and Gries 2009). Dates are shown at the bottom of the figure on the x-axis, and evolutionary stages according to Schneider (2007) are shown on the y-axis to the right.

Mukherjee and Gries (2009: 30) acknowledge that 'the evolutionary stages represent idealized states, and there may be considerable overlap between subsequent phases', so it is fair to accept that their straight arrows are generalizations of diachronic paths that might actually fluctuate up and down, diverge, and converge. The larger problem with the diagram is that Mukherjee and Gries (*ibid*.: 36) 'assume that the more advanced a New English variety is in its evolution, the more dissimilar it is to present-day British English'. That assumption is reflected in the arrow of 'British English' at the bottom of the diagram, from which the other arrows diverge. However, that assumption is not apparent in Schneider's (2007) theory, and is extremely problematic. An exo-normative variety can definitively be expected to follow the norms of an exo-

in the history of World Englishes. Schneider's (2007) model, while lately dominant in the field, is by no means the only theoretical framework available.

normative force such as British English, but an endo-normative variety may very well either differ from British English, or indeed resemble British English. The converse is also true: a difference between British English and another variety indicates lack of effective influence from British English in the given variety, but a similarity between British English and another variety does not necessarily indicate a direct influence from British English. Two varieties may include identical features for very different reasons, and with different historical causes.

The present study, then, investigates ICE-GB because British English occupies a unique place in theories of World Englishes, and it investigates ICE-SIN and ICE-HK because both Kachru (1985) and Schneider (2007) have hypothesized that Singapore and Hong Kong represent different categories of World Englishes. The present study focuses on only three varieties, in part because, unlike Mukherjee and Gries's (2009) study, which sampled only a small subset of the four ICE corpora and used automated methods, the present analysis was performed manually over every instance of each verb in question in each corpus; three varieties therefore represent a reasonable scope for a study of this size. In addition, Hong Kong and Singapore share a close geographical proximity, and both count Sinitic varieties as their main contact languages; that commonality can be seen as representing some added robustness in the experimental design, insofar as the additional variables of radically different contact languages and geographies have been minimized.

Of course, Hong Kong and Singapore do in fact represent very different linguistic landscapes in many ways. I have already stated that the present study is decidedly driven by corpus data of lexical semantics, rather than social and cultural facts in each region. To be sure, linguistic data and social and cultural data are complementary aspects of a larger picture, but a linguist must often decide whether to begin with the former or the latter. The present study begins with the former. I discuss each region a bit more in relation to its respective ICE corpus in 5.2, and in relation to the present findings in 9.2; in the remainder of this section, I would like to briefly cite some important work on

Singapore English and Hong Kong English. Research on British English is much more extensive, and work on the history and current state of British English is generally well known.

Leimgruber (2013), Deterding (2007), and Richards (1982) have written independent volumes devoted to Singapore English as a whole. Tickoo (1996) offers a shorter overview. Ansaldo (2010: 504) has written on the diverse and innovative linguistic situation of Singapore before and after the arrival of English. Many writers have engaged with *Singlish*, or Colloquial Singapore English, in various ways (Deterding 2007: 6; Ansaldo 2010: 499; Ling 2010: 232). Lim (2010) and Wee (2010) have written important pieces on socio-cultural linguistic variation within Singapore English. Ling (2010) has compared Singapore English to Malaysian English. Wee (2004) surveys Singapore English grammar. Singapore English is also discussed extensively by Schneider (2007) within his larger framework of World Englishes. For additional lexical studies of Singlish and Singapore English, see 2.2.

Bolton (2002) has edited perhaps the most important volume on Hong Kong English, devoted to the 'autonomy and creativity' of the variety, including an unusual argument by Vittachi (2002) for *Yinglish* as a local English/Cantonese contact variety comparable in some ways to Singlish, but grounded more in Cantonese than in English. Cummings and Wolf (2011) have produced a dictionary of Hong Kong English. Bobda (2003) compares Hong Kong English to Cameroon English as 'a tale of two extremes'. Schneider (2007) also discusses Hong Kong English extensively within his framework for World Englishes. For additional lexical studies of Hong Kong English, see 2.2.

5.2. *Corpora*

The present study investigates language in use in three corpora: ICE-GB, ICE-HK, and ICE-Singapore. A corpus is sampled to represent language in use in a particular context, and the corpus linguist must be aware of the details of the sample so as not to make overly general or overly specific conclusions. For example, ICE-Singapore is not equivalent to *Singapore English*. The actual,

non-observable entirety of language in a region is a statistical population from which a sample corpus is drawn. What exactly the sample contains and how that sample represents the population is something the analyst must bear in mind – that question requires a conscious and deliberate approach via the particular research questions and experimental design of any study.

The ICE corpora are parallel collections of sampled spoken and written English from each of an array of countries and regions (Greenbaum 1996: 3-5). Each million-word corpus contains 500 texts of approximately 2,000 words each, collected since 1990, and the texts represent a diverse array of highly specific categories (The ICE Project 2009). Table 2 displays full details of text types in the corpus.

Spoken (300)	Dialogues (180)	Private (100)	Face-to-face conversations (80) Phone calls (10)
		Public (80)	Classroom Lessons (20) Broadcast Discussions (20) Broadcast Interviews (10) Parliamentary Debates (10) Legal cross-examinations (10) Business Transactions (10)
	Monologues (120)	Unscripted (70)	Spontaneous commentaries (20) Unscripted Speeches (30) Demonstrations (10) Legal Presentations (10)
		Scripted (50)	Broadcast News (20) Broadcast Talks (20) Non-broadcast Talks (10)
Written (200)	Non-printed (50)	Student writing (20)	Student essays (10) Exam scripts (10)
		Letters (30)	Social letters (15) Business letters (15)
	Printed (150)	Academic writing (40)	Humanities (10) Social Sciences (10) Natural Sciences (10) Technology (10)
		Popular writing (40)	Humanities (10) Social Sciences (10) Natural Sciences (10) Technology (10)
		Reportage (20)	Press news reports (20)
		Instructional writing (20)	Administrative Writing (10) Skills/hobbies (10)
		Persuasive writing (10)	Press editorials (10)
		Creative writing (20)	Novels and short stories (20)

Table 2: Text types in the International Corpus of English. Numbers in parentheses indicate number of 2,000-word texts of each type in the corpus

The ICE corpora thus allow for precise comparisons between regions, using corpora that are controlled in similar ways vis-à-vis language data in use in particular contexts in each region. It is, however, important to acknowledge that

Singapore English as a real body of data (of astronomical proportions and composed of instances of use of millions of idiolects, constantly performed or enacted every day) is not necessarily 60% written and 40% spoken, nor is it 10% hand-written and 30% printed. We can, however, compare features for which the corpora are controlled: for example, we can compare writing in Singapore to writing in Hong Kong and Great Britain, as long as we bear in mind that more precise variables will affect the reliability of our conclusions – that is, a generalization about written language in the ICE corpora may not hold for a corpus of written language from each region sampled in some other way.

Moving from text type to speakers and writers, the language users represented in the ICE corpora are over 18 years of age, native to the locale represented, and 'have received formal education through the medium of English to the completion of secondary school', though English-speaking public personae who do not satisfy the education criterion are also included (Greenbaum 1996: 6). The corpus is therefore not a 'learners corpus' and does not represent students learning English: language users in the corpora have used English as a primary language for most of their lives. To reiterate once again, *Singapore English* insofar as it includes all English in use in Singapore, includes learner English and the English of fluent speakers from other locales, and many other speakers who would not be included in this corpus. The speakers in each corpus are meant to represent speakers of the local variety.

That said, ICE-HK has some unique issues. The compilers of ICE-HK aimed to include only speakers of Hong Kong English who were native speakers of Cantonese as well (Bolt and Bolton 1996: 199). In contrast, some language users in ICE-SIN and all users in ICE-GB will have used English as a primary language at home as well as at school and work, in some cases to the exclusion of any other languages; in ICE-SIN, some language users will have used English for formal education only, and will have used some other language at home or work. ICE-HK is the only ICE corpus that was designed and compiled to contain exclusively native Cantonese speakers who rely on Cantonese in their home and daily lives as well. That said, portions of ICE-HK

have not been reliably sourced (Bolt and Bolton 1996). Reliable sourcing is difficult in all of the ICE corpora: for example, with contemporary reportage, it is often impossible to confirm the backgrounds of the writer as well as all of the editors who have helped finalize a text. Bolt and Bolton (1996) assert that this was particularly difficult in ICE-HK, and that many writers simply were not sourced reliably. ICE-HK has also dramatically amended the requirement that speakers have English language education, because even schools officially designated 'English language schools' in Hong Kong often officially use English text books but unofficially lecture in Cantonese (*ibid*.). ICE-HK (like ICE-Fiji and ICE-New Zealand; *cf*. Biewer *et al*. 2010) also includes speakers who have spent considerable time overseas (Bolt and Bolton 1996). In addition, the spoken portion of ICE-HK contains interlocutors who are not from Hong Kong and who do not fit any of the criteria required by the ICE corpora. In such spoken texts, at least one interlocutor fits the corpus's criteria as a native speaker of Hong Kong English, but the other interlocutor or interlocutors do not. These outside interlocutors include speakers 'from a variety of "expat" backgrounds, including the US, UK, Australia, Europe, *etc*.' (p.c. Kingsley Bolton 2012), but details on particular speakers' backgrounds were not recorded (p.c. Gerald Nelson 2013). For example, the outside interlocutor in ICE-HK text S1A-065 identifies himself as Japanese (though it is unclear whether *Japanese* indicates nationality or ethnicity). The outside interlocutors' words are included in the corpus files but tagged as non-corpus text. Given our understandings of convergent and divergent behaviour (*cf*. Bell 2006), it is reasonable to expect that speakers of Hong Kong English might vary their language use as they interact with speakers from other locales. As far as I am aware, the issues raised by these non-native speakers in the corpus have not been discussed in any of the literature related to the ICE corpora. Thus, while we can accept that the ICE corpora are, as intended, broadly comparable, we must bear in mind that ICE-HK includes unique differences and problems. I return to this issue in 9.3.

Crucially, the three corpora are not controlled for numerous other variables: gender identification, age, education, social class or socio-economic status, racial identification, and so on. Data on gender, age and education are available for ICE-GB, though ICE-GB was not sampled in order to balance those features, but no data are available on those features in ICE-HK and ICE-SIN, and so no comparison is possible.

During the course of research, I have observed some additional variables that differ between the corpora, variables which have not been discussed in the literature. For example, ICE-GB includes in its private conversations at least one doctor-patient discourse (ICE-GB S1A-051); the other two corpora seem to include only casual private conversations between peers. The degree of formality, along with numerous linguistic variables, may therefore be expected to differ to some degree between ICE-GB's private conversations and the private conversations of the other two corpora. Regardless, the corpora are not controlled for 'formality', so the category of private conversations may be expected to include many different contexts, between individuals whose familiarity with each other or whose relative roles (social, familial, or otherwise) and power differentials may vary considerably. The category of 'Private conversations' is not necessarily a category of 'informal conversations' (and, similarly, the spoken section of the corpus is not systematically 'less formal' than the written section, particularly given that the spoken section includes parliamentary debates and courtroom proceedings, and the written section includes personal letters and fiction).

ICE-Singapore contains a duplicate text, which has not been reported elsewhere (ICE-SIN W1A-013 and ICE-SIN W1B-013). In addition, ICE-SIN

contains a duplicate subtext (comprising W2D-001 #1-16 and W2D-002 #74-93). I do not count the second instance of the text and subtext in my findings.[25]

ICE-SIN contains Singlish, particularly, for example, in some of the private dialogues (*cf*. ICE-SIN S1A-091 which includes *where got* constructions). Discourse particles such as *lah* are also quite common throughout the private dialogues. However, ICE-SIN is not a corpus of Singlish, and should not be viewed as such.

None of the corpora are controlled for topic or content. I have already argued that raw counts of words are a blunt tool that does not distinguish between usage preferences and text type or content, and I have described the case of HK W2F-005 as an example of this problem (see 2.2). That text is a fictional account of a baker 'making' bread, tarts, pastries, and so on, and it includes an unusually high number of instances of concrete *make*. Similarly, the ICE-HK selection of student essays includes a handful of essays on topics that explicitly include *make*. Rather remarkably, ICE-HK W1A-004 is an essay on the topic 'making decisions'; ICE-HK W1A-010 is an essay on the topic 'law-making'; ICE-HK W1A-016 is on the topic 'making bread'; and ICE-HK W1A-020 is on the topic 'making inferences'. Similarly, ICE-SIN S2B-040 is about 'match-making'. The raw numbers of *make* and its complements in each of these constructions reflects the topic under discussion. ICE-SIN, as a result of text S2B-040 contains an extremely high number of instances of *match-making*, much higher than ICE-HK or ICE-GB, but this is by no means indicative of a linguistic norm. This is an excellent example of the type of problem caused by a *pmw* baseline. The onomasiological approach circumvents major problems with these changes of topic, and a close reading of each corpus

---

[25] I have reported this duplication to the researchers who created ICE-Singapore (p.c. Gerald Nelson 2013), though the corpus has not, to my knowledge, been revised to correct the errors. The errors have certainly been included in numerous findings and conclusions in published studies based on the corpus, but the effect of the repeated text and subtext on any published findings is unknown, and quite possibly negligible.

and its contents (rather than an automated summation of features) also helps to identify these issues, as shown here.

5.3. Words and meanings

The present study investigates the lemmas *make*, *take*, and *give* in corpora representing three varieties of English. The words have been presented here beginning with *make*, which, of the three words, occurs the most times in each corpus; followed by *take*, which has the second most instances in each corpus; and finishing with *give*, which has the least instances in each corpus. Why are these three words worth investigating? First, they tend to be common words in corpora, which allows for a large amount of data for analysis. Second, they are considered highly polysemic, which allows for a great deal of semantic nuance and varying detail across the corpora. In those regards, they meet the call by Lambert (2012) and Platt *et al*. (1984) for studies into subtle, under-the-radar semantic variation in common lexis. Certainly, additional high-frequency, polysemic verbs such as *see* or *think*, and high-frequency, polysemic nouns such as *way* also warrant investigation, and future studies should be conducted on those words. The three words selected for the present study are particularly useful in that they can be used to engage with and respond to existing research into the frequencies and semantics of verbs with both concrete and light or delexical senses and uses, an important body of research that I believe can be complemented by the present onomasiological approach. Gilquin (2008) and Werner and Mukherjee (2012) have investigated *give* and *take* due to their high polysemy, and as representatives of verbs with both light and concrete senses.

In approaching these three words, a working hypothesis on the lexical semantics of each verb was initially formed after consulting the OED and the Collins COBUILD Dictionary as heuristics and comparators, and studying existing academic literature on the semantics of each verb. The present study is not an investigation of dictionaries or lexicographical practices, but as a reflective tool for formulating a hypothesis, dictionaries can be useful for the corpus semanticist, insofar as dictionaries convey the sense divisions

established (using a variety of methods) for the words in question, by other linguists. The OED is a thoroughly researched academic standard, and is therefore an appropriate comparator, and Collins is designed based on corpus research and usage frequencies (apparently, but not explicitly, *pmw*), and differs substantively from the OED, rendering it a useful additional comparator.[26] There is no reason to expect that corpus data representing Singapore, Hong Kong, or indeed Great Britain should conform to the conclusions of the OED. Data for Great Britain might be expected to affirm the conclusions of the COBUILD dictionary, which is corpus-based, but the COBUILD corpus is not designed to be balanced with ICE-GB, and the other two varieties might be expected to vary significantly in their regional norms – that potential variation is at the heart of the present study. Nonetheless, engaging with existing literature, lexicographical or otherwise, constitutes a legitimate (and indeed necessary) foundation for academic inquiry. On one hand, this portion of the work can be seen as *hypothesis formation*, which is a sort of inroad to the data-theory cycle discussed in 4.3. On the other hand, it can also be seen as Leech's (1974) 'prescientific' stage, consisting of reflections on existing discourse in a relatively subjective way, discussed in 3.1. The hypothesis allows for specific questions to be posed, such that the data can be analysed to find whether it supports the hypothesis or not. In the first lexical semantic analysis, performed on *make* in Chapter 6, I consider an alternative hypothesis, that of monosemy, as well as some additional issues in lexical semantic analysis. After discussing those issues thoroughly in Chapter 6, I refer back to them in subsequent chapters. The analysis of corpus data, which follows each statement of a hypothesis, moves from the prescientific to the scientific and minimizes the role of subjective reflection as much as possible; that investigation is the heart of the present work. Nonetheless, the use of a hypothesis in approaching corpus data

---

[26] Comparing definitions for the three verbs in the present study in additional dictionaries would be an interesting undertaking, but given the focus of the present work – on corpora rather than on lexicography – that undertaking is not feasible here.

(even one generated in a reflective, prescientific manner) is a necessary foundation for deductive reasoning in the scientific method. In the words of Gibbs and Matlock (1999: 263), 'linguists' intuitions can play an important role in forming testable hypotheses about linguistic structure and behavior'. For each word, I first discuss the sense divisions proposed by the dictionaries and academic studies. All of that exploration of the meanings of the given words contributes to the formation of a hypothesis for the words' meanings, which can be tested against the corpora.

Finally, I analyse the words in each corpus by manually cataloguing a variety of features for each verb, including, most importantly, its grammatical complements and the sense it represents. I use the corpus evidence to corroborate or refute my hypothesis on each word's polysemy.

After identifying and verifying the multiple senses of each word in use, I then go on to enumerate onomasiological alternates. To do that, I begin by listing all Direct Objects of each verb in question, and then searching for all other verbs in each corpus that occur with the same Direct Objects, and that convey generally the same sense. I discuss this process in more detail, along with the specific issues that arise for each verb, in chapters 6 through 8, respectively, for each verb. This identification of semantic alternates should facilitate a reasonably thorough onomasiological comparison for each sense of each verb. The basis in the corpora makes my approach particularly strong because it avoids the culturally and socially biased reflection that underlies thesauri, in favour of a list grounded directly in the evidence of actual use.

5.4. The methods summarized

Having discussed and justified the choice of data sources, the object of investigation, and the nature of the research questions, I now present in detail the methods employed in the present investigation. This stepwise presentation of the methods renders the study transparent and allows for careful reproducibility of the present experiment.

5.4.1 Hypothesis Formation

A literature review is conducted, including lexicographical representations of the three verbs in question in the OED and the COBUILD dictionaries, as well as in existing academic literature. That literature review, as an exploratory step, loosely informs the essentially prescientific, reflective establishment of hypothetical sense taxonomies to be tested in the corpora (see 5.3). The hypotheses formed for the present study do not include all senses proposed in existing literature – indeed, that would not be possible, as various studies have proposed various sense divisions that overlap or conflict with the sense divisions of other studies. Moreover, as I discuss in each instance, many existing studies have posited an unwieldly number of sense distinctions for the three verbs in question. Instead, I aim to generate hypothetical sense categories that can be consistently and reliably measured based strictly on evidence in corpus texts, and thus reproduced by other researchers. Such categories involve, for example, the concreteness or abstractness of the sense conveyed based on the concrete or abstract nature of the Direct Object; this quality is present in much of the existing literature and can be consistently measured using corpus evidence. In addition, I hypothesize the existence of a light or Delexical sense for each verb (see 3.4), in order to test the nature of that sense and to engage with existing work on light verbs. The hypothesized senses are also designed to avoid the problematic issues discussed in relation to some existing studies on each verb, and to balance generality with specificity.

5.4.2 Data annotation

Three components of the International Corpus of English, specifically ICE-Singapore, ICE-HK, and ICE-GB, constitute the data employed in the present study. The construction of those corpora throughout the past twenty-five years constituted the first stage of data collection relevant to this investigation. That data is not semantically annotated, and only ICE-GB is grammatically annotated. The first step in engaging with those corpora for the present study is therefore annotation of the corpus data.

All instances of each lemma, MAKE, TAKE, and GIVE, are identified in each corpus. For ICE-Singapore and ICE-HK, this was completed using the concordancing function in the software AntConc, based on lexical searches for each form of each lemma. Because semantic analysis is conducted manually in the present study, a lexical search using AntConc for all instances of all inflectional forms of the three verbs is an adequate first step, resulting in a list of all instances, which can be examined manually. Each instance of each verb is then catalogued in an Excel spreadsheet, along with an array of metadata. Metadata for each instance include the following:

- Location in the corpus of each instance. This is presented in standard ICE format, listing corpus (e.g. ICE-SIN); followed by text category, a hyphen, and the text number (e.g. S1A-001); followed by a hashtag and the line number (e.g. #1), for a final standard representation, e.g. ICE-SIN S1A-001 #1.

- The full line in which the instance occurs, including the words to the left and right of the verb, comprising a complete phrase, clause, or sentence.

- An indication of whether the example constitutes a derived form, for example *make* or *matchmaking* as nouns, or a phrasal verb, and a catalogue of the precise form of each. This data is not employed in the present study but is available for future use.

- An indication of which hypothesized sense the instance represents, if any, and a catalogue of instances which do not fit any hypothesized sense category. In cases where instances that do not fit a hypothesized sense category can be aggregated into a commonly occurring new sense category, those new sense categories are assessed and discussed. Most of these instances are too infrequent for further analysis, but this data is available for future use.

- A catalogue of grammatical complements for each instance of each hypothesized sense, including Direct Objects and Indirect Objects (for *make*, *take*, and *give*), predicate complements or clausal complements

(for *make*); and dative *to* constructions (for *give*). Intransitive instances are also catalogued as such.

- An indication of whether a Direct Object is concrete or abstract, or whether that distinction is ambiguous or impossible to discern in context.

Following the completion of each spreadsheet, the following steps are taken.

- Direct Objects of each verb are tallied.
- Direct Objects that occur at least three times with the given verb are identified.
- A lexical search of each corpus is conducted using AntConc for each inflectional form of each Direct Object that occurs at least three times with each verb in question.
- A new catalogue is created in Excel listing verbs that occur in the corpora with those Direct Objects, and with approximately similar meanings to the verbs in question.
- Finally, a spreadsheet is created for each instance of those near-synonymous verbs in each corpus, and a tally is created for the total number of occurrences of each near-synonymous verb.

In addition, light verb constructions are identified as follows:

- If the Direct Object has a related verb, and the meaning of the given verb construction with the Direct Object is roughly equivalent to the meaning of the related verb, then the example is catalogued as a light verb construction.
- Direct Objects of light verb constructions which occur at least three times are catalogued, and a lexical search is conducted using AntConc for each inflectional form of the related verb.

- A new spreadsheet is created listing related verbs in each corpus, and their locations in the corpus.

In addition to the features catalogued in the spreadsheets, instances are also noted separately in which each term in use exhibits attributes of classic polysemy tests, including antagonism evidence, autonomy evidence, identity evidence, and truth-condition evidence. The location of each instance of such evidence is catalogued.

5.4.3 Statistical analysis

The tallies collected in the data annotation process are then used to represent each sense of MAKE, TAKE, or GIVE as a proportion of total instances of MAKE, TAKE, or GIVE. This proportion represents a semasiological baseline, and the result is an observed probability measurement between 0 and 1. This data is presented as a bar graph in each chapter.

Next, each instance of each sense of MAKE, TAKE, or GIVE is represented as a proportion of total instances of the given sense, including near-synonyms identified in the corpus. This proportion represents an onomasiological baseline, and the result is an observed probability measurement between 0 and 1. Bar graphs are presented in the paper for these probabilities, and the raw tallies that underlie these probabilities are presented in the appendix at the end of the paper.

The probabilities are compared and analysed statistically by calculating Wilson intervals and conducting a Newcombe-Wilson test with continuity correction for each set of tallies, with $p \leq 0.05$. A Newcombe-Wilson test with continuity correction is appropriate for comparing multiple data sets, and in particular for analyzing the quantitative results of alternation studies involving the alternation of two or more features in two or more samples (Wallis *forthcoming*); Wallis (*ibid.*) has shown this test to be effective when comparing multiple corpora. Moreover, Wallis (*ibid.*) has shown that a Newcombe-Wilson test uniquely avoids the error of outputting confidence intervals that may extend

lower than 0 and higher than 1 (such errors are a potential pitfall of log-likelihood tests, for example). Newcombe-Wilson tests with continuity correction are performed using an Excel spreadsheet originally designed by Wallis (available at https://corplingstats.wordpress.com/) and further customized for the present study to measure tables larger than the two rows and two columns typical of a conventional chi-square test (customized spreadsheets available from the author). In what follows, I discuss the nature of these statistical measurements, and how they are represented visually.

The application of the Newcombe-Wilson test with continuity correction is a valid analytical choice which addresses the present research questions adequately. That said, there is an array of viable statistical approaches to analyzing the present data. Taking inspiration from Wallis (*forthcoming*), which compares the outputs of Newcombe-Wilson tests with chi-square tests, it could perhaps be useful to build on the present study by comparing the outputs of the Newcombe-Wilson tests conducted here and the outputs of linear regression analyses (with a single predictor). Such a comparison of mathematical tools is, however, beyond the scope of the present study. Speelman (2014) has shown that logistic regression is a useful tool for alternation studies involving multiple corpora, and particularly for identifying correlations between a dependent variable (such as, in the present study, the choice between MAKE and PRODUCE) and an array of independent variables (such as, in the present study, the regional corpus, or the spoken or written genre; if additional sociolinguistic or other independent variables existed in the present study, those could be addressed as well). Levshina *et al.* (2014) is a good example of logistic regression in an onomasiological study. Logistic regression is not the only multivariate approach available, but it has been shown to handle an array of independent variables effectively in this type of study. The present study addresses the independent variables of corpus (representing regional populations) and genre (written or spoken), so complex multivariate analysis is not entirely necessary. The present study manages the multiple independent variables in a mathematically simpler, if manually more engaged,

manner: the independent and dependent variables are tested using a Newcombe-Wilson test, and the data are aggregated in various ways for manual observation from different perspectives, in relation to different precise questions about the relationships between the variables. Results of many of those tests are presented in the body of the paper. This approach is, to reiterate, viable and powerful, but it is not the only possible approach. Indeed, future work might compare the present statistical outputs to those of logistic regression as well, though that essentially mathematical comparison is beyond the scope of the present work.

In the present study, error bars, or confidence intervals, are calculated in the form of Wilson intervals around a given observation, at $p \leq 0.05$ (this choice of p value is discussed further below). The error bars here indicate the probable margin of error around a given observation: for example, an observed probability of 0.6 with a confidence interval of +/-0.12 indicates with 95% confidence that the observation is reproducible with the same sample, give or take 0.12. That means that 19 out of 20 samples of the same population can be expected to yield observations of the same feature with an observed probability of 0.48 to 0.72. This statistical model presumes that if each variety were systematically resampled to create 20 new ICE corpora, no more than 1 of those new corpora (5%) would differ from the present corpus. These bounds can be understood as the probable variation within the local variety, and therefore the confidence that an observed difference between samples reflects an actual difference between populations. Confidence intervals are thus a way of visualising uncertainty in a statistically sound way (*cf*. Wallis, *forthcoming*).

When we claim that measurements of given samples like corpora are significantly different from each other, that claim is an assertion that, with an established level of confidence (and uncertainty) the differences measured in the samples reflect differences in the sampled populations. This statistical significance is a reflection of both effect size and sample size. Any measure of statistical significance incorporates those two factors. If an observed effect size (i.e. the difference between two observed probabilities) is very small, the difference between the two observed probabilities may still be statistically

significant given a large enough sample size. That is, our confidence of an actual difference between the populations may be high even if an observed effect size is small, given a large enough sample. The converse is also true. If a sample size is very small (e.g. a corpus of only a few hundred words), an observed difference between two probabilities may still be statistically significant given a large enough effect size. That is, our confidence of an actual difference between populations may be high even if the samples are very small, given a large enough effect size. With relatively large effect sizes and large samples, significant differences are very clearly discernible; with relatively small effect sizes and small samples, significant differences are generally not observable. Often, the most interesting material is the middle ground between the two extremes: those observations that fall near the boundary between significance and non-significance.

In order for two observed probabilities to be deemed significantly different, the present study establishes p as less than or equal to 0.05, in accordance with established convention. The convention is not arbitrary, but instead clearly motivated. Nonetheless, it is a convention rather than an indisputably correct choice. Various probability measurements, for various purposes, rely on various conventional p values, and some fields often employ a p value much smaller than 0.05. In linguistics, very small p values such as 0.01 or smaller often only affirm those differences which are obvious to the casual observer. Very large p values such as 0.1 render a claim for a significant difference fairly weak. The convention of p as 0.05 sets a boundary where some unpredictable and interesting differences may appear, and the convention is therefore reasonably motivated. As discussed in 2.2, however, statistical analyses are not mechanistic processes which output absolute truths, but instead tools for learning more about quantitative data and relationships within quantitative data. Observing a difference between two samples, and determining that the difference is significant at $p \leq 0.05$, is only a step in the analytical process. That observation must be critically interrogated, often in relation to the particular observations in the sample, and to particular examples

of language in use, in order for statistical significance measures to become meaningful information.

Interpreting graphs with error bars is generally quick and convenient. The difference between the bars in the bar graph indicates an effect size. To reiterate, graphs presented in the body of the present study display probabilities rather than raw tallies. So, a graph with two bars at 0.5 represents a 50/50 probability. The raw tallies, or the sample size, might each be 1, or they might each be 1,000. The size of the sample is discernible from the error bars. Wide error bars indicate a smaller sample (and hence low confidence), while narrow error bars represent a larger sample (and hence higher confidence). The specifics of the raw tallies, and the sample sizes, are presented in the appendix, and all statistical analyses can therefore be recalculated using other methods. Error bars also allow for convenient interpretation of significance at $p \leq 0.05$. If the ends of two error bars do not overlap with each other, the results are significantly different; if they overlap and the overlap includes the centre of either bar, the results are not significant; if error bars overlap, but the overlap does not include the centre of the bar, then a Newcombe-Wilson test with continuity correction is reported for disambiguation.

In interpreting graphs, it is important to reiterate that a relatively large apparent difference between the probabilities presented (e.g. a first given probability of 20% and a second given probability of 80%) may be statistically insignificant due to a small sample size. That is, we maintain a high level of uncertainty regarding the measurements due to the small sample size. Put differently, such a relatively large apparent difference in effect size between the probabilities shown in the graphs may be the result of relatively small difference in effect size between the raw tallies (e.g. two given probabilities of 20% and 80% reflecting actual raw tallies of 1 and 4 instances, respectively). Sample sizes vary greatly because the sample sizes in question are not the corpora in their entirety but the baseline of the measurement in question, e.g. the total number of instances of *give* or the total number of instances of words representing a particular sense. When viewing an apparently large effect size

(e.g. a first given probability of 20% and a second of 80%), the reader should also take note of the presented confidence intervals, or error bars. Very wide confidence intervals with a very large effect size indicate a relatively small sample size. Very narrow confidence intervals likewise indicate a relatively large sample size. In the present study, these confidence intervals are understood to represent both of those important factors: effect size and sample size. Effect size and sample size are therefore discernible from the data as presented and are rarely discussed independently. Readers can consult the appendix for the raw tallies of each measurement: those raw tallies will further affirm the actual effect sizes, as well as sample sizes, as represented by the confidence intervals in the graphs, and allow other researchers to recalculate statistical significance of those effect sizes and sample sizes using alternative measures.

5.4.4 Discussion and conclusions

Finally, the statistical analysis is examined and the findings are interpreted with respect to the semantics of *make*, *take*, and *give*; existing theories of World Englishes; theories of Cognitive Linguistics; and the value of the onomasiological approach. The research questions are reiterated and addressed in relation to the findings.

6. MAKE

6.1. Literature review and hypothesis

Of the three verbs investigated here, *make* is generally the most common in use *pmw*, as evidenced in the BNC, ICE-SIN, ICE-HK, and ICE-GB. It is a highly polysemic verb with both concrete and light senses, so it provides a good opportunity for studying existing ideas about light verbs and corpus frequencies (see 3.4). The high degree of polysemy as well as the common use across a variety of contexts and genres means that there is a great deal of data on *make*. As a result, a high degree of nuance can be applied to the semantic analysis. As described in 5.3, a working hypothesis on *make* was formulated, based on

consultation with the OED, the Collins COBUILD Dictionary, and existing academic studies on the semantics of *make*. The lexicographical work is summarized first, after which some academic studies are described. The working hypothesis is proposed at the conclusion of this section, and the corpus analysis then appears in 6.2.

6.1.1 Literature review

The OED divides *make* into 6 major senses and 75 sub-senses. Collins COBUILD lists 7 major senses. Both dictionaries clearly include a sense with a concrete Direct Object, in which the concrete Direct Object is a product or result. In the OED, this sense is listed first, as it is one of the earliest senses historically, while in Collins COBUILD, the concrete sense is sense 3, indicating that it does not occur as frequently (presumably *pmw*) in the COBUILD corpus as some other senses. The concrete Direct Object can be food, but is not necessarily food, as shown in Examples 37 and 38.

> 37. The factory **made** many things beside rope and tents…
>   (Mortimer 1962, qtd. in OED: *make*, v., 1a)
> 38. She **made** her own bread. (Collins 2012: *make*, v., 3)

In addition, both dictionaries describe a set of sub-senses in which the Direct Object of *make* is an abstract noun which represents a result of the action conveyed by *make*.

> 39. The theory that the laws were **made** or enacted by the king with
>   the consent of the lords… (Stubbs 1875, qtd. in OED: *make*, v.,
>   15a)
> 40. The police don't **make** the laws, they merely enforce them.
>   (Collins 2012: *make*, v., 3)

In other abstract constructions, there is often an implication of understanding or comprehension, as with the very common constructions *make sense* or *make a point*. With both *make sense* and *make a point*, it is debatable whether *sense* and *point* should be viewed as a result of *make*: I address that issue further in the present corpus analysis in 6.2. The OED identifies this implication of understanding as a unique sub-sense meaning 'to arrive at… as the result of calculation' (OED: *make*, v., 31a). Collins COBUILD lists *make sense* and *make a point* under *sense* and *point*, respectively, rather than under *make*, suggesting that the key word in these constructions is not *make* at all, and that *make* in these constructions does not have a discrete sense at all.

Both dictionaries also include a sense with a Direct Object and a Predicative Complement of the Direct Object, in which the Predicative Complement assigns some attribute to the Direct Object. The Predicative Complement can be a Noun Phrase or an Adjective Phrase, as shown in Examples 41 and 42. This is the second major sense in both the OED and Collins COBUILD.

> 41. The massa had told him he was only around seventeen when he won the bird. That would **make** him around fifty-six or fifty-seven now. (Haley 1977: 535, qtd. in OED: *make*, v., 33a)
> 42. She's **made** it obvious that she's appalled by me. (Collins 2012: *make*, v., 2)

Additionally, both dictionaries include a sense with a Direct Object and a Clausal Complement, indicating that the Direct Object is caused to perform the action described in the Clausal Complement, either with or without coercion,[27] as shown in Examples 43 and 44. This is sense 3 in the OED, and sense 2 in Collins COBUILD.

---

[27] The question of coercion in this construction has been discussed (*cf*. Quirk *et al*. 1985: 1205) and investigated via corpora (*cf*. Chatti 2011).

43. His slimness **made** him look tall. (Maugham 1937: 116, qtd. in OED: *make*, v., 39a)

44. You can't **make** me do anything. (Collins 2012: *make*, v., 2)

Finally, both dictionaries also include a sense with a Direct Object which has a related verb form, such that the entire predicate (*make* + Direct Object) is equivalent in meaning to the related verb form, as in examples 45 and 46.

45. His mind was incapable of **making** comparisons. (Barker 1991: 38, qtd. in OED: *make*, v.)

46. I'd just like to **make** a comment. (Collins 2012: *make*, v., 1)

In Example 45, *make a comparison* is equivalent to *compare*, and in Example 46, *make a comment* is equivalent to *comment* (v.). This is the light verb usage (Jespersen 1954: 117), or the 'Delexical sense' described by Gilquin (2008) and Mukherjee *et al*. (2012), discussed in 3.4, and it is particularly relevant for the present research questions. The OED explicitly states the criterion in sub-sense 45 that the *make* construction and the Direct Object's related verb are substitutable – a criterion which has a bearing on this study's approach moving forward. In the OED, this is sense 4 out of 7, while Collins COBUILD lists this sense first, presumably because it is the most frequent sense semasiologically.

The OED differs from the COBUILD Dictionary in numerous ways: the OED is a historical, diachronic dictionary while the COBUILD Dictionary is intended for learners; the OED presents senses in the order that they appeared historically, while the COBUILD Dictionary presents senses ordered by their frequency in use (presumably *pmw*; see 2.2 and 2.3 for more on the *pmw* analysis). Nonetheless, the similarities between the dictionaries' definitions (if not between their ordering of senses) are remarkable, and those similarities

provide a reasonably reliable starting point for establishing hypothetical sense categories for the present study.

In her approach to *make* (v.), Levin (1993) relates verb semantics to syntax in defining a broad array of verb classes based on meaning and complementation patterns. She identifies *make* as a 'Build verb' throughout her study, arguing that it alternates with verbs including *develop*, *arrange*, *assemble*, and *bake* to express 'building' in the broadest sense, often related to materials and products (*ibid*.: 56) and to creation and transformation (*ibid*.: 57, 172-3). Levin (*ibid*.) effectively connects syntax and semantics, and by gathering and synthesising the results of a huge number of previous studies on verb meaning, complementation patterns, and alternations, her work constitutes a useful reference. Semantically, her categories are much broader than the sub-senses of the lexicographical work cited above, but also clearly supported by (particularly syntactic) evidence.

Gilquin and Viberg (2009) conduct a corpus study of *make* and *do* and their translational equivalents in German, Dutch, Swedish, and French. They identify five 'main meanings' for *make*, as follows:

I.   Support verb – 'I can't make a decision yet.'
II.   Production – 'She made her own wedding dress.'
III.   Transformation – 'It was this movie which made him a star.'
IV.   Causative – VP – 'I like him because he makes me laugh.'
V.   Causative – ADJ – 'What makes you happy?'

(Gilquin and Viberg 2009: 69)

Interestingly, Gilquin and Viberg (*ibid*.) opt for the term *support verb* here, rather than *delexical construction* as in Gilquin (2008). They provide no framework, 'prescientific' or otherwise, in distinguishing one sense from another here, and the approach appears relatively reflective and subjective. Gilquin and Viberg (2009: 73) find that sense I is by far the most common *pmw*, followed in order by senses V, IV, and II (with little difference between

them). Sense III is far less common than the rest. This study is definitively semasiological and the analysis is definitively *pmw*.

Gibbs and Matlock (2001: 226-8) establish eight senses of MAKE, as follows:

    I.   Intention + Human Effort – 'Bob made a pie.'

    II.   End result of specific action – 'Bob made a dent.'

    III.   Language – 'The poet made a plea.'

    IV.   Temporary state – 'The clown made the child happy.'

    V.   Inherent attribute – 'Antonio Banderas makes a good father.'

    VI.   Necessary ingredient – 'Black beans make a good burrito.'

   VII.   Inclusion – 'A surprise ending makes a good court case.'

  VIII.   Substitution – 'A Bible makes a good paper weight.'

The authors note that their sense II here is an example of an ACTION FOR END RESULT metonymy (*ibid*.: 227). They also note that these senses 'are not entirely distinct from each other' (*ibid*.: 228), in accordance with the Cognitive Linguistic assumption of the non-discreteness of categories, but their experiment tests the psycholinguistic reality of these sense distinctions with a group of experimental subjects. More specifically, they attempt to link speakers' conceptual knowledge related to, for example, 'intention' or 'temporary states' to specific examples of *make* in use. These sense categories are, therefore, effectively topical, generally conceptual, and related to context. The legitimacy of these sense categories is part of the research question being tested in the study. I would argue that these sense distinctions relate more to contingent meaning than constant meaning, though those categories are of course called into question in Cognitive Linguistic work like Gibbs and Matlock's (*ibid*.). Senses V through VIII could perhaps be subsumed under a single sense category, in which *make* relates a Subject to a Predicate Complement, conveying that the Subject referent constitutes the Predicate

Complement referent, or is a definitive attribute of the Predicate Complement referent.

Similarly, Aubois (2008) has explored the polysemy of *make* from a cognitive perspective, and relates her definitions to contingent meaning by employing pragmatic and encyclopedic information in her description. I summarize her sense distinctions below.

I. Direct manipulation of a concrete thing, including creating a concrete thing or changing the state of a concrete thing – 'She makes her own clothes.' 'Why can't you kids make your own beds?'

II. Reification, in which 'abstract entities are reified and conceptualized as manufactured objects' (ibid: 43) – 'We must make a decision by tomorrow.'

III. Actualizing force, in which 'there is some identity in the speaker's mind between the Subject and the Object' (ibid: 45) – 'Susan will make a very good teacher.'

IV. Focus on the result, in which the Direct Object is a goal to be attained – 'We could make the city before nightfall.'

Aubois (*ibid*.: 42) asserts that the light *make* construction, exemplified in her Sense II, 'focuses on the result… and on the change of state', while its alternate related verb 'focuses on the process'. This claim is apparently reflective and subjective – no specific evidence is provided. In addition, she includes traditional light constructions as examples in each of her four categories, and her description of light constructions as focusing on a result obviously overlaps with her definition for sense category IV (an issue she does not discuss). Like Gibbs and Matlock (2001), these sense distinctions are general, topical, and subjective, and Aubois (2008) does not specify how they were determined.

Gilquin and Viberg (2009), Gibbs and Matlock (2001), and Aubois (2008) are all working in a Cognitive Linguistics (CL) framework, and it is important to note that CL has nurtured a new generation of polysemy research.

It is no coincidence that existing studies on the polysemy of *make* tend to come from CL researchers. It is also no coincidence that this research does not apparently aim to describe a word with a single semantic analysis that can claim unique integrity or objectivity. Geeraerts (2006 [1993]) has commented on the anti-objectivist nature of CL: Lakoff (1987) suggested that lexical semantics (*i.e.* word meaning) is not 'objective' but subjective, insofar as word meaning emerges from cognitive processes related to subjective, embodied experience. But Geeraerts (2006 [1993]: 138) points out that, within CL, it is not just lexical semantics (word meaning) that often appears anti-objectivist, but also the research methods applied to lexical semantic studies as well. That is, the claim that 'lexical semantics is not objective' has often been a vague claim, failing to specify whether it is word meaning itself that is not objective, or the study of word meaning that is not objective. It may be that Gilquin and Viberg (2009), Gibbs and Matlock (2001), and Aubois (2008), in their intuitive and generally non-verified and non-reproducible descriptions of the semantics of *make*, reflect this second type of anti-objectivism: they may be embracing anti-objectivist methods in addition to accepting the non-objectivity of word meaning itself.

6.1.2 Hypothesis

I propose the five sense categories for *make* below as a working hypothesis for the semantics of *make*.[28]

---

[28] In the course of the study, I also observed instances of Gibbs and Matlock's (2001) sense distinctions V through VIII (see above) and Aubois's (2008) sense III, although those sense distinctions were not included in the originally hypothesized senses. Those instances were catalogued, and are available for future investigation, but they did not occur frequently enough to make any viable generalizations about their usage preferences or semantics in any of the three corpora. Similarly, examples resembling Aubois's (2008) sense IV occur in the corpora, but too rarely for any viable generalizations or conclusions. Those senses, too, are catalogued, and that data is available for future study.

I. Produce (Concrete) - Senses in which the Direct Object referent is a concrete product or result of the action described by *make*

II. Produce (Abstract) – Senses in which the Direct Object referent is an abstract product or result of the action described by *make*

III. Delexical - Senses in which the Direct Object referent has a cognate verb, and the make construction is roughly equivalent in meaning to that cognate verb

IV. Causal - Senses with a Direct Object and a Clausal Complement, in which the Direct Object referent is caused to perform the act described in the Clausal Complement

V. Complex Transitive – Senses with a Direct Object and Predicative Complement in which the state or attribute described by the Predicative Complement is assigned to the Direct Object referent

These five sense categories also seem to strike a reasonable balance between the general and the specific, avoiding unnecessary propagation of senses but also avoiding over-generalized monosemy. Like the definitions offered by the OED, the COBUILD dictionary, and the academic studies discussed above, the senses offered here may overlap, but I have aimed for minimal overlap, as I discuss below. The concrete and abstract sense categories are effectively forwarded by both the OED and the COBUILD dictionaries, and they are reflected in some way, to a greater or lesser degree, in the academic studies described above. Proposing these categories as a working hypothesis allows the present analysis to engage with existing work, lexicographical and otherwise, and to corroborate or refute existing analyses on the semantics of *make*. I justify this starting hypothesis further, and explore counter-arguments, below.

To work from a hypothesis is not to aim for a hypothesis to be true (or even to 'believe' it to be true). A hypothesis allows a researcher to pose specific questions and then determine how the data at hand answer those questions. In this case, I began by determining whether individual instances of *make* could be categorized into these five sense distinctions, and what additional distinctions

needed to be added. In fact, many additional sense distinctions needed to be added, and the hypothetical sense distinctions could not be maintained in their entirety, as I discuss in 4.1.2. An alternative hypothesis will be considered below – that of monosemy – but it is noteworthy that lexicographical work and existing academic studies, as cited above, accept polysemy as a theoretical and descriptive approach to lexical semantics.

Senses I, II, IV, and V seem relatively straightforward, and are in fact generally easy to identify. A concrete Direct Object represents something that can be directly observed by any of the five senses. In this case, *concrete* is therefore not equivalent to *tangible*, as 'sound' or 'noise', for example, are directly perceivable by human senses but are not tangible. 'Ideas', on the other hand, are not perceivable by the five senses, and are therefore categorized as abstract. In practice, very few examples in the corpora are ambiguous between concrete and abstract – such ambiguities are discussed in their place in each chapter.

Senses II and III raise particular issues. I worked from the hypothesis that a delexical sense exists in which the Direct Object has a related verb, and the *make* construction is roughly equivalent in meaning to that related verb. This working hypothesis reflects the tradition of light verb research, and therefore allows for engagement with that tradition while working to corroborate or refute the arguments in that tradition. Senses II and III are not necessarily mutually exclusive. In the broadest sense, nouns denoting acts can be seen as a subset of abstract nouns. Indeed, Huddleston and Pullum (2002: 337) categorize nouns denoting actions as abstract. In the present study, by default, examples that could be identified as either Abstract or Delexical were considered Delexical. By defaulting to Delexical when the category is unclear, the present study might be seen as establishing a coherent sub-group of Abstract constructions which includes all Delexical constructions. This is certainly not the only valid approach to the problem, but it is a reasonable, coherent, and consistent one. I return to the nature of this approach and its implications in my analyses, particularly in 6.2.7. Then, in Chapter 9, I compare

make to *take* and *give* in this regard, and show that Delexical senses and usage are not identical for the three verbs, and that, contrary to established ideas regarding light verbs (see 3.4), there may not be a consistent Delexical usage that is common to all three verbs.

As described in 5.3, I looked for evidence in the corpus of the types of constructions represented by the traditional polysemy tests forwarded in 3.3. That is, I looked for corpus evidence of antagonism, autonomy, identity, and truth conditions. Hypothetical examples to be drawn from the corpus might include instances like the following:

47. He **made** cakes and plates to put them on.

If constructions similar to that in Example 47, with a Direct Object referring to food and another Direct Object referring to non-food, occur in natural use in the corpora, then those corpus examples may suggest that *make* does not have the two distinct senses 'produce (concrete, food)' and 'produce (concrete, non-food)'. If such constructions are more likely to occur in one variety than another, then that variation might constitute differences in semantic norms, and in the polysemy of *make* around the world. Alternatively, it may be that such examples occur as intentional zeugma – in fact, it seems that intentional zeugma does occur with *make*, as discussed in 6.2.5 and 6.2.6 – and contextual information can help to establish intentional zeugma as such. Ultimately, confirmation of intentional or non-intentional zeugma in a given variety may require input from local native speakers, perhaps via surveys or psycholinguistic experimentation.

It might be argued that the distinction between senses I and II is contingent rather than constant, as that distinction is determined by the concrete or abstract nature of the Direct Object. The test for this hypothesis will lie in examples of identity evidence in the corpus, such as the hypothetical example below:

48. [?]The state **makes** both laws and courthouses.

For many speakers, this example likely results in zeugma, and a linguist relying on his or her own reflection alone might decide that this is either acceptable or unacceptable. More important than reflective data for the present study is corpus data. In corpora, we find actual evidence for the feasibility of such a construction (even if we do not necessarily find confirmation of the intentionality of zeugma). This, then, allows for a valuable elaboration of the research questions presented here: are constructions such as that in Example 48 more likely to occur in some varieties of English than in others? Such a difference would qualify as the sort of subtle fluctuation in semantics and usage between regions that the present study is examining. However, it is necessary to bear in mind that corpora cannot prove a negative: the absence of a feature in a corpus does not mean that the feature cannot occur. There are many other possible ways that *make* can fit into the types of constructions represented by polysemy tests, and looking for examples that in some way resemble some polysemy tests has in fact proven fruitful.

As discussed 3.3, a monosemic bias is a reasonable approach to lexical semantics: it is justifiable for an analyst to avoid over-profileration of senses in favour of ontological parsimony, aiming for simplicity and elegance in describing meaning. The simplest possibility of all is complete monosemy, and it is therefore worth discussing how *make* might be seen as monosemous. This constitutes an alternative to the hypothesis presented above: the range of *make* constructions could in fact represent a single, broad sense of *make*. Is it possible that the following examples actually represent a single sense?

49. They **make** furniture.
50. They **make** laws.
51. They **make** decisions.
52. They **make** her angry.
53. They **make** her do it.

In Examples 49 through 53, the words and constructions following *make* might all be seen as conveying some kind of result of the action described by *make*: in Example 49 the result is a concrete object; in 50 an abstract thing; in 51 an act; in 52 an assigned state; and in 53 an action undertaken by a specified actor. In that case, in Example 53, it is not *her* that is the result of the action represented by *make* but *her do it*, or perhaps 'her doing it' or the proposition 'she does/did it'. One interesting argument on the grammar of *make* that might be seen as implying a sort of monosemy comes from Aarts (2011; *cf.* Aarts 2013: 84, 302-4 on *consider* with a Predicative Complement). Aarts (2011: 218-19) argues that in examples like 53, the 'NP + bare infinitive clause' (in this example, *her do it*) is the Direct Object of *make*. This is perhaps not intuitive grammatically: the NP *her* would become the Subject of a passivized version of the sentence, which is a key criterion for typical Direct Objects (*cf.* Aarts 2011; Quirk *et al*. 1985; Huddleston and Pullum 2002).

> 54. She was **made** to do it.
> 55. *Her do it was **made**.
> 56. *She do it was **made**.

Aarts (ibid: 219) explains that although the NP becomes the subject of the passive, *to* must be inserted into that passive to be grammatical,[29] as shown in Example 57.

---

[29] Aarts (2011), taking a synchronic view, argues that the canonical active clause is transformed by inserting *to* in the passive construction. That perspective is complementary with the OED's presentation of diachronic evidence, but the difference between Aarts (2011) and the OED is noteworthy. The OED suggests that the inserted *to* is in fact an artefact of the *to* that was once required in active constructions as well, rather than something additional that is inserted contemporaneously in the passive construction. The OED also observes that *to* in such constructions is still standard usage in Indian English. The present study has found that ICE-SIN and ICE-HK include numerous examples of *to* in such constructions as well.

57. She was **made** to do it.

We might likewise argue that *her angry* is the Direct Object of *make* in Example 52. Aarts (2013: 84-5, 302-6) makes a similar argument with the verb *consider*.

58. Larry considers my brother a genius. (*ibid*.: 85)

Aarts (*ibid*.: 85) argues that the Direct Object of *consider* is the string *my brother a genius*. Because Aarts's work is a textbook for students, he presents semantic evidence for this position first (a sound pedagogical approach), arguing that Larry is not *considering my brother*, but is instead *considering* the proposition that my brother is a genius; Aarts (*ibid*.: 302-6) then finds that the syntactic evidence is less conclusive, but weighs the syntactic and semantic evidence together to support his point. The precise semantics of *consider* are not explored or defined, which is quite defensible given that the work is a grammar textbook and not a semantics one. However, Aarts's (*ibid*.) semantic argument does depend on an implicit definition of *consider*. This issue is

---

But what **made** you to put those principles in such stirring terms? [ICE-SIN S1B-039 #8]
…the deliberate violation or flouting of maxim can **make** the listener to infer the conversational implicature. [ICE-HK W1A-011 #54]

The possibility of of including *to* also gives rise to syntactic ambiguity:

What we need however are those people who are able to **make** policy to represent public interest [ICE-HK S2B-040 #111]

Example iii might be glossed either as a causative construction obsolete in British English, 'cause the policy to represent public interest', or as an existing British English construction that can be glossed as 'create policy, in order to represent public interest'.

relevant to the analysis of *make* as monosemic, particularly for Example 52. Again, it does not seem obvious that *her angry* could be the Direct Object of *make*, given that *her* (*she*) is the Subject of the corresponding passive (Example 59) rather than *her angry* or *she angry* (Examples 60 and 61).

> 59. She was **made** angry.
> 60. *Her angry was **made**.
> 61. *She angry was **made**.

Maintaining the argument for *her do it* or *her angry* as Direct Objects would seem to rest to a significant degree on the semantics of *make*, and would seem to presume a clear understanding of the semantics of *make* as meaning 'cause a result'. This would perhaps be acceptable for a study that is primarily grammatical, but for a semantic study, it is precisely the meaning or meanings of *make* that are under investigation. *Her angry* can be seen as the result (and Direct Object) of *make* only if *make* means 'cause a result'. If *make* instead means 'coerce' (*cf*. Chatti 2011), then the patient must be *him* rather than the entire proposition conveyed by *him angry*: we probably cannot coerce propositions like *him angry* but we can coerce agents like *him*. If we hypothesize that *make* is monosemic, generally meaning 'cause a result', then we might in turn consider that the causative uses of *make* constitute a subset of the Abstract sense, such that the proposition or act conveyed by the clausal complement is an abstract result of *make*. According to the methods laid out in the present study, if *make* is monosemic, we can expect to find constructions like the hypothetical ones in 62 through 64:

> 62. [?]They **made** furniture and him do it.
> 63. [?]They **made** cakes and him angry.
> 64. [?]They **made** him angry and him do it.

Examples 62 through 64 contain not just questionable semantics but also certainly nonstandard grammar. Intuitively, these examples would likely be unacceptable on multiple levels. Nonetheless, the questions in the present study are: Do these constructions occur in a British corpus? Are they more or less likely to occur in a British corpus than a Singapore or Hong Kong corpus? Moreover, if they do occur regularly in, for example, a Hong Kong corpus, does this suggest monosemy for *make*, or at least unique polysemy for *make*, that is particular to English users in Hong Kong? In this case, polysemy tests have provided a research question and an approach to a corpus study that would not otherwise have been discernible.

Ultimately, then, the present research questions ask what sense distinctions can be established for *make*, and whether those sense distinctions vary from one region to another, and my working hypothesis on the sense distinctions of *make* is constructed in response to existing research. The approach is a valid one, even if other working hypotheses could have been forwarded instead. In addition to the sense distinctions hypothesized above, other senses of *make* have also been catalogued in this study. That data, while not immediately pertinent to the present onomasiological study, is available for future research. Also, in the interest of comparing *make* with both *take* and *give*, and in focusing on *make* in its light or Delexical use, the present study does not include a full analysis of Causal or Complex Transitive *make*, which involve additional grammatical complementation patterns and are therefore not comparable with *take* and *give*. However, Data on Causal and Complex Transitive *make* were collected, and are addressed briefly in 6.2.11; those data are also available for future research.

6.2. Findings

6.2.1. Make: The lemma

Having examined existing lexicographical and linguistic work, the present study proceeded to examine the 6,019 instances of *make* in the three corpora aiming to find whether semantic variation exists between the three regions.

Instances of *make* in the corpora were examined to affirm or refute the working hypotheses (see 6.1); to determine whether unique sense categories for *make* appeared in any of the corpora; and to look for particular types of evidence derived from polysemy tests (see 3.3). Analysing a feature in a corpus for the first time usually presents challenges to any theoretical framework, and the present study is no exception. Each example of *make* in the corpus was manually analysed; where possible, the example was classified within a hypothetical category, and where not possible, new categories were created. This process is not automatable, and is neither simple nor straightforward. Nonetheless, many instances fit neatly into the hypothesized sense categories.

The lemma *make*, in the forms *make*, *makes*, *making*, *made* can be said to be quite common. The ICE corpora, like the BNC and other corpora, suggest that an English listener/reader can reasonably expect to encounter *make* relatively often, with variation expected depending on text topic and genre, and, perhaps, regional variety: the raw numbers for *make* in the corpora suggest high exposure rates for these words. Those raw numbers do not account for variation in exposure rates due to variation in text topic or real-world context; these exposure rates are also an epiphenomenon of psycholinguistic selection processes, even if they do not provide detailed information on those selection processes (see 2.2 and 2.3). Total instances of *make* for each corpus appear in Table 3.

	Instances of MAKE
ICE-Singapore	1,917
ICE-HK	2,102
ICE-GB	2,000

Table 3: Instances of *make* in ICE-Singapore, ICE-HK, and ICE-GB

Semasiological data on the frequencies of each of the five hypothesized senses are presented in Figure 3 and Figure 4.[30] Error bars represent Wilson intervals with p≤0.05.[31] These data answer the question: given that a listener or reader is exposed to *make*, what is the probability of encountering each sense as a proportion of the total number of occurrences of all senses? This might be useful for a lexicographer arranging the order of entries in a dictionary like the Collins COBUILD dictionary, or a learners' dictionary designed for ease of use.

---

[30] Subsequent semasiological and onomasiological data is presented as probabilities in bar graphs. The exact numbers for each graph are presented in detail in the Appendix. See 5.4.3.

[31] See 5.4.3 on error bars.
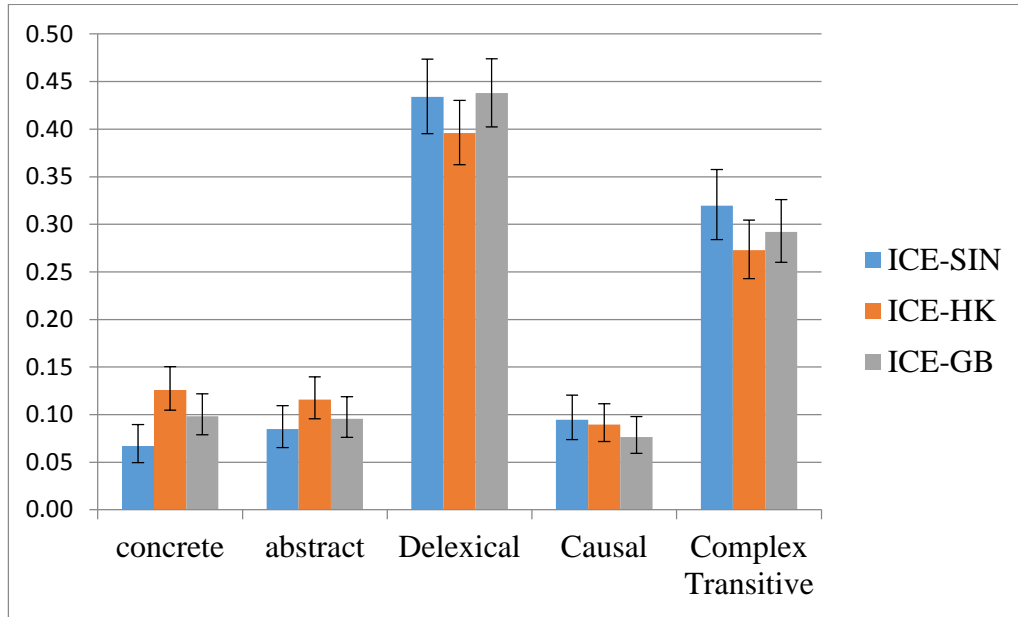
Figure 3: Probabilities of encountering *make* with each of the five hypothesized senses in the written portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
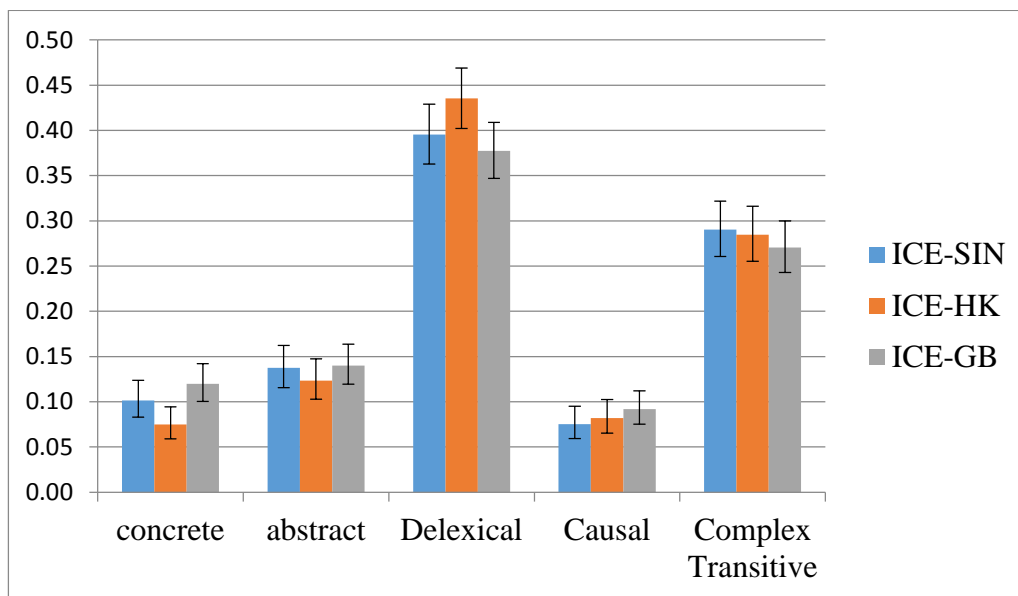
Figure 4: Probabilities of encountering *make* with each of the five hypothesized senses in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

Semasiologically, it is clear that Delexical instances of *make* are the most common. The semasiological measurement is also the only presentation of the data that aggregates Delexical occurrences of the verb – the onomasiological presentation presents individual Delexical constructions and their individual related verb alternates without aggregating the total number of all Delexical constructions. It is clear from the semasiological graphs that exposure rates to Delexical constructions are quite high. These measurements corroborate Gilquin's (2008) observation that the light sense is the most common semasiologically; moreover, these findings seem to agree with the general sense organization in the Collins COBUILD Dictionary, which lists Delexical examples first, followed by complex transitive examples. The following sections present analyses of individual senses, with examples.

6.2.2. Make: Produce (Concrete)

Many of the most clear and certain instances of *make* represent the sense
'Produce (Concrete)' (which I also refer to as 'concrete *make*'). In practice,
there are very few ambiguous cases for concrete *make*. Examples include the
following:

> 65. You don't have to **make** poster lah for your case. [ICE-SIN
>   S1A-020 #68]
> 66. Custard pie is when you **make** it with condensed milk. [ICE-
>   SIN S1A-039 #214]

Example 65 describes the production of a poster, and Example 66 describes the
production or preparation of food.

One instance in the three corpora, Example 67, drawn from ICE-HK,
combines food and non-food Direct Objects, and therefore constitutes identity
evidence between the possible senses 'Produce (Concrete, Food)' and 'Produce
(Concrete, Non-food)'.

> 67. On that day, we shall sell the handicrafts and snacks which are
>   **made** by our students. [ICE-HK W1B-017 #17]

One instance, however, is not enough evidence to make any claims about the
significance of identity between concrete (food) and concrete (non-food)
senses, either in the Hong Kong corpus or in English more broadly. It cannot,
therefore, be concluded based on this evidence whether concrete (food) and
concrete (non-food) are distinct or non-distinct senses of *make*. In fact, the one
instance might be seen as rare enough to indicate that identity between concrete
(food) and concrete (non-food) senses is not the norm. In addition, there may be
grammatical features of Example 67 that relate to the acceptability of the
coordinated Direct Object: the passive nature of the construction might be more
acceptable than an alternative active construction. That question cannot be

answered with the given data, but it serves as an illustration of all of the factors that might influence the selection process and the acceptability of a given construction.

6.2.3. Make: Produce (Abstract)

Examples of 'Produce (Abstract)' are sometimes difficult to identify. Examples 68 and 69 show sentences that can be analysed as 'Produce (Abstract)'.

68. The individuals who have **made** a difference have always understood this link and exploited it to the fullest. [ICE-SIN S2B-050 #82]
69. They must **make** an effort to build better relations and resolve differences rather than create fear. [ICE-HK W2E-007 #32]

The Direct Objects in both of the examples above are abstract nouns. In Example 68, *difference* is an abstract noun whose referent can be seen as being produced in the act described by *make*. Likewise, in Example 69, *effort* is an abstract noun whose referent can be seen as being produced in the act described by *make*. Also, importantly, *make a difference* is not equivalent, at least in this example, to *differ* or *differentiate*. Is this interpretation the one with the most integrity? Is this the most plausible interpretation? As I show below, the onomasiological portion of the study demands that we question not just this initial analysis, but also the hypothetical category 'Produce (Abstract)'; for now, these examples suffice.

6.2.4. Make: Delexical

Many cases of Delexical *make* are perfectly clear and certain.

70. All they have to do is use a small remote control to **make** their selections which will be sent via telephone line to a computerised switch. [ICE-SIN W2B-037 #21]

71. The exercise enables your body to **make** better use of the calcium you eat, but it will be more effective at building bone if you also take plenty of calcium (at least 1000 mg a day). [ICE-GB W2B-022 #45]

72. … Dr Tay Eng Soon led eight uniformed youth groups under the Education Ministry in **making** donations at the Well. [ICE-SIN S2B-016 #52]

In Example 70, *make a selection* is equivalent to *select*; in Example 71, *make use* is equivalent to *use* (v.); and in Example 72, *make donations* is equivalent to *donate*.[32] In Example 70, however, it is worth noting that *select* in standard British English must take a Direct Object, and the alternation between *make a selection* and *select* here would therefore be dubious in standard British English – or would at least require the insertion of a Direct Object which is not presented in the original example. This is a recurring issue with Delexical examples. In the present study, I accept such alternations for two reasons: first, from a strictly lexical semantic perspective (*i.e.* isolating semantics as a particular variable), *make a selection* and *select* seem to be equivalent. From a broader grammatical, or even pragmatic perspective, it would be possible to identify Delexical examples only as those examples that express a patient or Direct Object for the related verb construction. However, that approach very quickly leads to very murky territory dependent upon intuitions regarding regional pragmatic norms. It seems that the patient for the related verb can be expressed in many ways in the corpora: it can be suggested in an Adjunct or Modifier in the form of an adverb, prepositional phrase, or clause, or it can be implied pragmatically or suggested from the surrounding discourse. As a non-native speaker of any of the three varieties in question, relying on my own

---

[32] Intuitively, it may be that *make a donation* would seem to occur more commonly in British English in reference to monetary donations, whereas *donate* need not be restricted in that way. The construction does not, however, occur frequently enough in the corpora to corroborate or refute that intuition.

reflection in analysing the presence or absence of pragmatic or discursive implications of this sort would be an unsound approach. Instead, distilling the question to a sort of 'meaning slot' with an alternation seems to be the most consistent and reliable approach. Besides this first semantic reason, there are also clearly varying norms for transitivity between regions. For example, *make* often occurs intransitively in ICE-SIN.

> 73. A: Uh, beef rendang.
>
> B: Wah, no time to **make** ah. [ICE-SIN S1A-091 #219-20]

The presence of a Direct Object, and the acceptance in a region of various verbs (such as *make* or *select*) in intransitive constructions, may relate to the selection process between alternates. However, only by isolating variables one at a time can we build a complete picture of the selection process. The present study isolates the variable of lexical semantics and looks at the semantic 'slot' filled by *make* (and the other verbs in question), rather than at the syntactic requirements of the verbs.

Passivization is another potential variable in selecting Delexical *make* over an alternate. For example, it may be that the active construction *make modifications* is readily interpreted as an alternate for *modify*, while *modifications were made* is not so readily interpreted as an alternate for *modify* (or as an alternation for *X was modified*). In that case, Example 74 might be particularly difficult to categorize as either Abstract or Delexical.

> 74. Friedman (1984), for example, appears to suggest that the
> evolution of different contract strategies are fundamentally

modifications **made** in the marketing direction to suit the

clients' requirements. [ICE-SIN W2A-003 #21] [33]

In the present study, passivized Delexical constructions are analysed as Delexical constructions. Again, the particular additional issues with light verbs, including these specific grammatical questions regarding alternations, are not the focus of the present study, but these variables must be born in mind.

Additionally, some Direct Objects in Delexical *make* constructions may relate to obsolete verb forms, as in Example 75.

75. The superficial trabeculectomy scleral Qap was dissected and the partial thickness cataract incision was **made**. [ICE-SIN W2A-026#101]

Incise is obsolete according to the OED. However, as with the inserted *to* in causative *make* constructions (see 4.1), it is possible that a linguistic element which is obsolete in British English may not be obsolete in other varieties. In the present corpus-based study, if the related verb is obsolete in British English according to the OED, and also not evidenced in any of the corpora, I consider it obsolete in general, and therefore categorize the example as Abstract, with no existing related verb, rather than Delexical.

Delexical *make* may overlap with Delexical *take*. There is one interesting instance of such overlap in the corpora, shown in Example 76.

76. The staff member shall not: 1) Take or permit to be **made** any alterations in the internal construction or arrangements or in the

---

[33] I follow Huddleston and Pullum's (2002) analysis and consider the *–ed* participle after the noun *modifications* in *modifications made* a bare passive, such that this qualifies as a passive *make* construction.

external appearance or in the present scheme of decoration of the premises. [ICE-SIN W2D-003 #130-1]

Example 76 conjoins *make* and *take* with a single Direct Object, *alteration*. In this case, *to make an alteration* is *to alter*; and, apparently, *to take an alteration* is *to alter*. Alternative expressions might have included the following:

77. ...**take** or permit to be **taken** any alterations…
78. …**make** or permit to be **made** any alterations…

The conjunction of *take* and *make* in this case may indicate some degree of identity between Delexical *make* and *take*, though this evidence is extremely limited and is not at all conclusive.

There is limited evidence for *make a difference* in ICE-SIN and ICE-HK as a Delexical construction with the meaning 'differentiate', rather than as an abstract construction with the meaning 'produce a difference'. The OED lists this sense as well (OED, *difference*, n., 5), but it is not evidenced in ICE-GB.

79. A difference must be **made** between simultaneous and successive bilingualism. [ICE-SIN W1A-011 #91]
80. A: But is it [the term *Mandarin*] just commonest or is it, uh, any difference in the interpretation [between the terms *Mandarin* and *gwok yuh*]? …
B: I can **make** a difference, but I would prefer *Mandarin*. [ICE-HK S1B-079 #22-27]

In Example 79, the writer is demanding that simultaneous and successive bilingualism be differentiated; in Example 80, the second speaker claims that he or she can differentiate between the two terms. In Example 80, the coordinator *but* complicates the sentence, as it would seem to indicate that the second clause contradicts the first; the contradiction is not apparent. The important note

here is that ICE-GB contains no examples of *make a difference* with this meaning. Nonetheless, the evidence for this construction in ICE-SIN and ICE-HK is too limited to conclude any significant difference between the corpora in this regard.

6.2.5. Make: Concrete/Abstract ambiguities

Examples that are ambiguous for Concrete/Abstract include only the Direct Object *mistake*, and this situation occurs only three times in one corpus, ICE-HK, including Example 81, in which a teacher discusses student mistakes. It is unclear whether the teacher is referring to concrete errors (like words misspelled on the page due to the slip of a finger on the keyboard even when a typist knows the correct spelling) or conceptual errors (like not knowing the correct spelling).

> 81. Sometimes because they **make** the same mistake over and over
>   again I want to see why. [ICE-HK S1A-040 #165]

This ambiguity obviously relates to the ambiguous nature of the Direct Object. *Mistake* can relate to a concrete thing (an error printed on paper, for example), or to an abstraction in the mind (a conceptual error), and some, but not all, concrete mistakes are necessarily conceptual mistakes as well. In Example 81, it is unclear whether *mistake* refers to a mistake that is observable by any of the five senses (concrete) or purely conceptual (abstract), or both. There are no contextual clues in the corpus to solve that problem in this instance. In some instances, *make a mistake* is clearly concrete or abstract.

> 82. In fact I **made** a mistake once. I was coming from KL I threw
>   away some of my pots and pans and those were the best ones
>   you know [ICE-SIN:S1A-037#227-28]
> 83. Well the thing is that I **made** a terrible mistake in typo a typo
>   error [ICE-HK S1B-078 #239]

Example 82 is clearly a discussion of a conceptual error, as the physical act of throwing away the pots and pans was deliberate, but the decision to throw them away was erroneous; and in Example 83, *typo* indicates that the error was a concrete product of mistaken typography despite the fact that the writer was not mistaken in her conceptual understanding. This issue is an exemplar of the type of problem that can arise in manual analysis. However, it must be borne in mind that such instances are relatively rare, and examples of *make a mistake* like those above do not occur frequently enough to impact statistical analyses in any way. Without data on corpus frequencies, it would be easy to be overly distracted by such difficult cases. Corpora allow us to note that these difficult cases arise, but are in fact very unusual. For *make*, the distinction between concrete and abstract senses is generally maintainable.

There are no instances in the corpus of *make* occurring with coordinated Direct Objects such that one is abstract and the other concrete. Such examples would provide identity evidence for monosemy between Concrete and Abstract senses. The absence of such examples is in remarkable contrast to the most common alternate for *make*: *produce*. *Produce* often occurs with the two Direct Objects *goods* and *services*, the former concrete and the latter abstract. It is conceivable that in this coordinated Direct Object, the individual characters of *goods* as concrete and *services* as abstract are psycholinguistically obscured; such a possibility is beyond the scope of the present investigation. Nonetheless, it remains the case that *make goods and services* does not occur in use. There is no autonomy evidence with abstract and concrete *make* either, in which one sense is used while the other is explicitly denied. Nor is there independence evidence. The absence of such evidence does not prove that such examples cannot occur in natural language, but it is fair to assert that if they do occur in natural language, they are relatively rare, and that the concrete and abstract senses are relatively discrete in common instances of use.

There is one example which evidences antagonism between concrete and abstract senses of *make*, and that example depends on the polysemy of the

Direct Object *racket*, resulting in a deliberately humourous pun, or intentional zeugma. An article on the popularity of tennis in Singapore includes the section heading in Example 84.

84. **Making** a racket. [ICE-SIN W2D-014#51]

Example 84 suggests both 'building a tennis racket' and 'producing an uproar'. The conflation of the two meanings of *racket* here also requires a conflation of the Abstract and Concrete senses of *make*. The example in context affirms that this conflation is not typically acceptable, and is intended as humorous and zeugmatic (though native speaker input would be helpful to corroborate this interpretation). The fact that a writer and editor, and perhaps multiple editors, approved this title as an accessible and comprehensible bit of humour for a broad population of readers in Singapore probably lends the example more weight as evidence of the discreteness of meaning between the concrete and abstract senses of *make* in Singapore.

6.2.6. Make: Concrete/Delexical ambiguities

Just as *make* is rarely ambiguous between Concrete and Abstract senses, *make* is also rarely ambiguous between Concrete and Delexical senses. Ambiguity between delexical and concrete senses occurs in the following proportions:

	Instances of concrete *make*	Instances of Delexical *make*	Instances of Concrete/Delexical Ambiguity for *make*	Instances of Concrete/Delexical Ambiguity with *make a copy*
ICE-SIN	126	597	26	11
ICE-HK	164	685	12	4
ICE-GB	184	674	22	6

Table 4: Instances of Concerete/Delexical Ambiguity for make in ICE-SIN, ICE-HK, and ICE-GB

The single most common ambiguous usage in all corpora involves the Direct Object *copy*, such that *to make a copy* can be understood as 'to produce a concrete copy' or 'to copy'. Less frequent ambiguous Direct Objects include *note*, *mark*, *recording*, *draft*, *list*, and *crease*.

One zeugmatic example evidences antagonism between the Concrete and Delexical senses of *make*. ICE-GB includes the following headline:

85. Voters **make** all kinds of marks. [ICE-GB W2E-006 #1]

Example 85, like Example 84, seems to represent intentional zeugma. The article goes on to discuss the concrete marks made on concrete ballots and the abstract social and political marks, or impacts, of the votes. In fact, this example may be antagonistic three ways, between the Concrete, Abstract, and Delexical senses, isofar as *make a mark* is equivalent not only to *mark* (*v.*) but also to *produce an abstract mark* and *produce a concrete mark*. Judging from Examples 84 and 85, it may be that headlines are a common place to find intentional zeugma. In that case, a corpus of headlines might provide an interesting resource for polysemy evidence.

6.2.7. Make: Abstract/Delexical ambiguities

Abstract and Delexical examples are often difficult to distinguish from each other – there is frequent ambiguity. Indeed, as noted in 4.1, the Delexical sense might be seen as a subcategory of the Abstract sense. Ambiguities arise in part because nouns that represent acts also commonly represent the results of those acts. Consider the construction *make a synthesis* in Example 86.

86. …to combine Jewish values and a secular education and
somehow to **make** a creative synthesis out of the two. [ICE-GB,
S1B-047#030

In this example, *synthesis* may denote either the act of synthesising or the
result, the abstract synthesized product. That is, *to make a synthesis* may be
understood as either 'to synthesize' or 'to produce an abstract synthesized
product'. Other frequent Direct Objects are similarly problematic: *to make a
decision* is definitely 'to decide' but also quite possibly 'to produce a decision';
to make a change is 'to change' but also 'to produce a change'. These
ambiguities are connected to the ambiguous nature of the Direct Objects, and
were noted in 6.1.

There are a few instances in the three corpora of potential identity
evidence for abstract and delexical *make*, shown in Example 87.

87. We have **made** a pact. A new start. [ICE-GB W2F-008 #17]

Make a pact cannot be Delexical and must be understood as 'Produce
(Abstract)': *pact* (v.) is unacceptable in all three corpora so the example must
be understood as *produce a pact*. *Make a start* seems to represent the Delexical
sense as it is equivalent to *start* (v.), but it might also represent the abstract
sense insofar as *a start* is the result of the act described by *make*. It is therefore
possible to interpret this example as a straightforward case of abstract *make*; it
is also possible to interpret this example as a juxtaposition of abstract *make* and
Delexical *make*, suggesting identity between those two senses. It is noteworthy
that this example is drawn from written language and that the full stops are part
of the original text. The full stop after *pact* divides the construction between the
two Direct Objects, perhaps rendering it more acceptable. Alternatively, *a new
start* might be seen as a sort of appositive, such that the full stop might be
replaced by a colon or a comma. This is a curious example of, debatably,

identity evidence arising naturally in the corpus, but not an example that provides conclusive evidence for or against the present hypothesis.

Similarly, ICE-SIN includes the following:

88. …Red Guards would also **make** minor ambushes and small-scale battles on the Nationalists. [ICE-SIN W1A-020 #114

Example 88 is particularly tricky. *Make an ambush* and *make a battle* only occur in this single example. Are they delexical or abstract? The prepositional phrase *on the Nationalists* adds an additional variable. Strictly following my criteria, they are both Delexical: *make an ambush* is equivalent to *ambush* (*v.*) and *make a battle* is equivalent to *battle* (*v.*), but this is certainly disputable, and it illustrates the complexity of the issue at stake.

The important point to be taken away from these examples is, I believe, that in the three corpora examined here, speakers only very rarely coordinate an abstract Direct Object and a Delexical Direct Object in actual use. In contrast to that rarity, instances of coordinated Delexical Direct Objects are common. (On the other hand, instances of two clearly abstract Direct Objects are extremely rare; I follow up on this point in 6.2.9).

	Make with Two Concrete Direct Objects	*Make* with Two Abstract Direct Objects	Instances of *Make* with Two Delexical Direct Objects
ICE-SIN	1	0	5
ICE-HK	6	1	3
ICE-GB	4	0	2

Table 5: Instances of *make* with two concrete Direct Objects, two delexical Direct Objects, and two abstract Direct Objects in ICE-SIN, ICE-HK, and ICE-GB.

Examples of Delexical *make* with two Direct Objects (or in the case of Example 90, which is a passive construction, two Subjects which are also patients) include the following:

> 89. And we would help to facilitate to **make** an assessment and evaluation of the building… [ICE-SIN S1B-041#10]
> 90. We should bear in mind that all decisions and actions should be **made** with the animal welfare as the first priority… [ICE-HK W2B-027#131]
> 91. …uh so that they can **make** career switches or changes… [ICE-SIN S1B-022#19]
> 92. I think in the future when government **makes** big statements, announcements you should do what ministers do in the UK. [ICE-HK S2A-033 #115]

The data show that speakers do not avoid coordinated Direct Objects with *make*, but they do avoid coordinating Direct Objects from different hypothesized semantic categories of *make*. This seems to be evidence for the relative discreteness of the Delexical and Abstract semantic categories in actual use.

Straying interestingly from standard Delexical usage is Example 93, in which *make a complaint* is abstract, with the meaning 'produce a complaint', rather than Delexical, with the meaning 'complain'.

> 93. It has been noticed that the standard of the water supply of the above building is found to be unacceptable for a long period and it **made** a lot of complaints from our occupants. [ICE-HK W1B-019 #181]

The creative flexibility underlying Example 93 is striking.

Ultimately, Delexical instances of *make* are not entirely straightforward to identify. The corpus would seem to support the existence of a relatively distinct category in Delexical *make*. I conclude that it is reasonable to consistently interpret those Direct Objects that have related verbs as Delexical.

6.2.8. Onomasiological analysis: Produce (Concrete)

Given that instances of concrete *make* are generally very clear, and given that limited antagonism evidence suggests discreteness for concrete *make* in use, the present study now proceeds to identify semantic alternates; to test the feasibility of the sense 'Produce (Concrete)' onomasiologically; and to investigate preferences for semantic alternates in actual use. First, all Direct Objects of concrete *make* were identified. Then, Direct Objects that occur at least twice per corpus were selected, and all other verbs in each corpus which occur with the same Direct Objects, and which convey generally the same meaning, were identified. That is, the corpus shows that people discuss *making cakes* and *preparing cakes*, and these suffice as alternates, but *ruining cakes* or *dropping cakes* do not qualify as semantic alternates. In fact, all semantic alternates were effectively identified via Direct Objects that occurred three or more times; Direct Objects of *make* that occurred only twice yielded no new evidence for semantic alternates. It is reasonable to conclude that investigating Direct Objects that occur only once would not yield significant additional evidence of alternates. The count of alternates, as conducted, should facilitate a reasonably thorough, evidence-based onomasiological comparison for each sense. It is certainly possible that other alternates exist, but this approach, based in the corpora, is justified and reasonably complete. In fact, the basis in the corpora makes this approach particularly strong because it avoids the culturally and socially biased reflection that underlies thesauri in favour of a list grounded directly in the evidence of actual use in each locale. Concrete *make* in the three corpora has the alternates shown in Table 6.

Alternates for concrete make:	produce, create, bake, prepare, manufacture, cook, generate, form, build, emit, construct, develop, draw, yield, erect, compile, dig

Table 6: Alternates for concrete *make*, as evidenced in ICE-SIN, ICE-HK, and ICE-GB

We might have expected that *make* is a relatively general term semantically, and that its alternates would tend to be more specific, and that expectation seems to be met by the data. Other rarer alternates might be evidenced in larger corpora, and reflection by native speakers in each region might yield numerous additional alternates, including metaphorical alternates and creative extensions.

Contrary to the hypothesis presented in 4.1, onomasiological evidence suggests separate senses for 'Produce (Concrete, Food)' and 'Produce (Concrete, Non-food)'. I had hypothesized one semantic category for 'Produce (Concrete)', including both food and non-food Direct Objects, but I encountered only one instance of identity evidence for that single group (see Example 67). The difference between 'Concrete, Food' and 'Concrete, Non-food' might seem to be related to the Direct Object more than to the semantics of *make*. However, there is a clear and consistent divide between the alternates that occur with Direct Objects referring to food and to non-food items. There is no such divide with other groups of Direct Objects. That is, individual Direct Objects may occur only with *make* and one other verb in these three corpora: for example, we can *make a hole* or *dig a hole*, but there is no evidence in the corpus for *manufacturing a hole*, much less for *erecting a hole* (though a larger corpus might provide evidence for *boring a hole* or *drilling a hole* as well). Nonetheless, there is apparently no systematic semantic divide as marked as the food/non-food distinction. The alternates *bake* and *cook* in the corpora take exclusively Direct Objects referring to food, while *prepare* largely takes Direct Objects referring to food. No other alternates take Direct Objects referring to food. This seems to be compelling evidence that there is a food/non-food sense distinction. In turn, this suggests that the selection processes between food and

non-food use of *make* may in fact be distinct as well: with food, users select between *make* and *cook*, for example, but not *produce*, whereas with non-food Direct Objects, users select between *make* and *produce*, but not *cook*. Equally importantly, it seems inappropriate to compare usage tendencies of *cook* and *bake* to the set of other verbs listed in Table 6. That is, users will not generally select between *cook*, *bake*, and *manufacture*. I therefore opt to analyse 'Produce (Concrete, Food)' and 'Produce (Concrete, Non-food)' separately.

Figure 5 and Figure 6 present preferences for *make* and its alternates with the sense 'Produce (Concrete, Food)' as probabilities of selecting each alternate in each variety, with error bars representing Wilson intervals.
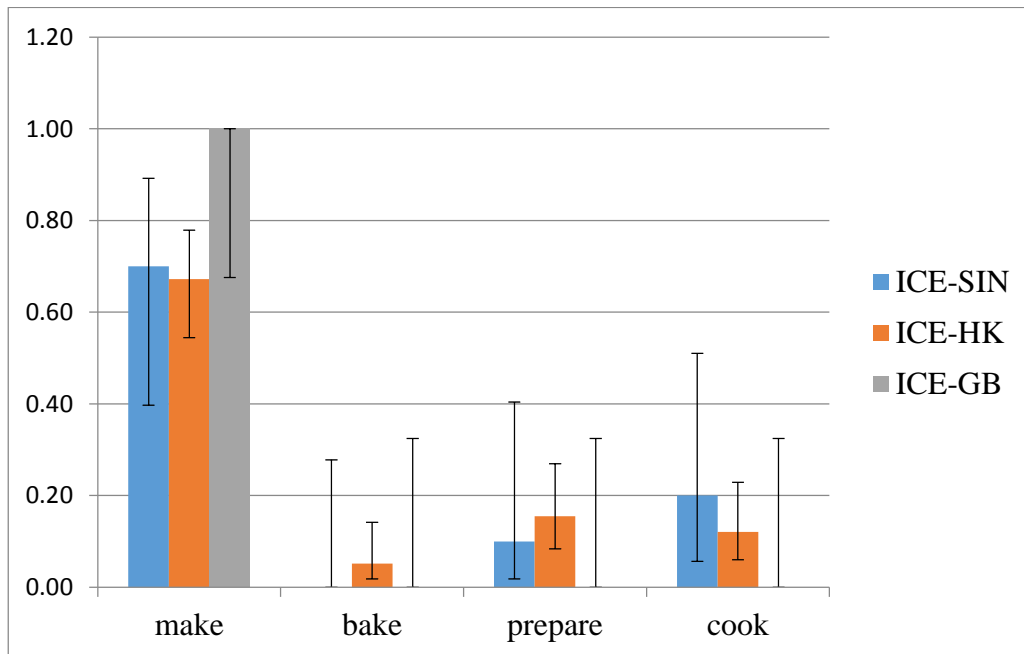
Figure 5: Verbs with the sense 'Produce (Concrete, Food)' in the written portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
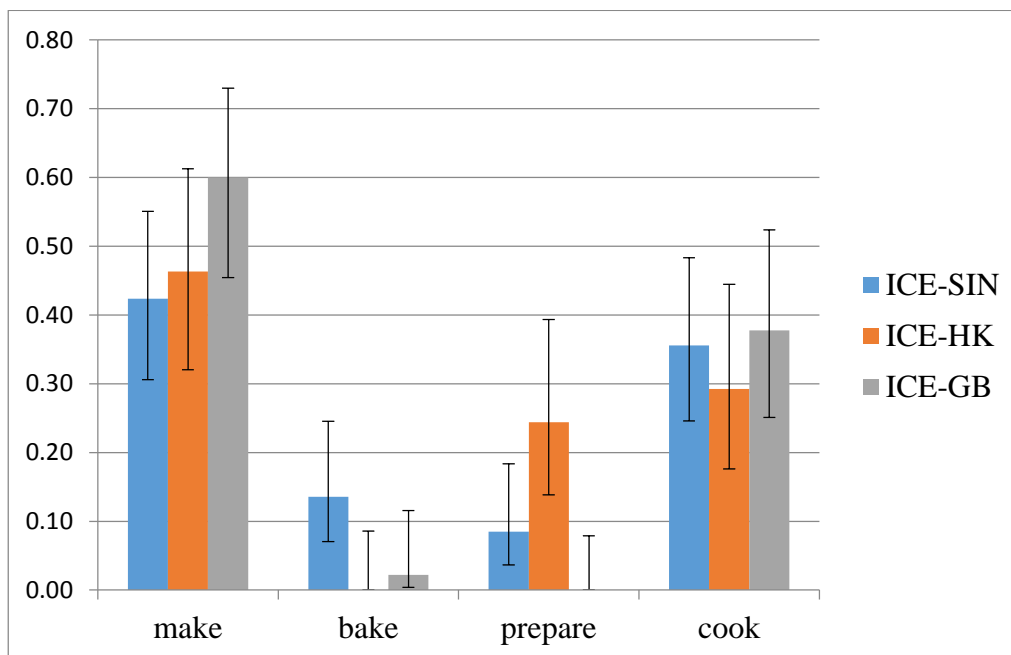
Figure 6: Verbs with the sense 'Produce (Concrete, Food)' in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

The probabilities presented in these figures, along with their error bars, indicate various measurements of significant differences between words and corpora at $p \leq 0.05$ (see 5.4.3). For the presentation discussion, I would like to focus on a more fundamental fact that is apparent in this data visualization: it is immediately striking that the general amount of data for the sense 'Produce (Concrete, Food)' is relatively low: for example, there are no examples of *bake*, *cook*, or *prepare* in the written portion of ICE-GB. The other issue that immediately arises is that *cook* and *bake* are not necessarily alternates for each other, and will not always be alternates for *make* with Direct Objects representing food. That is, *make* and *bake* are not near-synonyms. *Prepare*, on the other hand, is the most semantically general alternate, and it seems to alternate with *make* in the corpora nearly universally with Direct Objects representing food. Figure 7 and Figure 8 compare *make* and *prepare* with the sense 'Produce (Concrete, Food)', which establishes a strong baseline for nearly universal alternation.

Figure 7: *Make* and *prepare* with the sense 'Produce (Concrete, Food)' in the written portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

Figure 8: *Make* and *prepare* with the sense 'Produce (Concrete, Food)' in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

In writing, all three corpora show a preference for *make* over *prepare*. In speech, however, the one significant difference between the corpora is that *prepare* is not dispreferred to *make* in ICE-HK, but is dispreferred in the other two corpora (according to a Newcombe-Wilson test with continuity correction).

The separation of food and non-food senses raises the question of constraints upon onomasiological selection processes, insofar as every semantic alternate of *make* is not necessarily a semantic alternate in every case of *make*, and all alternates are not necessarily alternates for each other. Generally, it is necessary to examine the various constraints upon the choice between *make* and its semantic alternates. This problem might be approached from multiple directions. I have already stated one research question thusly: given that language users want to express the concept 'Produce (Concrete)', how do they tend to do so? Starting from that general approach, we can consider all

alternates, as I have suggested above, and then explore the constraints on each, and consider what circumstances disallow certain alternates. On the other hand, we could also pose a much more specific question: Given that language users want to express the concept 'Produce (Concrete)' with a specific Direct Object, how do they tend to express that concept? For example, how do language users tend to express 'Produce (Concrete)' with the Direct Object *product*? Put differently, do British English speakers tend to *make products*, while Singapore English speakers tend to *produce products*, and so on? Such a difference could represent a semantic norm, insofar as speakers from different regions might tend to fill a specific semantic 'slot' in a specific construction in very different ways. Table 7 shows all alternates that appear with the common Direct Object *product*, the most common non-food concrete Direct Object across all three corpora, with frequencies for each alternate in parentheses; Figure 9 displays those numbers as probabilities, with error bars representing Wilson intervals.

ICE-SIN	ICE-HK	ICE-GB
make (8)	*produce* (6)	*produce* (6)
produce (3)	*make* (2)	*make* (5)
create (2)	*build* (1)	*develop* (3)
manufacture (1)	*create* (1)	*create* (2)
	generate (1)	*generate* (1)
		manufacture (1)

Table 7: Verbs with the sense 'Produce (Concrete, Non-Food)' occurring with the highly frequent concrete Direct Object PRODUCT in ICE-SIN, ICE-HK, and ICE-GB. Numbers in parentheses represent the number of instances of each verb in each corpus.
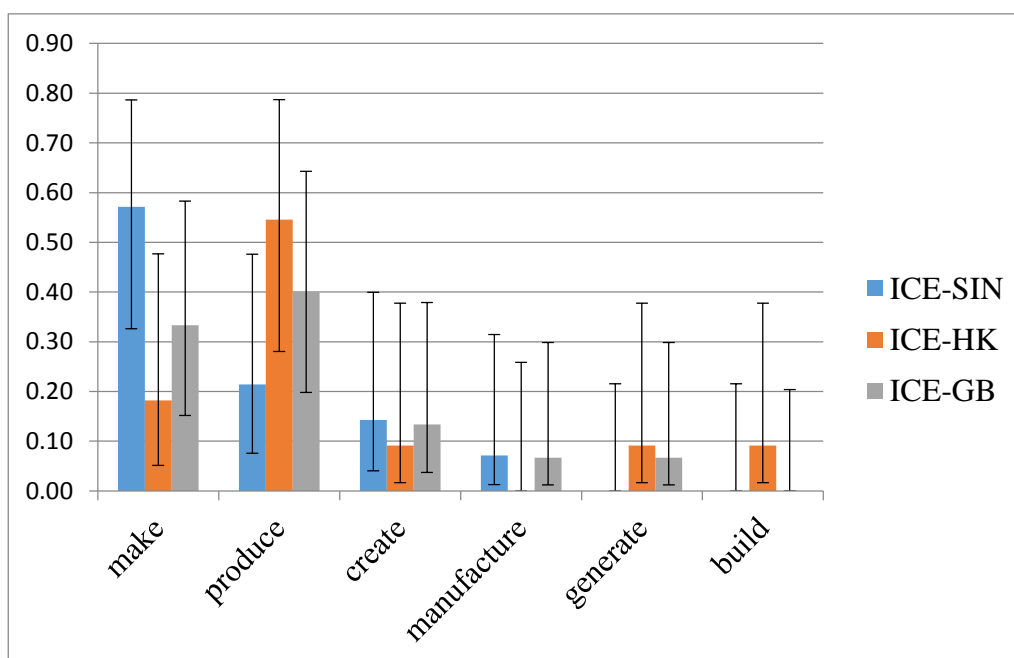
Figure 9: Verbs with the sense 'Produce (Concrete)' occurring with the highly frequent concrete Direct Object PRODUCT in ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

In Figure 9, some apparently large effect sizes are evident, e.g. the difference in the probabilities between MAKE and PRODUCE in ICE-HK and ICE-SIN. Upon reflection, some observers might wonder whether the effect size between MAKE and PRODUCE in ICE-GB is not also large. The confidence intervals on these measurements reflect both effect size and sample size (see 5.4.3), and they show that, given the confluence of effect size and sample size, our confidence in these observed probabilities is relatively low given the sample size. Even if we compare the two highest frequency verb choices with the Direct Object PRODUCT (that is, if we compare MAKE + PRODUCT and PRODUCE + PRODUCT, as in Figure 10), the constructions in question do not display any variation in statistical significance, in part because they are simply too infrequent to conclude with confidence that the populations in question will actually differ in 19 out of 20 samples (i.e. $p \leq 0.05$). Considering significance measures based on 18 out of 20 (or 1 out of 10, i.e. $p \leq 0.1$) is relatively rare in statistics. Such an approach would accept a considerably lower confidence in research findings than is generally acceptable (see 5.4.3). Moreover, even when examining a high-frequency Direct Object such as PRODUCT, it is clear that many example constructions, such as *manufacture a product*, occur too infrequently to make definitive statistical (or logical) conclusions about their probability in the population. CREATE, MANUFACTURE, GENERATE, and BUILD occur at statistically indistinguishable probabilities.
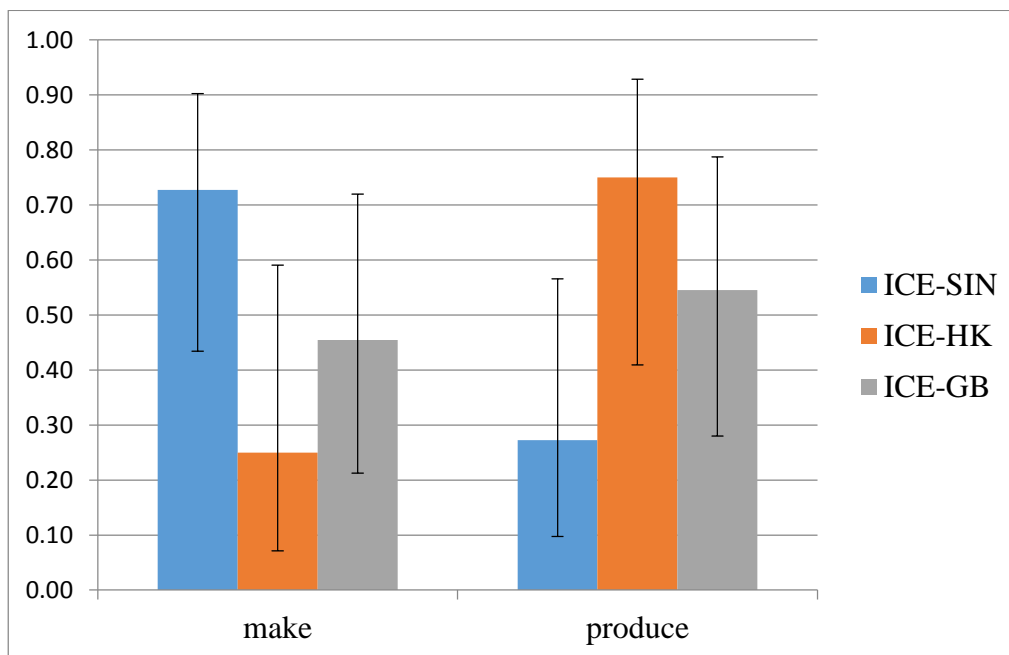
Figure 10: *Make* and *produce* with the sense 'Produce (Concrete, Non-Food)' occurring with the highly frequent concrete Direct Object *product* in ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

The specific constructions MAKE + PRODUCT and PRODUCE + PRODUCT are more common than constructions with other semantic alternates for *make* and other Direct Objects, and these specific constructions still show a scarcity of data for practical purposes. The same problem arises, but to an even greater degree, in comparing *make* and *bake*, for example, with the Direct Object *bread*. The error bars in Figure 10 reflect both effect size and sample size; the error bows show that given the combination of effect size and sample size, we cannot be confident that 19 out of 20 samples from the same populations would yield similar results (see 5.4.3). This example is in fact very similar to the hypothetical example presented in 5.4.3: the effect sizes might appear very large: e.g. the difference in probabilities between MAKE + PRODUCT and PRODUCE + PRODUCT in ICE-HK may, upon reflection, seem rather large. The raw tallies, however, are few enough, the difference between them small enough, that they are not significantly different at *p*≤0.05 (raw tallies appear in

the appendix). Considering significance measures based on 18 out of 20 (or 1 out of 10, i.e. $p \leq 0.1$) is not advisable: such an approach would accept a considerably lower confidence in research findings than is generally acceptable (see 5.4.3).

Given the difficulty of analysing individual constructions due to their low frequency, it is necessary to consider aggregating all concrete (non-food) Direct Objects and all semantic alternates together, and analysing them as a general mass representing, roughly, the sense 'Produce (Concrete, Non-food)' in all of its forms. However, it is clear that aggregating these instances will add noise to the analysis by including some examples that are not universally replaceable with each other – the baseline for the probabilities will be a bit too high. For example, *produce* is a very general term that appears in the corpora with a wide array of Direct Objects, but *dig* is a very specific term that alternates with *make* in only a few very specific cases. Moving forward, I first aggregate the data for all alternates for a very rough picture of the phenomena in question, and then analyse more closely particular alternates that alternate more consistently. For example, Figure 11 and Figure 12 display all semantic alternates with the sense 'Produce (Concrete, Non-food)' in all corpora, as probabilities for selecting each alternate in each corpus. After considering this rough picture, I then go on to look more closely at the alternate *produce*, which is nearly universally replaceable with *make*, and which also occurs frequently enough in all corpora to draw meaningful conclusions.
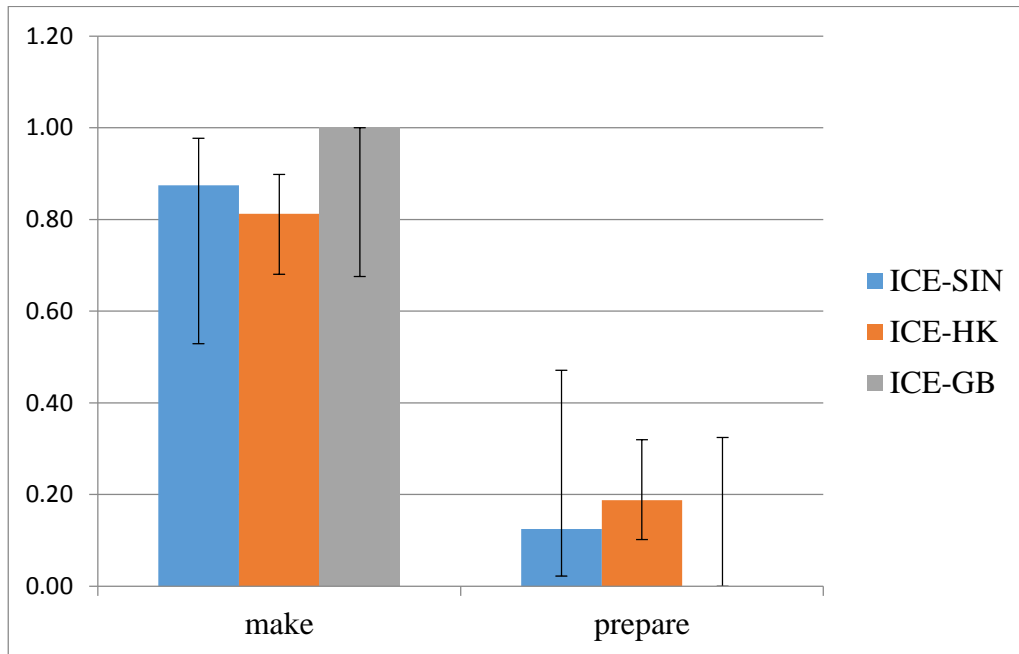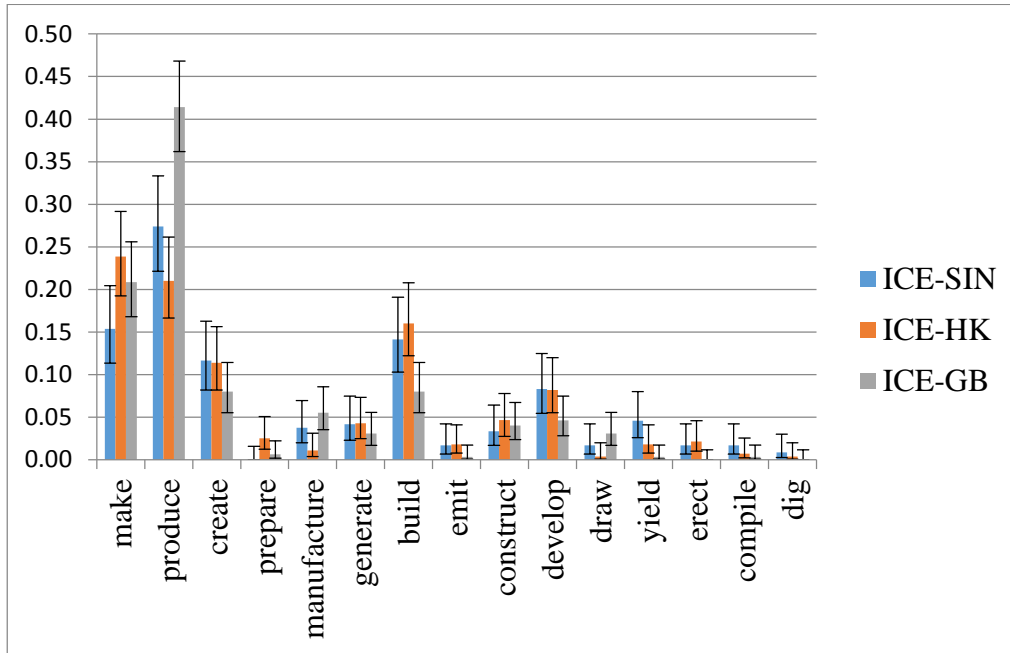
Figure 11: Verbs with the sense 'Produce (Concrete, Non-food)' in the written portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
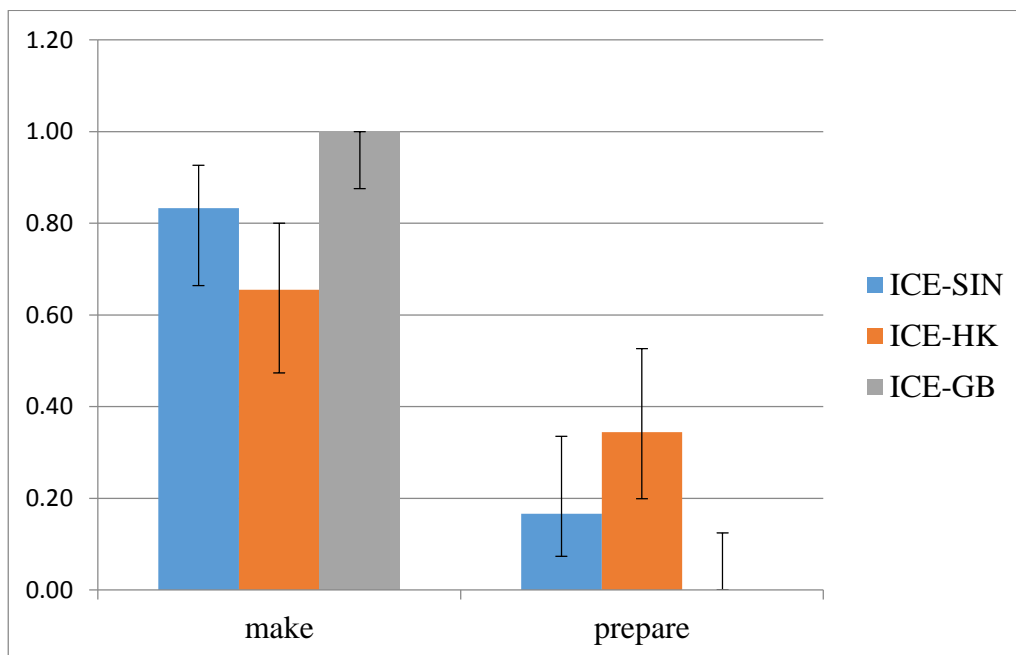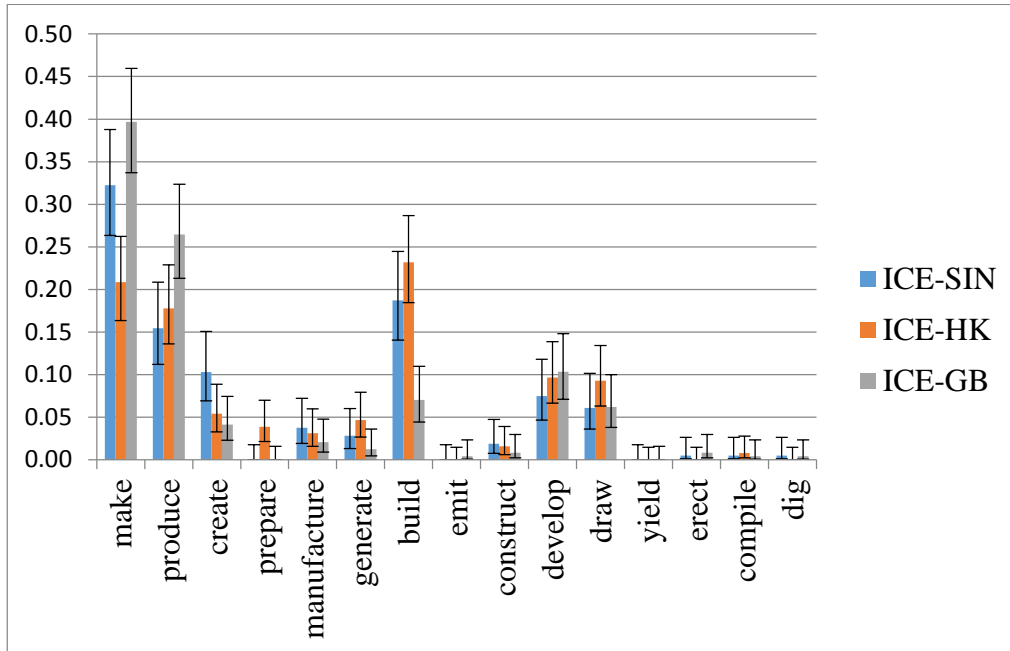
Figure 12: Verbs with the sense 'Produce (Concrete, Non-food)' in the spoken portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

It is apparent from Figure 11 and Figure 12 that preferences for *produce*, *create*, and *build* vary significantly across written and spoken data. We can look even more closely at these selection phenomena in the corpora by comparing only *make*, *produce*, *create*, and *build* without reference to the other alternates. Aggregating verbs that do not in fact universally alternate with each other introduces noise to the analysis. Eliminating the other alternates reduces some of this noise. It also allows for a more precise research question: 'Does any variety prefer *produce*, *make*, *create*, or *build* in a unique way and in what contexts does that preference hold?' This is a valid question, particularly because we can see from Figure 11 and Figure 12 that the other alternates do not vary significantly. Figure 13 and Figure 14 show this phenomenon more clearly.
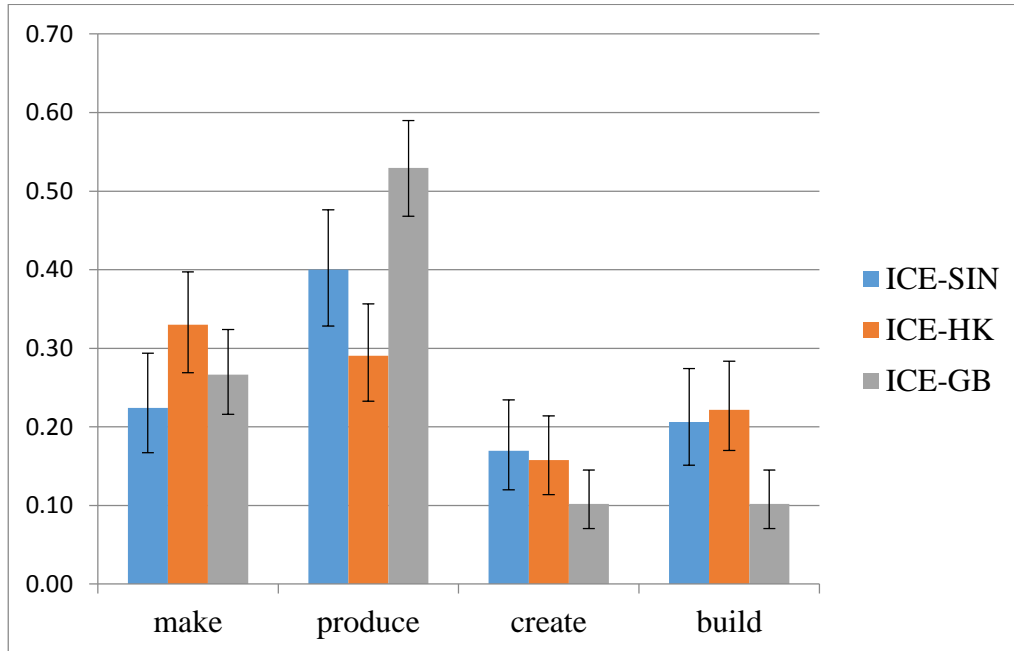
Figure 13: Instances of *make*, *produce*, *create*, and *build* with the sense 'Produce (Concrete, Non-food)', in the written portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
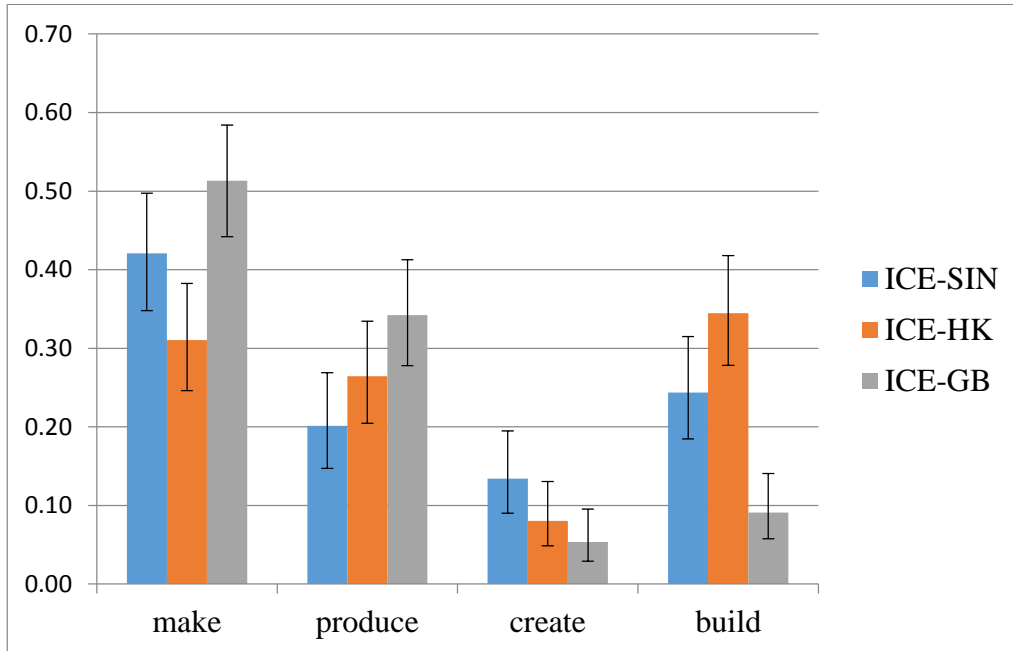
Figure 14: Probability of selecting *make*, *produce*, *create*, and *build* with the sense 'Produce (Concrete, Non-food)', in the spoken portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

I first consider the variation in preference for *produce*. Perhaps unsurprisingly, ICE-GB exhibits a strong preference for *make* in spoken language and for *produce* in written language. Figure 15 and Figure 16 show this more clearly.
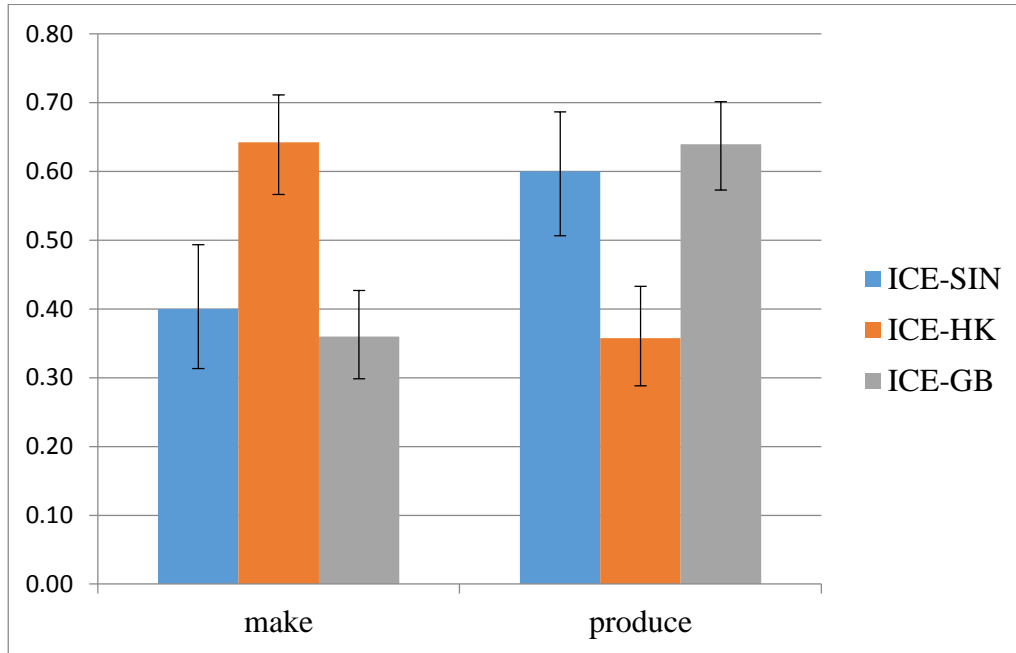
Figure 15: Probability of selecting *make* and *produce* with the sense 'Produce (Concrete, Non-food)', in the written portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
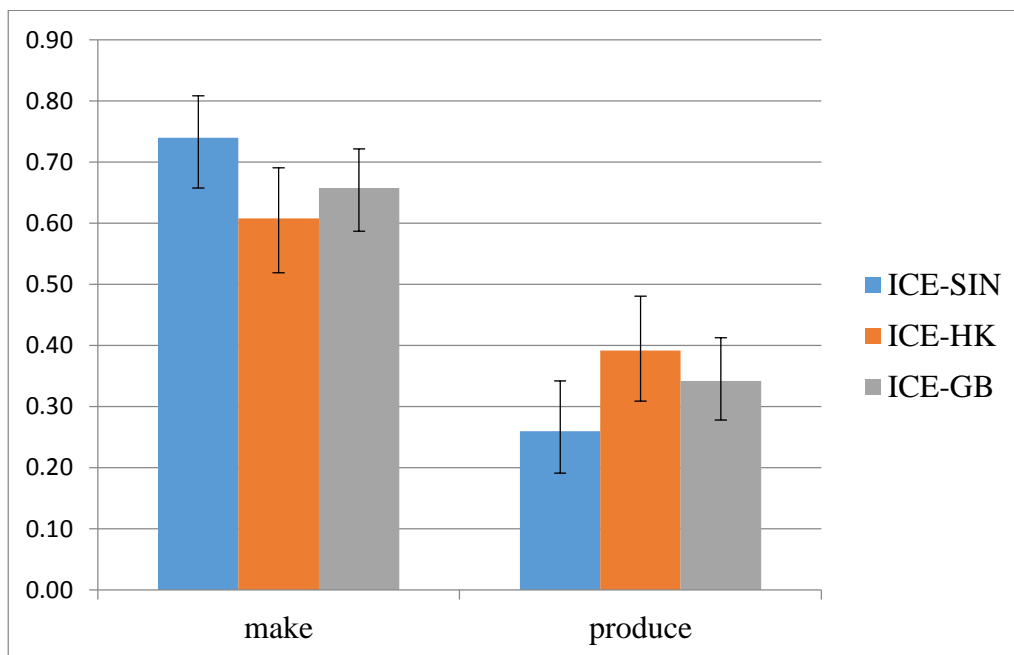
Figure 16: Probability of selecting *make* and *produce* with the sense 'Produce (Concrete, Non-food)', in the spoken portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

Make, as a monosyllabic Germanic lexical item, may generally be considered more colloquial than its polysyllabic Latinate alternate *produce*. Not surprisingly, then, all corpora show a preference for *make* over *produce* in speech. ICE-GB and ICE-SIN, unsurprisingly, prefer *produce* over *make* in writing. It is noteworthy, then, that ICE-HK prefers *make* over *produce* in writing. Just as ICE-HK exhibited unique onomasiological preferences between *make* and its polysllabic, Latinate alternate *prepare* with the sense 'Produce (Concrete, Food)', ICE-HK is singled out as unique in its onomasiological preferences with the sense 'Produce (Concrete, Non-food)' as well. It is clear that ICE-SIN and ICE-GB exhibit the expected preference for *make* in speech and *produce* in writing, while ICE-HK exhibits no such preference. This difference is not due to any particular text with an unusually high or low number of instances of either alternate, so the finding appears to be robust.

Produce may also exhibit different polysemy from *make* in one important way. *Make* tends to resist coordinating a concrete Direct Object with an abstract Direct Object. As discussed in 6.2.5, *produce* regularly coordinates a concrete Direct Object with an abstract one in the Direct Object *goods and services*.

Moving from *produce* to *build*, ICE-GB shows a unique, statistically significant preference against *build* in both speech and writing. That preference does not seem to be attributable to any particular texts in ICE-SIN or ICE-HK that might skew the data towards *build* in those corpora.[34] The semantic specificity of *build*, however, in contrast to the generality of *make* and *produce*, must be born in mind here, and this conclusion is only tentative.

Finally, it would seem from Figure 13 and Figure 14 that ICE-SIN might have shown a unique preference for *create* in both speech and writing, but an application of a Newcombe-Wilson test with continuity correction shows no significant difference in that regard – ICE-SIN does not in fact display a preference for *create*.

6.2.9. Onomasiological analysis: Produce (Abstract)

I hypothesized that abstract *make* would constitute a coherent category (see 6.2.3). However, I found no autonomy evidence, antagonism evidence, identity evidence, or truth-condition evidence to confirm the coherence of Produce (Abstract) as a discrete sense of *make*. The onomasiological study suggests that most constructions with abstract Direct Objects are actually unique and also semantically non-transparent, insofar as almost none of them allows an

---

[34] ICE-HK was compiled during the debate over the building of a new airport in Hong Kong. While *build an airport* does occur numerous times in ICE-HK, those instances do not significantly skew the data; removing those instances from the analysis yields no change in the determination of statistical significance. This issue illustrates the importance of tracking real-world context in corpus linguistics, but also shows that relatively small million-word corpora like the ICE corpora are not necessarily overly susceptible to idiosyncratic contextual fluctuations.

alternate construction in which *make* is replaced with another verb, much less replaced with a verb semantically comparable to *make* such as *produce*. Table 8 displays constructions that I had originally categorized as Produce (Abstract) which occur at least twice in the corpora. Some were more tenuous than others in the first place (see 6.2.3), but none of them hold up onomasiologically, as no consistent alternates occur in the corpora.

ICE-SIN	ICE-HK	ICE-GB
point (33)	*sense* (24)	*point* (32)
sense (18)	*point* (17)	*difference* (30)
difference (18)	*mistake* (17)	*mistake* (24)
effort (15)	*effort* (17)	*effort* (18)
mistake (11)	*difference* (13)	*sense* (15)
trip (6)	*case* (6)	*trip* (6)
appointment (4)	*law* (5)	*appointment* (5)
mark (3)	*deal* (5)	*mess* (4)
loss (3)	*policy* (4)	*thing* (3)
headway (3)	*appointment* (4)	*history* (3)
virtue (2)	*trouble* (3)	*case* (3)
issue (2)	*trip* (3)	*variable* (2)
harmony (2)	*name* (3)	*weather* (2)
	mess (2)	*secret* (2)
	fallacy (2)	*loss* (2)
	condition (2)	*inroad* (2)
	climax (2)	*headway* (2)
	balance (2)	*deal* (2)
		analogy (2)

Table 8: Abstract Direct Objects of *make* in ICE-SIN, ICE-HK, and ICE-GB. These constructions were originally categorized as representing the sense 'Produce (Abstract)', but that sense proved not to be an internally coherent category. Numbers in parentheses represent number of occurrences of each construction in each corpus.

Only very few alternates for *make* occur at all in the corpora for any of these constructions. *Cause trouble* occurs twice in ICE-HK and twice in ICE-GB, while *create trouble* occurs once in ICE-HK, so they are relatively infrequent. Most importantly, neither *cause* nor *create* occurs with any of the other abstract

Direct Objects in Table 8. *Make an effort* gives rise to some interesting alternates for *make*; constructions in the corpora include *put in an effort*, *put an effort*, *pay an effort*, *expend an effort*, and *give an effort* in ICE-HK only; *take efforts* and *spend efforts* occur in both ICE-HK and ICE-SIN.[35] Figure 17 displays these alternates as probabilities.
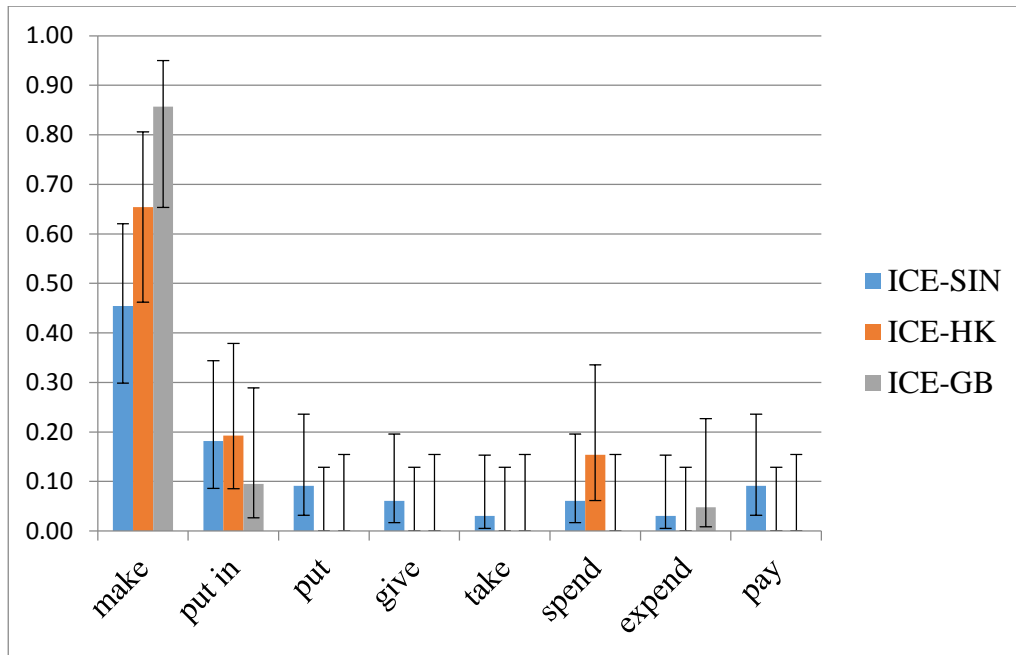
Figure 17: Alternates for *make* in the construction *make an effort*, in ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

Each of the alternates is less common than *make an effort*, and none of the other alternates occur frequently enough to reach any firm conclusions about usage preferences between them. More importantly, none of those alternates is an alternate for *make* in any other construction with an abstract Direct Object. The

---

[35] ICE-SIN W2C-014 is a text about a company named Effort Holdings. It is worth noting that an automated collocational analysis of *effort* in this case would result in a significant number of collocates related to real-world facts about that company. The present approach, which includes manually analysing each instance of *effort*, avoids that problem.

crucial point here is that the onomasiological analysis shows that *make* itself is not consistently replaceable by any other verb with comparable meaning, such as *produce*. Instead, each of these constructions is likely best catalogued as conveying a unique meaning as an entire construction, and that unique meaning is not transparent or predictable based on the possible meanings of *make* and the abstract Direct Object. It would have been conceivable that a variety of English might have developed alternates for these constructions such as *produce sense* rather than *make sense*, or *create a point* rather than *make a point*. In fact, Hong Kong may have developed unique alternates for *make an effort*, in the form of *put an effort*, *give an effort*, and *pay an effort*, while Hong Kong and Singapore may both have developed a new alternate in *spend an effort*. None of those constructions appears in ICE-GB. Additional corpus data would be necessary to corroborate the uniqueness and productivity of those constructions in Hong Kong and Singapore. In addition, alternates such as *generate mistakes* or *yield differences* are conceivable, but neither occurs in any of the corpora, so neither seems to be in regular use in any region. This consistency across the corpora for *make* with an abstract Direct Object is remarkable.

6.2.10. Onomasiological analysis: Delexical

There are no clear and consistent semantic alternates for the Delexical or light use of *make*. Each delexical construction instead alternates with a verb that is related to the Direct Object; indeed, that relationship is definitive of the category as presented here. Delexical *make* seems to be more semantically transparent and predictable than the Produce (Abstract) sense of *make* (see 6.2.9), in that Delexical *make* constructions consistently alternate with a verb related to the Direct Object. There is a consistent and internally coherent system of alternation, even if the end result is a unique alternate for each construction. This section proceeds with an analysis of Delexical *make* and its alternates, which contributes to existing discourse on corpus frequencies of light verbs (see 3.4). Existing studies have not approached light verbs onomasiologically, and I

would like to show that an onomasiological approach is in this case extremely useful.

Direct Objects that occur frequently enough, and whose alternates occur frequently enough, for statistical analysis include: *make use*, *make a decision*, *make a change*, *make contact*, and *make a contribution*. A larger corpus would certainly provide more data for a fuller analysis of a broader array of constructions, but a larger corpus would also require considerable additional manual analysis in recognizing and categorizing each individual instance of Delexical *make*, which could be prohibitively costly. Figure 18 through Figure 22 display probabilities of selecting these common Delexical constructions and their related verb alternates in each corpus. Because spoken and written sections of the corpus do not differ significantly from each other, the figures present the data for each corpus as a whole rather than for spoken and written sections separately.
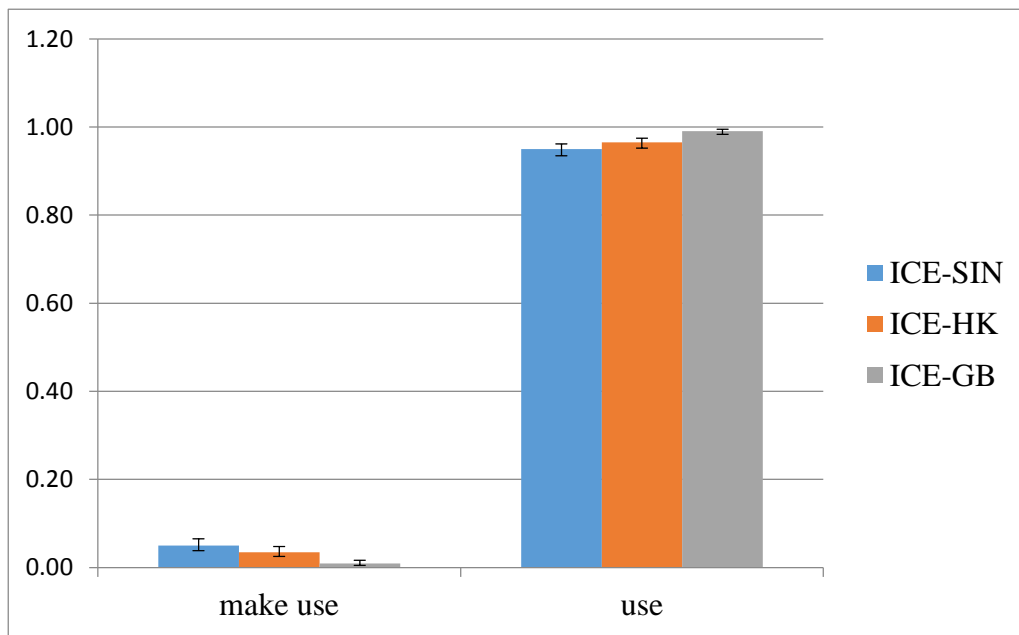
Figure 18: Instances of *make use* and *use* (*v.*) in ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
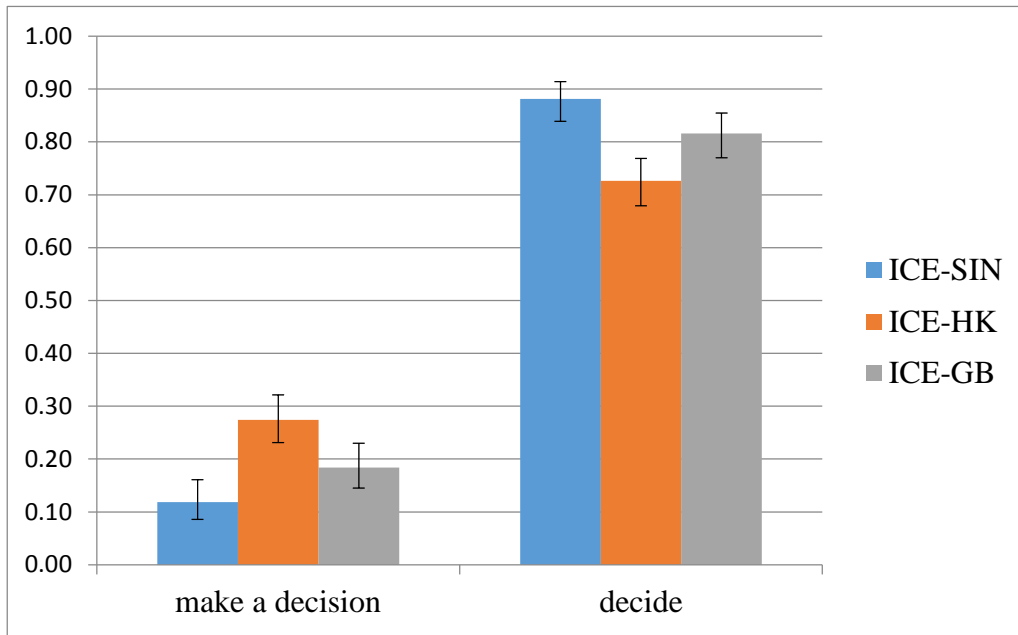
Figure 19: Instances of *make a decision* and *decide* in ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
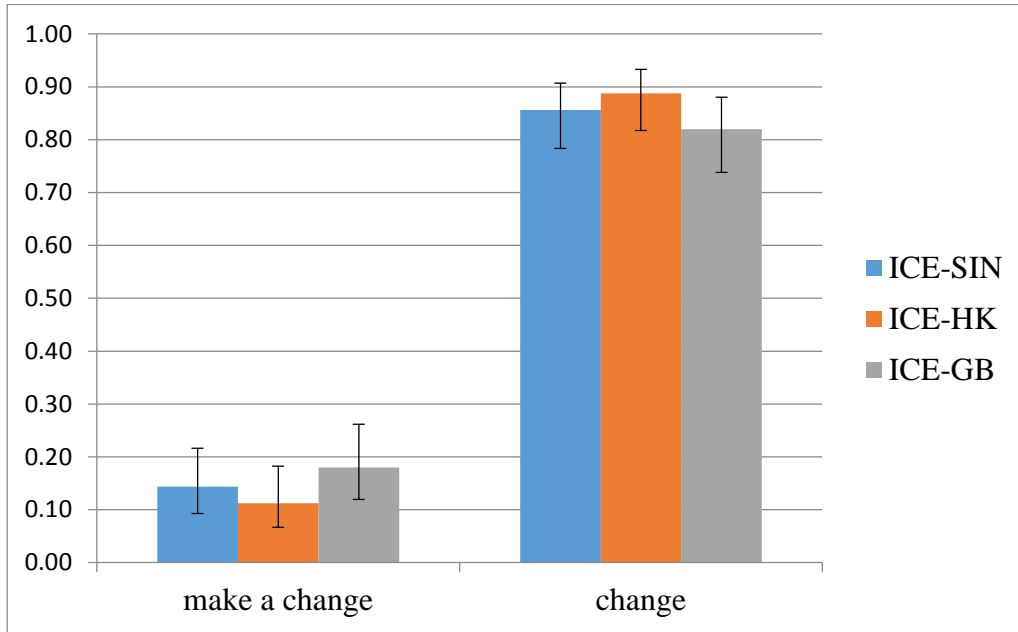
Figure 20: Instances of *make a change* and *change* (*v.*) in ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
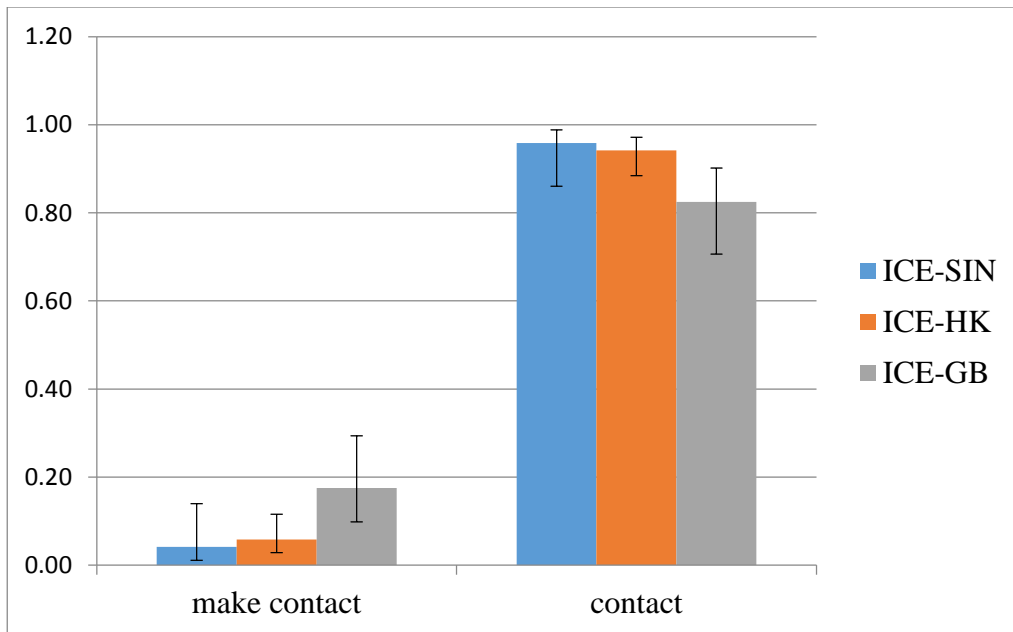
Figure 21: Instances of *make contact* and *contact* (*v.*) in ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
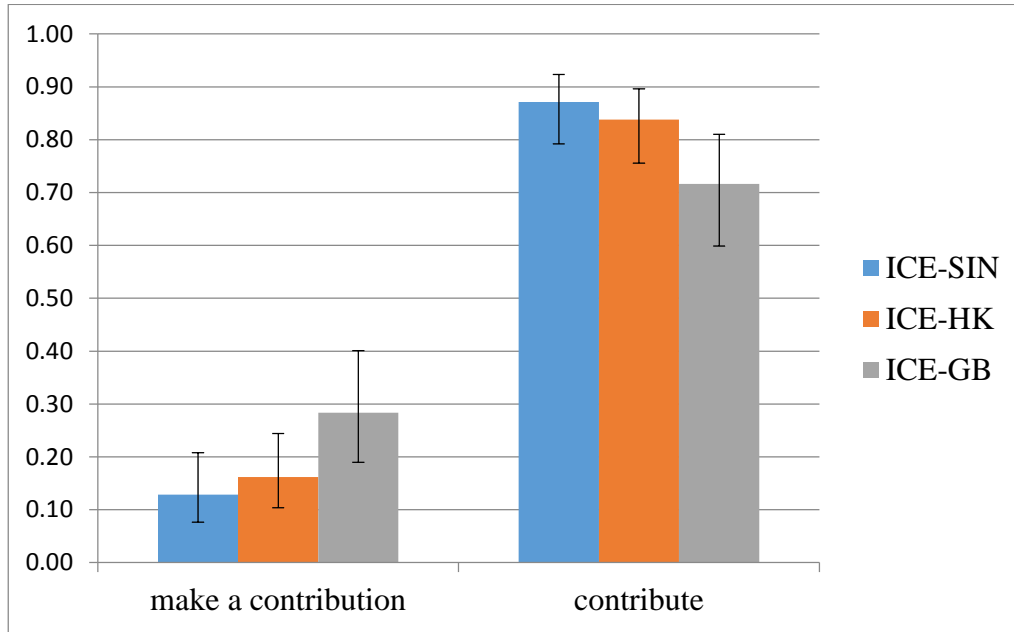
Figure 22: Instances of *make a contribution* and *contribute* in ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

In all cases of Delexical use, the alternate of the Delexical *make* construction is significantly preferred over the Delexical *make* construction. This is also true of spoken and written sections of the corpora when analysed separately. The preference for the verbal cognate over the *make* construction does not vary across the three corpora. It would have been conceivable that a particular variety of English might develop a unique preference for or against the Delexical construction with *make*, or even for or against Delexical constructions in general, but that does not appear to be the case for *make*. The significance of this finding in relation to existing studies on light verbs, as well as to investigations of cognitive salience and the hypothesis of onomasiological salience, is discussed in 9.1 and 9.3.

6.2.11. Make: Additional factual evidence

As noted in 2.1, corpora provide not just frequency evidence, but also factual evidence, via individual examples in use. Factual evidence shows what *can*

happen in actual use by showing at least one example of a feature, even if the feature occurs too infrequently to make conclusions about usage trends. Some noteworthy factual evidence can be drawn from the corpora in relation to *make*.

One significant grammatical construction occurs multiple times in ICE-SIN and ICE-HK; although it falls outside the focus of the present study, it is remarkable that it has not been previously reported or discussed. It will be referred to here as the *make me difficult* construction.

94. We think that this language barrier **make** us difficult to communicate. [ICE-HK S1A-070 #232]

95. Because I may have come from a background which **makes** me very difficult to understand uh comedy… [ICE-HK S1B-009 #98]

96. …we work with uh Custom to educate them and to cooperate them with to recognize them uh to **make** them easier to know what which is CD, which is pirate CD I mean. [ICE-HK S1B-029 #101]

97. money does not make us happier it'll only **make** us more difficult to be happier [ICE-SIN S1A-074#123]

These constructions seem to be parallel to alternate constructions with an extraposed Direct Object. Example 94 might be expressed in standard British English as follows:

98. We think that this language barrier **makes** it difficult for us to communicate.

This construction represents a grammatical difference from standard British English but not a semantic one – there is no evidence that *make* here represents a different meaning from the Predicative Complement usage established in 6.1. While it is beyond the scope of the present work to further investigate this

construction, its acceptability or pervasiveness as a linguistic norm, or its origins, it is the most common of several low-frequency *make* constructions observed in the corpora during the course of this study, and it deserves additional attention in future research.

Numerous compound derivatives from *make* occur in the corpora, and most are nouns (e.g. *decision-making*, *lawmakers*, *filmmaker*). One compound verb occurs that is unique to ICE-HK: *fashion-make*.

99. The final jurisdiction in mediation is incumbent upon the parties themselves, and they may **fashion-make** a particular form of solution for themselves under no influence of precedent cases in common law. [ICE-HK W2A-014 #99]

The instance of *fashion-make* is particularly noteworthy given that it appears in academic writing, which can be expected to be edited and vetted for adherence to standard usage guidelines. That said, it may also be seen as a deliberate neologism indicative of a sophisticated tone.

One fascinating piece of written language from Hong Kong exemplifies identity evidence between the Predicative Complement usage and the Causative usage. Neither of those usages is investigated in depth here, but the example is noteworthy.

100. Medicine in this aspect may be really helpful because the effect of the medication has **made** the hyperkinetic child dull and feel drowsy. [ICE-HK W1A-012 #72]

In Example 100, *make* is followed by a Direct Object, *hyperkinetic child*, and then both two Complements, the adjective *dull* and the clause *feel drowsy*. This type of coordination does not occur elsewhere in any of the three corpora. The construction is from a student essay, and it may be that it represents an error or an individual idiosyncrasy rather than an acceptable norm – more research

would be necessary to reach a final conclusion. Nevertheless, this is potential evidence for a unique polysemy for *make* in use in Hong Kong, such that the causative sense of *make* shares some identity with the predicative complement use of *make*.

6.2.12. Make: Summary

The most striking finding in the analysis presented throughout this chapter is the significant similarity in meaning and use of *make* across the three regions. No novel senses for *make* appear in the corpora, and examples generally fit neatly into the same sense categories in each corpus. Four of the five hypothetical senses forwarded for *make* seem to apply for all three regions; the fifth sense, 'Produce (Abstract)', was not found to be viable in any of the three corpora. This is in stark contrast to *get*, for example, which occurs in ICE-SIN, and in Singapore English more broadly, with many senses unique to Singapore (see 4.3); this is also in contrast to the semantic variation hypothesized in literature on lexicography and World Englishes (see 4.3).

Similarities across the corpora include not only the consistent applicability of the same sense categories for the three regions, but also many similarities in onomasiological preferences. In particular, Delexical usage universally reflects a strong preference against the Delexical construction in favour of its related verb. If this finding is considered in relation to Geeraerts's (2010) hypothesis of onomasiological salience, then we might expect Delexical constructions to exhibit low cognitive salience; Gilquin (2008) found just that. For other senses of *make*, spoken usage seems to be generally uniform across the corpora. That speech would vary less than writing between regions is a surprising finding, given the general assumption that innovation should first occur in speech, and also given the assumption that a sort of common core exists in written English worldwide.

Significant variation does occur in writing across the corpora. When differences do arise, ICE-HK seems to be the unusual or unique dataset. All corpora prefer concrete *make* over concrete *produce* in speech, while only ICE-

HK prefers concrete *make* over concrete *produce* in writing; ICE-SIN and ICE-GB, on the other hand, prefer the Latinate, polysllabic *produce* over *make* in writing. That preference perhaps renders even more interesting the unique preferences for *make* and *prepare* exhibited in ICE-HK with Direct Objects representing food: ICE-HK prefers *prepare* with the sense 'Produce (Concrete, Food)' in speech. ICE-HK is also the only corpus that includes *prepare* at all with a concrete, non-food Direct Object. The unique preference for *prepare* and against *produce* are similar insofar as they both involve a Latinate alternate in relation to a monosyllabic Germanic alternate in speech and writing. That is, ICE-HK seems to exhibit its own unique genre norms.

ICE-HK also includes unique onomasiological alternates for the construction *make an effort*, including *put an effort*, *give an effort*, and *pay an effort*, while both ICE-HK and ICE-SIN include the alternate *spend an effort*. All of those alternates are significantly less common than *make an effort*, and none is attested in ICE-GB.

Some other variation in the corpora has not been reported in existing literature, and is worth noting, even though it is not the focus of the present study, including the *make me difficult* construction in ICE-HK and ICE-SIN, and *make a difference* with the meaning 'differentiate' in ICE-HK and ICE-SIN. Both phenomena seem to evidence creative flexibility or innovation in ICE-HK and ICE-SIN in contrast to ICE-GB.

7. TAKE

7.1. Hypothesis

Take is slightly less common than *make* in use *pmw*, as evidenced in the BNC, ICE-SIN, ICE-HK, and ICE-GB. Like *make*, *take* is a highly polysemic verb with both concrete and light senses, so it provides a good opportunity for studying existing ideas about light verbs and corpus frequencies (see 3.4). The high number of occurrences of *take* in all three corpora, as well as the high degree of polysemy of *take*, facilitates a high degree of nuance in semantic analysis over a relatively large amount of data. As with *make*, a working hypothesis for the semantics of *take* was established by consulting the OED, the Collins COBUILD Dictionary, and academic work on the semantics of *take*. Lexicographical and academic work on *take* is summarized here first, and then a hypothesis is presented. Corpus evidence appears in 7.2.

The OED divides *take* into 9 major senses and 63 sub-senses. Collins COBUILD lists only two senses: the first with nouns representing actions (including light constructions), and the second an aggregate of 'other uses'. To begin with the light construction: both dictionaries include a sense in which the Direct Object has a related verb form, and the entire predicate (*take* + Direct Object) is equivalent in meaning to that related verb form. This relationship is explicit in the OED (in sense VIII), and indicative of the sense, while Collins COBUILD explains that the meaning of the expression is carried by the Direct Object, without explaining the grammatical correlation with a related verb. In addition, Collins COBUILD lists some apparently light constructions, including *take care*, under the entries for the particular Direct Object, e.g. *care*, further affirming the stance that it is the Direct Object which carries the significant meaning in the construction.

> 101.   The salmon **took** a great leap. (Kennard 1889, quoted in
>   OED, *take*, v., 52a)
> 102.   She **took** a shower. (Collins 2012: *take*, v., 1)

In Example 101, *take a leap* is equivalent to *leap* (v.), and in Example 102, *take a shower* is equivalent to *shower* (v.). This sense represents the light verb usage (Jespersen 1954: 117), or the 'Delexical sense' described by Gilquin (2008) and Mukherjee *et al*. (2012), as discussed in 3.4 and in relation to *make* in Chapter 4.

The OED's first sense of *take* is a concrete sense, in which a concrete Direct Object referent is transferred to an agent or Subject referent by a physical act. Collins COBUILD lists examples of this sense at the beginning of its aggregated 'other uses'.

> 103.   He could **take** his hat and go. (Hook 1833, quoted in OED, *take*, v., 12a)
> 104.   Here, let me **take** your coat. (Collins 2012: *take*, v., 2)

The concrete Direct Object referent can be a person or people as well.

> 105.   The school bus **takes** them to school and brings them back. (Collins 2012: *take*, v., 2)
> 106.   Being obliged to **take** four of us in his carriage to wait upon his majesty. (Thackeray 1848, quoted in OED, *take*, v., 57a)

The OED also describes a sense with a concrete Direct Object in which the Direct Object is not physically transferred to the Subject, but is instead attributed or assigned to the subject (including attribution of ownership). Collins COBUILD does not clearly distinguish such a meaning.

> 107.   The undertakers… had power to **take** lands compulsorily. (Law Times Rep. 1883, quoted in OED, *take*, v., 15a)

The OED and Collins both also identify uses of *take* with an abstract Direct Object, such that the abstract Direct Object is assigned or attributed to the Subject, or the Subject 'adopts' the thing represented by the abstract Direct Object. *Adopt* and *assume* are given as synonyms in both the OED and Collins COBUILD. The OED lists distinct sub-senses with the meaning 'Adopt/Assume' (16 through 21), while Collins COBUILD subsumes 'Adopt/Assume' under the 'action' sense, alongside light constructions.

> 108.   Captain Mayer… was compelled by circumstances to **take** the responsibility. (Speaker 1892, quoted in OED, *take*, v., 17a)
>
> 109.   I felt it was important for women to join and **take** a leading role. (Collins 2012: *take*, v., 1)

In OED sense VII, *take* relates to mental comprehension, and Collins COBUILD similarly lists numerous examples suggesting mental comprehension.

> 110.   An audience… quick to **take** his points. (*National Observer* 1893, quoted in OED, *take*, v.)
>
> 111.   Unfortunately, no one **took** my messages seriously. (Collins 2012: *take*, v., 2)
>
> 112.   **Take** this office, for example. (Collins 2012: *take*, v., 2)

Example 112 is an imperative clause in which *take* could alternate with *consider*. This imperative type of construction is the most common construction with this sense in the three corpora considered here (see 7.2.5).

Norvig and Lakoff (1987: 9-14) describe 7 senses of *take*.

I. Grab - 'The baby took the toy from its mother.'

II. Take Patient to Recipient - 'The messenger took the book to Mary.'

III. Take Patient to Destination - 'I took the book home.'

IV. Take action at Patient – 'I took a punch at him.'

V. Take action from Agent - 'I took a punch from him.'

VI. Going-to-Destination - 'Max took Sadie to the theatre.'

VII. Take as *perceive* – 'Take a glance at…'

Norvig and Lakoff (*ibid*.) relate each successive sense back to sense I in a systematic way, defining exactly the features of each sense as they vary from sense I. Senses are also defined in terms of Agents, Patients, and Recipients, as well as Sources and Instruments, and Origins and Destinations. Like Brugman and Lakoff (1988; see 3.3), the aim is not to confirm the distinctness of these senses, or even to explore the degrees of distinctness between these senses, but to explore the semantic relations between each sense as a variation of sense I. Norvig and Lakoff's (1987) sense distinctions can be seen to fit within the present hypothesis below, though the present hypothesis differs from Norvig and Lakoff's categories, largely in order to be more comparable to the hypothesis and findings for *make*.

Levin (1993) describes the semantics and syntax, and the relations between the two, of an array of verb classes. She assigns four different senses of *take* to four different categories of verbs (see 6.1 for further discussion of Levin, 1993):

I. 'Verbs of continuous causation of accompanied motion in a deictically specified direction';

II. 'Steal' verbs;

III. 'Characterize' verbs as in 'Angela characterized Shelly as a lifesaver', alternating with *regard* and *treat*;

IV. 'Fit verbs', alternating with *fit*, *hold*, and *contain*.

Levin (*ibid*.) does not attempt to comprehensively describe the semantics and syntax of *take*, but her synthesis of semantics and syntax, and her complete consideration of an extremely wide array of published studies, are useful. Levin's (*ibid.*) sense III does not in fact occur in the corpora.

Newman (1996) briefly discusses the meaning of *take* and various similar verbs in other languages in relation to his thorough and useful analysis of *give* (see 8.1 for a discussion of Newman, 1996, on *give*). He (*ibid*.: 58) argues that *take* is a natural semantic converse of *give*. Newman's approach is 'encyclopedic' (*ibid*.: 37), examining not only the semantics of *give* and *take* in a narrow sense, but also contexts of use, as well as cultural and social rituals, from a Cognitive Linguistic perspective. That said, his (*ibid*.) discussion of *take* is not meant to be comprehensive, but instead to provide a brief complement to the more thorough discussion of *give*. Newman (*ibid*.: 58) identifies *take* as involving the movement of a thing, initiated by a person, such that the thing ends up in the 'sphere of control' of that person; the movement of the thing is towards the Subject referent. The concrete sense is identified as typical within a Cognitive Linguistics framework (*ibid*.: 2): because cognition derives from embodied experience, concrete senses are primary. In fact, Newman (*ibid*.: 2) notes that human beings often interpret abstract notions in terms of more concrete things: thus, we might see the concrete sense as prototypical by Rosch's (1973, 1975a, 1975b) criterion that non-prototypical items are interpreted in terms of prototypical ones. Even with concrete uses of *take*, Newman (1996: 58) claims that there is a clear 'energy flow' which originates with the taker, moves towards the thing, and then returns to the taker. It is unclear, however, how this energy flow might be tested or evidenced, and it is certainly not a scientific argument. Newman (*ibid*.) also notes the following examples of *take* and discusses their meaning:

I.   *Take* can occur with a Direct Object referring to a person, in which case *take* conveys not just movement but also accompaniment (*ibid*.: 243).

II.   *Take advice*, according to Newman (*ibid*.: 244), involves not just receiving advice or listening to advice, but also accepting and acting on advice.

III.   *Take* can relate to 'assuming control' as in *take responsibility* and *take charge* (*ibid*.: 244).

IV.   *Take* can represent an abstract energy flow to a person from an indeterminate source, as in *take a beating*, *take a punch*, *take a knock* (*ibid*.: 245)

I would argue that the sense 'accompaniment' is a reasonable interpretation of *taking someone somewhere*, but the specific meaning 'accompaniment' seems to depend on the nature of the Direct Object as a person. My hypothesis takes this issue into account, and I discuss it further in relation to my findings in 7.2.2. I discuss the implications of *taking advice* further in 7.2.5, as this issue also arises in the corpora observed here. Newman's sense III is central to my own hypothesis. Sense IV is deemed too vague to facilitate useful empirical methods, and is not considered further in the present study. It is perhaps worth noting that Newman's sense IV is extremely rare in use in all three corpora.

Gilquin (2008) identifies seventeen senses of *take*; she states that these are based on the sense categories in five different learner's dictionaries, but the details of the learner's dictionaries are not explained or discussed. Just as the present study refers to lexicographical research as a starting point in designing a hypothesis, Gilquin's reference to learner's dictionaries is a reasonably strong approach. Her study asks whether the salience of a word's various senses in the mind, as evinced by elicitation tests, corresponds to the semasiological frequency *per million words* of that word's various senses, as evinced by corpus analysis. I discuss this study and its findings in much greater depth in Chapter 9. Gilquin's (*ibid*.) semantic classifications are as follows:

I. Grab – 'Let me take your coats.'

II. Move – 'Our guide took us around the cathedral.'

III. Buy – 'What newspaper do you take?'

IV. Ingest - 'Take two aspirins and go to bed.'

V. Require – 'It takes strength and stamina to be a long-distance runner.'

VI. Do (Delexical sense) – 'Let's take a walk down the river.'

VII. Record – 'A nurse took his temperature every hour.'

VIII. Engage in – 'Shelley is taking economics at university.'

IX. Consider – 'She took his remarks as a compliment.'

X. Accept – 'Do they take credit cards in this shop?'

XI. Assume – 'I did all the work, but Gill took all the credit.'

XII. Experience – 'The school took the full force of the explosion.'

XIII. Use – 'It's more interesting to take the coast road.'

XIV. Capture – 'The rebels succeeded in taking the town.'

XV. Work – 'If the cortisone doesn't take, I may have to have surgery.'

XVI. Idioms – 'The Olympics take place every four years.'

XVII. Phrasal Verbs – 'The plane should take off on time.'

XVIII. Other

<div align="center">(Gilquin 2008: 244)</div>

The senses listed here are quite numerous, and as a semantic hypothesis, this outline seems relatively difficult to manage. I avoid such overproliferation of senses in my own hypothesis below. It may be that Gilquin's sense III involves the same constant meaning as sense I, but different contingent meaning. Similarly, sense IX might be seen as adding contingent meaning to sense XI, both of which take abstract Direct Objects that are received by the Subject or agent. Gilquin's sense I and II raise the interesting question whether a Direct Object that is an inanimate object (as in *take your coats*) suggests a different sense of *take* than a Direct Object that is a person (as in *take us around the cathedral*); Newman (1996) identified them as different senses, but I am less convinced (see above and 7.2.2). In fact, most of the senses above occur in the

three corpora examined here (the exception being sense XV), but most of those are relatively infrequent.

Werner and Mukherjee (2012) build on Gilquin's (2008) study, using the same semantic classifications for a sort of replica semasiological study of the semantics of *take* in ICE-GB, ICE-India, and ICE-Sri Lanka. They do not elaborate on the semantic categories themselves, however: their categories do not raise any new issues not raised by Gilquin (2008). I discuss Werner and Mukerjee's (2012) work at greater length in 9.3, in relation to the impact of the present findings for Cognitive Linguistics research.

Informed loosely by the academic studies described above, and building in particular on the hypotheses and evidence for *make* presented in chapter 6, the present study proposes five hypothetical sense categories for *take*. These sense categories allow for comparison with the study of *make* already conducted and presented in chapter 6; they are suggested in particular by both the OED and COBUILD dictionaries; they are reflected to some degree by the academic studies discussed above; even more importantly, they avoid the problematic issues identified in the academic studies discussed above; and they are designed to balance generality and specificity. As with *make*, these senses may overlap, and the issues in that overlap are discussed further in relation to corpus evidence in 7.2. Also as with *make*, this hypothesis formation must be considered a pre-scientific, reflective process, which can then be tested and anlaysed more rigorously in the scientific process that follows in the corpus study.

I. Transfer (Concrete) - Senses in which the concrete Direct Object referent is transferred to or received by the Subject referent or agent

II. Transfer (Abstract) – Senses in which the abstract Direct Object referent is transferred to or received by the Subject referent or agent

III. Adopt/Assume – Senses in which the Subject referent or agent adopts or assumes a Direct Object referent; put differently, the Direct Object referent is attributed to the Subject referent or agent

IV.   Consider/Apprehend – Senses in which the Direct Object referent is considered or apprehended by the Subject referent or agent

V.   Delexical - Senses in which the Direct Object referent has a related verb, and the *take* construction is roughly equivalent in meaning to that related verb

In a data-driven way, as with *make*, the present study proceeded to address this semantic hypothesis by attempting to categorize instances of *take* into these five sense categories, and determining additional distinctions that needed to be added. In addition to the three categories listed above, other senses of *take*, including phrasal verb usage and other constructions, were also recorded, and that data is available for future research.

Section 6.1 provides a detailed description of the establishment of the parallel hypothetical sense categories for *make*, and much of that discussion applies here as well. Moreover, because I have hypothesized two abstract and two concrete senses for *take* (unlike the single abstract and single concrete sense for *make*), the validity of that division must be tested as well. The use of antagonism evidence, autonomy evidence, identity evidence, and truth-condition evidence was used in the analysis of *take*, as with *make*, and section 6.1 describes that approach in detail. The analysis in the following section proceeds to ask what sense distinctions can be established for *take*, and whether those sense distinctions vary between the three corpora.

7.2. Findings

7.2.1. Take: The lemma

A total of 5,477 instances of *take* occur in the three corpora combined, and each instance was manually analysed with the aim of determining whether variation exists in frequencies and usage of each sense and its semantic alternates among the three corpora. First, the instances of *take* in the corpora were examined to determine whether they affirmed or refuted the hypotheses regarding the sense distinctions of *take* (see 7.1); and by looking for particular types of evidence

derived from polysemy tests (see 5.3). Then, an onomasiological analysis was performed, in which alternates for each sense were identified and preferences for alternates were analysed in each corpus.

The lemma *take*, in the forms *take*, *takes*, *taking*, *took* is quite common in use, and the corpora provide evidence that an English listener/reader can expect to encounter *take* quite often, depending on text topic and genre, and, perhaps, regional variety. Total instances of *take* in each corpus appear in Table 9.

	Instances of *take*
ICE-Singapore	1,807
ICE-HK	2,010
ICE-GB	1,660

Table 9: Instances of *take* in ICE-Singapore, ICE-HK, and ICE-GB

Semasiological data on the frequencies of each of the five hypothesized senses are presented in Figure 23 and Figure 24. This data answers the question: given that a listener or reader is exposed to *take*, what is the probability of encountering each sense as a proportion of the total number of occurrences of *take*? This might be useful for a lexicographer arranging the order of entries in a dictionary like the Collins COBUILD dictionary, or a learner's dictionary designed for ease of use.
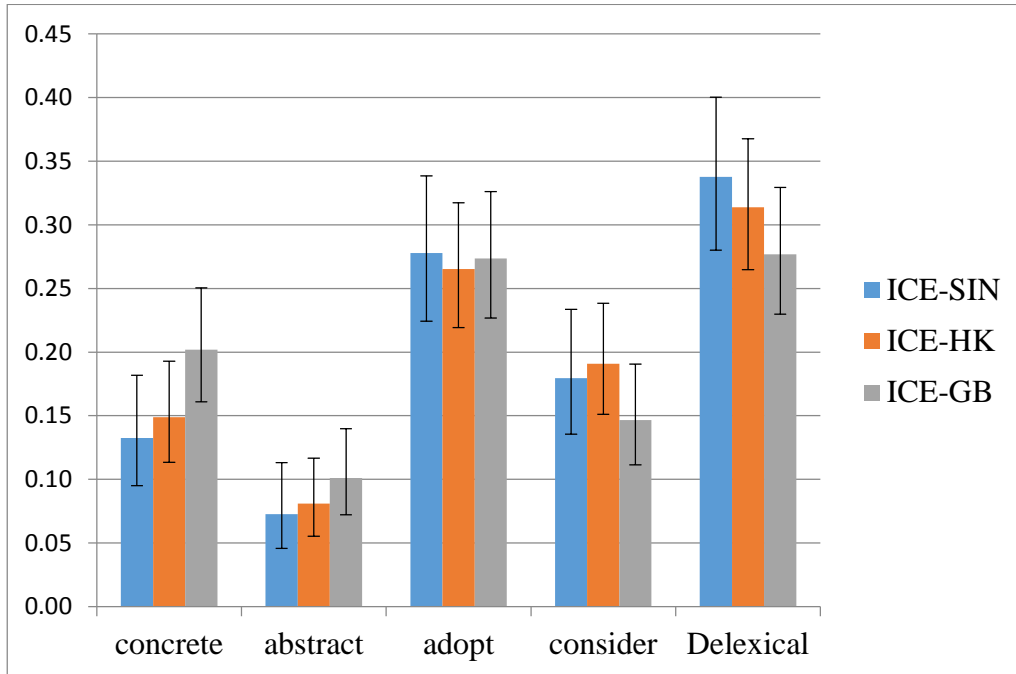
Figure 23: Probabilities of encountering *take* with each of the five hypothesized senses in the written portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
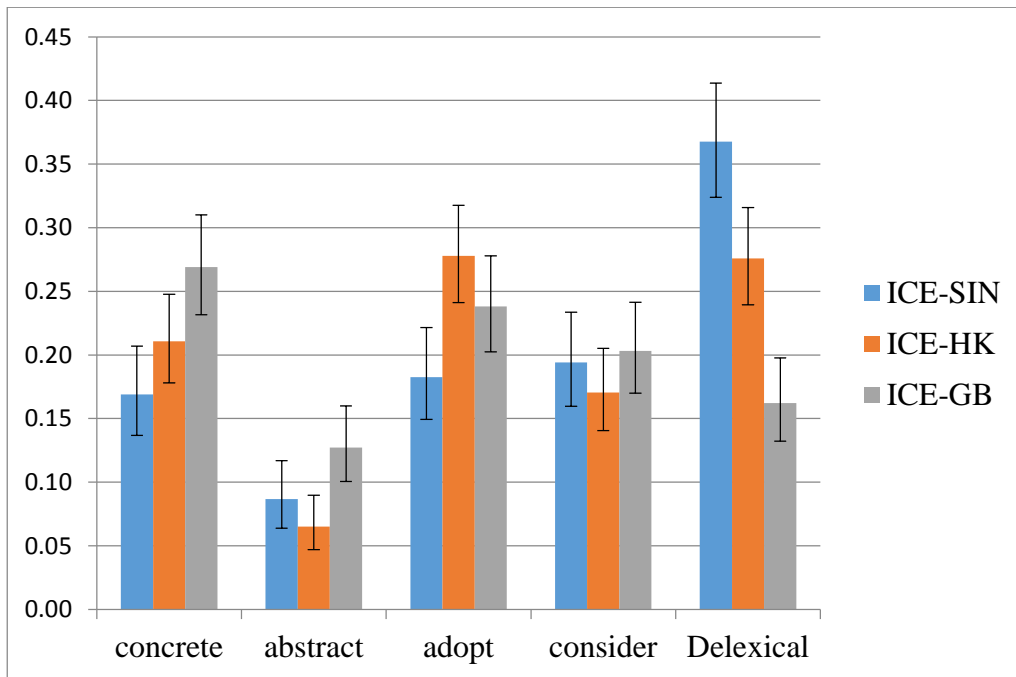
Figure 24: Probabilities of encountering *take* with each of the five hypothesized senses in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

It is particularly noteworthy that Delexical senses are not the most frequent senses in speech in ICE-GB – Delexical usage is in fact significantly less frequent than concrete usage. This refutes the established knowledge that Delexical senses are more common than concrete senses (*cf.* Gilquin 2008, Collins COBUILD 2013), but it echoes Werner and Mukherjee's (2012) findings for *take* in a sub-section of ICE-GB. It is clear from the semasiological graphs that exposure rates to Delexical constructions are high, but so are exposure rates to the concrete sense and the sense 'Adopt'. I would hypothesize that this unique feature of ICE-GB does not represent a linguistic norm, but is likely to reflect influences from text topic and real-world context. The onomasiological analysis of these senses later in this chapter refines the probabilistic analysis by asking how often language users speak or write *take* with each sense, given the opportunity.

7.2.2. Take: Transfer (Concrete)

Many of the most clear and certain instances of *take* represent the Concrete sense (also referred to here as 'concrete *take*'). Very few ambiguous cases for concrete *take* arise in the corpora. Clear examples of concrete *take* include the following:

> 113.   If you want to **take** the vehicle abroad within 3 weeks of registration please ask the dealer to inform the VRO when he registers the vehicle. [ICE-GB W2D-010 #12]
> 114.   Wen Ei passed her the Chinese pen and **took** the book. [ICE-SIN W2F-008 #136]

Example 113 describes the transfer of a concrete Direct Object, *vehicle*, from one place to another. Example 114 describes the transfer of a concrete object, a book, from one person to another. In both cases, there is no assignation of ownership, and *take* is not replaceable with *adopt* or *assume*, as in hypothetical sense III (see 7.2.4). Whether the example presents a transfer between places or between people is derived from the surrounding context and can reasonably be seen as contingent meaning. *Take* commonly refers to the transfer of a person from one place to another as well.

115.   Two men and six women aged between 25 and 70 were **taken** to the United Christian Hospital… [ICE-HK W2C-013 #53]

116.   Why would they have **taken** me to a hall somewhere? [ICE-GB S1b-064 #145]

117.   Julie and I **took** Emily and her two boys, who are aged 5 and nearly 3, swimming this morning. [ICE-GB W1B-008 #34]

Direct Objects representing people, in which the person, as a body, is transferred from one place to another, are identified here as instances of concrete *take*, in contrast to Newman (1996) and Gilquin (2008).[36] There is one instance of identity evidence in the corpora in which *take* coordinates a concrete inanimate Direct Object with a Direct Object representing people.

118.   In the summer you can **take** a car and four people [on the ferry] for a hundred and twenty pounds. [ICE-GB S1A-021 #105]

---

[36] The *Oxford Adanced Learner's Dictionary* (New Edition, 1989, *take* (v.), 1), like the present study, combines the transfer of people or things into a single sense category. Further lexicographical investigation, however, is not the aim of the present work.

Example 118 suggests that only one meaning of *take* is at stake when describe the transfer of people and inanimate objects. One instance of such identity evidence is not conclusive proof of the identity of the senses 'Transfer (Concrete, Person)' and 'Transfer (Concrete, Non-Person)'. Nevertheless, no evidence arises that suggests that these two senses must be separated, and it seems justifiable that the movement of a human being from one place to another should qualify as 'Transfer (Concrete)'.

7.2.3. Take: Transfer (Abstract)

Examples of abstract *take* are shown in Examples 119, 120, and 121.

119.   …it **took** the theory in a direction that Firth perhaps may not have fully expected… [ICE-GB S2B-049 #32]

120.   It's **taking** us we hope towards a better and more honest and more humane system of divorce law. [ICE-GB S2B-019 #68]

121.   …**take** this process an important step forward. [ICE-HK S2B-032 #86]

The Direct Objects in the examples above are abstract nouns. In Example 119, the abstract Direct Object *theory* is moved or transferred figuratively from a starting point to an end point. In Example 120, *us* refers not to concrete bodies of people moving from one concrete place to another (as in Examples 115 and 116), but to abstractions moving figuratively from one system to another. In Example 121, the abstract Direct Object *process*, not directly perceivable by any of the five senses, is transferred metaphorically towards some future state. In none of these instances can the Direct Object be interpreted as concrete. Likewise, none of these instances can be categorized as either of the other proposed abstract senses: 'Adopt/Assume' or 'Consider'. Nonetheless, the metaphorical 'transfer' reading is slippery, and it is necessary to ask whether this interpretation is the one with the most integrity. It is at least valid and

defensible from the present evidence, and is investigated further later in this chapter.

7.2.4. Take: Adopt/Assume

The sense 'Adopt/Assume' occurs quite frequently in all three corpora. The vast majority of instances occur with abstract Direct Objects. Clear examples include the following:

> 122.   …the neurosurgeon in charge **takes** the primary responsibility. [ICE-HK W2A-034 #24]
>
> 123.   …I understand what motivates one to **take** such drastic measures to ease inner pain. [ICE-HK W1A-006 #233]
>
> 124.   She **took** an independent line on feminism. [ICE-GB W2F-009 #143]
>
> 125.   She should **take** her mother's side. [ICE-GB W2F-011 #113]

In each example above, the Direct Object is an abstract noun, and *take* indicates that the Subject adopts or assumes the Direct Object, or that the Direct Object is attributed or assigned to the Subject. In Example 123, *take measures* is not equivalent to *measure*, and is therefore not Delexical.

Instances of the sense 'Adopt/Assume' with a concrete Direct Object are quite rare.

	Instances of *take* with the sense 'Adopt/Assume (Abstract)'	Instances of *take* with the sense 'Adopt/Assume (Concrete)'
ICE-SIN	130	15
ICE-HK	219	8
ICE-GB	189	11

Table 10: Instances of *take* with the sense Attribute (Abstract) and Attribute (Concrete) in ICE-SIN, ICE-HK, and ICE-GB.

Instances of *take* with the sense 'Adopt/Assume (Concrete)' generally relate to the attribution of ownership over concrete items that are not physically transferrable, such as land.

> 126.   He **takes** that land back for himself. [ICE-SIN S1B-006 #18]

In Example 126, an interpretation of a physical transfer from person to person, or place to place, is impossible; the interpretation here is generally of attributing the land, for example, to the Subject, with some kind of legal ownership often either implicit or explicit in context. There is no antagonism evidence, autonomy evidence, identity evidence, or truth-condition evidence to corroborate identity of the sense 'Adopt/Assume' across abstract and concrete Direct Objects; for example, there are no instances of this sense with coordinated Direct Objects in which one Direct Object is concrete, the other abstract.

7.2.5. Take: Consider

Take often occurs with the sense 'Consider'.

> 127.   **Take** an example: people without a voice box who cannot speak. [ICE-HK S1B-025#36]

128. **Take** five hundred years ago: they talk about
Indianisation of this region. [ICE-SIN S1A-073 #118]

129. **Take** for example, in the Three Little Pigs, the wolf
would have a gruff voice… [ICE-SIN W2D-020 #138]

Some of the clearest uses of *take* with this sense are in imperative clauses like Examples 127 through 129. Example 127 can be glossed as a command: 'Consider an example'; Example 128 as 'Consider five hundred years ago'; and Example 129 as 'Consider for example, in the Three Little Pigs…'. Uses of *take* with this sense can include additional modifiers and adjuncts that may sometimes occur quite regularly and, while grammatically omissible, are necessary for the given meaning to be conveyed.

130. Such efforts were **taken** for granted. [ICE-GB W2B-006 #102]

In addition to *take for granted*, other common constructions in all corpora include *take lightly* and *take seriously*, *take into account* and *take into consideration*. The pragmatics of some constructions may imply not only the sense 'Consider', but also the fact that the Subject referent accepts as true that which is being considered. This implication was noted by Newman (1996).

131. Have you ever **taken** any medical advice about this
before? [ICE-SIN S1A-051 #331]

Example 131 likely suggests not just the question 'have you considered any advice?' but also 'have you followed any advice?'. This possible pragmatic implicature seems to arise occasionally and could be expected to vary from one region to another, but there is little conclusive evidence in the corpus for or against the implicature. Native speaker input, and probably psycholinguistic testing, would likely be necessary to investigate such variation further.

7.2.6. Take: Delexical

As with Delexical *make*, many cases of Delexical *take* are quite straightforward.

132.   …and a waiter who just the minute you **took** a sip out of the glass replenished it. [ICE-GB S1A-011 #258]

133.   Winston Lee , pleased to meetcha an **take** a lookee rite heer at this lil beauty… [ICE-HK W2F-018 #119]

134.   …when he behaves so abnormally, we **take** notice. [ICE-SIN W1B-009#52]

In Example 132, *take a sip* is equivalent to *sip* (v.); in Example 133, *take a lookee* (or *take a look*) is equivalent to *look* (v.); and in Example 134, *take notice* is equivalent to *notice* (v.).

As with Delexical *make* (see 6.2.4), Delexical *take* occurs in passive constructions or with modifiers that might affect how readily it is interpreted as a Delexical construction equivalent to a related verb.

135.   Make the most of these years to **take** plenty of regular vigorous exercise. [ICE-GB W2B-022 #63]

In Example 135, *take exercise* is equivalent to *exercise* (v.), but the modifiers *regular*
and *vigorous* (and, arguably, *plenty of*) in the Delexical construction might affect how readily that equivalency is perceived. The psycholinguistic side of this question is potentially fruitful ground for future experiments. Example 136 includes modifiers, and is also a passive construction.

136.   It stressed that firm action will be **taken** against those who behave in a disorderly manner… [ICE-SIN S2B-001 #37]

In Example 136, *take action* is equivalent to *act*, but again the modifier *firm* and the passivization of the clause might affect how readily that equivalency is perceived. As with *make*, the present study identifies Delexical constructions by the existence of a related verb with equivalent meaning, regardless of whether the Delexical construction itself is passive or active, modified or unmodified (see also 6.2.4). Those additional grammatical variables constitute important future research questions, but the present study focuses on, and isolates, the semantic variable, rather than those grammatical ones.

Sometimes, the related verb of a Delexical construction might be expected to require a Direct Object that is not expressed in the Delexical construction itself.

> 137.   …he insists on a narrow definition of an executive lead government and **take** an exclusionary approach towards the legislature… [ICE-HK S1B-039 #141]
>
> 138.   When Harunobu Inukai is a guest chef, he insists on making everything, unlike Vittorio Lucariello who **takes** a more laid-back approach. [ICE-HK W2D-011 #75]

Take an approach relates to the equivalent verb *approach*. In Example 137, *take an approach towards the legislature* is therefore equivalent to *approach the legislature* (though the modifier *exclusionary* may complicate this equivalency, as discussed above). In Example 138, *take an approach* is again equivalent to the related verb *approach*, but the Delexical construction does not suggest a Direct Object for the related verb. In the present study, such instances are considered Delexical constructions, despite the troubled question of grammatical equivalency, as discussed in 6.2.4. This is in part because isolating the variable of semantics allows us to focus solely on semantic equivalency. In addition, transitivity can vary significantly from region to region, and many verbs that are necessarily transitive in standard British English can be intransitive in Singapore and Hong Kong (including *take* and *make*). The choice

to express a Direct Object or not in these constructions may also be part of the selection process, but that additional variable in the selection process, a more specifically grammatical one which would seem to be one among many relevant additional variables, is not the key variable under investigation in the present study. Instead, this study looks at the broader arc of the selection process among onomasiological alternates, conceived as filling a given semantic slot in expressing a particular meaning.

Some Direct Objects in Delexical *take* constructions relate to obsolete verb forms, as in Example 139.

139.   The second missile attack on Israel came after a night of false alarms during which the population had donned gas masks and **taken** refuge in sealed rooms three times. [ICE-GB S2B-015 #101]

According to the OED, *take refuge* is equivalent to the obsolete verb *refuge*. It is conceivable that a verb which is obsolete according to the OED's evidence is not obsolete in all varieties of World English. Taking evidence from the corpus, in the present study, if the OED claims that the related verb is obsolete in British English and if the related verb is not evidenced in any of the corpora, it is considered obsolete in general, and therefore not categorized as Delexical but as abstract. In future research, a larger corpus could further corroborate the obsolescence of such verbs in each region.

Delexical *take* may overlap with delexical *make*, as discussed in 6.2.4. The coordination of *take* and *make* in Example 76 (reprinted below as Example 140) might indicate a degree of identity between the two Delexical constructions, though the evidence is extremely limited and ultimately inconclusive.

140.   The staff member shall not: 1) Take or permit to be **made** any alterations in the internal construction or

arrangements or in the external appearance or in the present scheme of decoration of the premises. [ICE-SIN W2D-003 #130-1]

In addition, *take a decision* alternates in the corpora not only with *decide*, but also with *make a decision*, again suggesting some kind of equivalency between the two Delexical verbs.

Of the many interesting and not easily categorized instances of Delexical *take*, *take care*, serves as a useful illustration of the nuance required in manual semantic analysis.

141. My mother sort of **took** care of him. [ICE-SIN S1A-048 #103]

142. I expect to receive as many letters as I did in paris! **Take** care and lots of love, Anne Marie. [ICE-GB W1B-002 #158-160]

143. And then we add the twelve percent handling charge as appears in McHeaver 's invoices. I suppose that would **take** care of trucking… [ICE-HK S1B-062 #160-61]

In Example 141, *take care (of)* is equivalent to *care (for)* and is identified as Delexical. Both constructions require a preposition, but a different preposition in each case, which might be seen as complicating the equivalency. Nonetheless, the parallel seems viable, particularly given that numerous other Delexical constructions involve the addition or substitution of prepositions (e.g. *take a sip of water* is equivalent to *sip water*). In Example 142, *take care* is a construction used to sign off a letter, and is not replaceable with *care* in any of the corpus evidence or in standard British English. If one of the ICE corpora had contained instances of writers signing off letters simply with *care*, that might have been fascinating evidence of creative backformation in a region, but such evidence was not found. Instances like Example 142 are particularly

numerous in the sampled personal letters in the ICE corpora; these instances have been categorized separately from the hypothesized senses, as a unique construction and usage. In Example 143, *take care (of)* is not equivalent to *care* or *care for*: there is no corpus evidence to support such an equivalency. In this example *take care of trucking* cannot be rephrased as *care for trucking*. Instead, *take care (of)* here conveys, roughly, the meaning 'deal with' without the meaning of the related verb *care*. Example 143 is also categorized as a unique construction and semantic category. It is only a fraction of all instances of *take care*, therefore, that are in fact Delexical. The three semantic categories necessary for the single construction *take care (of)* illustrate the necessity of manual analysis, and also the difficulties that arise.

7.2.7. Take: Transfer (Concrete)/Transfer (Abstract) ambiguities

There are no instances of ambiguity in any of the corpora for these two senses. This is in contrast to *make*, which exhibited a small number of ambiguous cases. These two senses seem to be very discrete.

7.2.8. Take: Transfer (Concrete)/Adopt ambiguities

The overlap between these two senses is readily conceivable, and it might seem quite reasonable that *take* describes the transfer of a physical object and the attribution of ownership over that object, whether through donation, purchase, or theft (indeed, Levin, 1993, identifies *take* as a 'steal' verb; see 7.1). Interestingly, then, the combination of these two senses occurs only very rarely in actual use.

> 144.   I don't mean just wartime loot **taken** from the Chinese
> but from all peoples. [ICE-HK S2B-040#59]

In Example 144, both 'Transfer (Concrete)' and 'Adopt' are invoked, insofar as 'loot' is physically transferred and ownership is re-assigned. This is conveyed via the meaning of the Direct Object, *loot*, which is stolen during war. In most

instances, 'Transfer (Concrete)' clearly describes a transfer without ownership, as with simply handing or passing an object from one person to another.

	Instances of *take* with the sense 'Transfer (Concrete)'	Instances of *take* with the sense 'Transfer (Concrete) + Adopt/Assume (Concrete)'	Instances of *take* with the sense 'Adopt/Assume (Concrete)'
ICE-SIN	105	0	15
ICE-HK	159	11	8
ICE-GB	193	3	11

Table 11: Instances of *take* with the sense Transfer (Concrete); with the combined senses Transfer (Concrete) and Adopt (Assume Ownership); and with the sense Adopt (Assume Ownership, Concrete) in ICE-SIN, ICE-HK, and ICE-GB.

Table 11 displays just how rare this overlap of meanings is in the three corpora. The generally low frequency of the sense 'Adopt/Assume (Concrete)' has already been discussed (see 7.2.4). In use, the concrete transfer sense seems to be most often discrete from the concrete adopt sense. Table 11 also shows that the overlap is most common in ICE-HK, suggesting that language users in ICE-HK are more likely to be exposed to the sense 'Transfer (Concrete) + Adopt/Assume (Concrete)' than to the sense 'Adopt (Concrete)'. Although this is a measure of exposure rates rather than selection processes, it is noteworthy that, yet again, ICE-HK is the dataset that stands out. This could be relevant to local or World English lexicographers – users in Hong Kong may be more likely to encounter *take* with this combined sense, quite possibly interpretable as 'steal', than users in other regions. On the other hand, it may be that ICE-HK simply includes more discussions of stealing than the other corpora – this

particular measurement in Table 11 may be an epiphenomenon of text topic and real-world context rather than language norms.

7.2.9. Take: Transfer (Concrete)/Consider ambiguities

There are no instances of ambiguity or overlap between the two senses 'Transfer (Concrete)' and 'Consider'; the discreteness of these two senses seems valid from the present evidence.

7.2.10. Take: Transfer (Concrete)/Delexical ambiguities

With *make*, ambiguity between the concrete sense and the Delexical sense was relatively rare, but did occur. With *take*, such ambiguity is even rarer.

> 145.   So that is another difference that we have to **take** note.
>   [ICE-HK S1B-026 #41]
> 146.   He **took** away some notes and some extracts from their
>   accounts. [ICE-GB S2A-070 #025]

In Example 145, *take note* is equivalent to *note* (v.), which is quite distinct in use from 'transfer notes'. In Example 146, the modification with *away* as well as the coordination between *notes* and *extracts* all contribute to the discrete sense 'Transfer (Concrete)'. There is, similarly, no ambiguity in use with any of the most common Direct Objects of the Delexical *take* construction. There is one relatively unusual construction that seems to exemplify overlap between these senses.

> 147.   A blood sample can be **taken**. [W2A-023 #86]

As in Example 147, *take a sample* can be seen as equivalent to both *sample* (*v.*) and as an act of transferring a concrete thing, the sample, towards a Subject referent or Agent. Examples of *take samples* occur in all three corpora, though only rarely: once in ICE-SIN; 8 times in ICE-HK; and 6 times in ICE-GB. In

fact, they occur rarely enough that they do not seem to challenge the relative discreteness between these senses. In the analysis, nonetheless, these cases have been categorized separately, counted as neither concrete nor Delexical.

There is one useful example in the corpora that might be interpreted as suggesting discreteness between the Concrete and Delexical senses.

148.   Not only do most women in Britain from the age of about 50 onwards **take** far too little calcium, they also tend to **take** far too little exercise. [ICE-GB W2B-022 #22]

The OED categorizes the sense 'ingest' separately from other senses of concrete *take* (OED, *take* (v.) 13a), but it seems defensible to include the transfer of concrete substances into the body alongside other uses of concrete *take*, which I have done. (In any case, instances of *take* meaning 'ingest' are rare enough that they do not significantly alter the total numbers of concrete *take*; even if they are removed from the data, the significance measurements remain the same.) It might be that Example 148 is broken into two clauses so as to avoid coordinating *calcium* and *exercise*, because the result would be zeugmatic, as in Example 149.

149.   ?Most women in Britain from the age of about 50 onwards **take** far too little calcium and exercise.

It would be unwise to draw strong conclusions from this example, but it might be seen to imply the discreteness of these senses.

It would seem that Delexical *take* differs remarkably from Delexical *make*, insofar as Delexical *take* resists ambiguity or overlap with Concrete *take* more strongly than Delexical *make* resists ambiguity or overlap with Concrete *make*. The one instance of coordination between Delexical *take* and *make* suggests that Delexical *take* and *make* are in some cases substitutable, but

Delexical *take* seems to be more discrete from other senses of *take*. Sections 7.2.14 and 7.2.15 further corroborate this, and it is discussed in 9.1.

7.2.11. Take: Transfer (Abstract)/Adopt ambiguities

This ambiguity arises only rarely, in discussions of, for example, someone *taking responsibility* or *power* from someone else. It seems that there is a metaphorical transfer and also an attribution or assignation as a result. In the corpora, there are actually very few cases of this overlap. The ambiguity seems to be unresolvable, and these examples have been removed from both categories and categorized instead as ambiguous. They occur rarely enough that these two categories still seem to be relatively distinct in use.

7.2.12. Take: Transfer (Abstract)/Consider ambiguities

There is a unique instance of ambiguity between these two senses in the corpora.

> 150.   …to **take** the issue of translation and power further.
>   [ICE-HK W2A-009#11]

In Example 150, there is a sort of abstract metaphorical transfer, in *taking* an issue further, but it is also possible to gloss the construction as 'further consider the issue'. The ambiguity seems to depend in part on the complicated meaning of *further*, which can suggest abstract space on one hand, or 'more intensely' or 'more extensively' on the other hand. Given the presence of only this anomalous case, the distinctness in use of these two senses in the corpora seems to be firm.

7.2.13. Take: Transfer (Abstract)/Delexical ambiguities

No instances of ambiguity or overlap between these two senses occur in the corpora. This represents a remarkable contrast with Delexical and abstract *make*, for which a huge proportion of instances (most instances, in fact) could be glossed as either. Because of that issue with *make*, the present study opted

by default to categorize all ambiguous cases between abstract and Delexical *make* as Delexical. That issue does not arise with *take*, suggesting that the Delexical category can be realized in various words in various ways: the Delexical sense may be more or less discrete from other senses of a given word, such that varying light verbs show varying degrees of Delexicality. The possibility of degrees of Delexicality is discussed further in 9.1, and *give* provides even more evidence of this possibility (see 6.2.7).

7.2.14. Take: Adopt/Consider ambiguities

No ambiguities occur in any of the three corpora between these two senses in use. They seem to be, in practice, quite firmly distinct.

7.2.15. Take: Adopt/Delexical ambiguities

Some interesting issues in categorization arise between the Delexical sense and the sense 'Adopt/Assume'. In particular, the frequent descriptions of free kicks and throw-ins in running commentary of football matches is complicated. The senses seem to be discrete, but this example serves as an illustration of the care that must be taken in manual semantic analysis of these verbs. *Take a kick* and *take a throw* might be interpreted as *kick* (*v.*) and *throw* (*v.*), respectively, in some instances of use, but such simple instances in fact only rarely occur in any of the corpora. The constructions *take a free kick* and *take a throw in*, which are quite common, are complicated by the words *free* and *in* (respectively). Is *take a free kick* equivalent to *kick* (*v.*), *free-kick* (*v.*), or *kick free*? It seems that it is not. There are no examples of the latter two alternates in the corpora. In practice, what often happens is that the referee *gives* free kicks, and the players or teams then *take* free kicks, and only afterwards is the ball actually kicked. In most corpus examples, context suggests that the construction *take a free kick* is a report of attribution or assignation, not of the act of kicking itself. So, the player *takes the free kick* (accepting the assignation to perform the free kick) and only afterwards actually kicks the ball. Nonetheless, there are numerous cases where context does not disambiguate whether *take a free kick* describes

the assignation of the kick or the act of kicking. Moreover, although these examples often seem to represent the sense 'Adopt/Assume', there is no alternate in use for *take*; there are no examples in the corpus of, for example, *assuming a free kick*. Because *take* in *take a free kick* does not actually alternate with *adopt*, *assume*, or any other verb, the construction is is not counted in the onomasiological study of alternation patterns between *take* and *adopt*. Regardless, eliminating these instances from both categories does not effect measures of statistical significance due to their low frequency in relation to the totals for each category. This is an excellent illustration of the way that an onomasiological approach allows for the discovery of additional variables that restrict alternation. To reiterate, this approach does not in any way assume absolute alternation, but functions as a ground for testing alternation in practice (see 3.2).

One interesting example might be seen as suggesting distinctness between these two senses.

151. But this does not justify the United States and Britain **taking** the law into their own hands and **taking** military action to topple him because the leaders of these two countries do not like him. [ICE-HK W2E-002#53]

It is possible that Example 151 repeats *take* rather than coordinating two Direct Objects because coordinated Direct Objects would introduce zeugma.

152. ?But this does not justify the United States and Britain **taking** the law into their own hands and military action to topple him…

While this evidence might suggest the discreteness of these two senses, it is by no means conclusive. Nonetheless these two senses seem to be relatively distinct.

7.2.16. Take: Consider/Delexical ambiguities

No ambiguities arise between these two senses; they seem to be firmly distinct in use.

7.2.17. Onomasiological analysis: Transfer (Concrete)

The present study now proceeds to identify semantic alternates of concrete *take* and to evaluate the feasibility of the sense 'Transfer (Concrete)' onomasiologically. As with concrete *make* (see 6.2.8), in order to measure preferences for semantic alternates in use in the corpora, Direct Objects of concrete *take* were identified first. Then, Direct Objects that occur at least twice per corpus were identified, and all other verbs in each corpus which occur with the same Direct Objects, and which convey generally the same meaning, were subsequently catalogued. This count of alternates can be expected to provide a reasonably thorough, evidence-based onomasiological comparison for concrete *take*. Concrete *take* in the three corpora exhibits the alternates shown in Table 12.

Alternates for concrete *take*:	collect, carry, push, borrow, transport, seize, snatch, pull, heave, grab, haul, extract, transfer, withdraw, confiscate, drag

Table 12: Alternates for concrete *take*, as evidenced in ICE-SIN, ICE-HK, and ICE-GB

Like *make*, *take* is semantically a very general term, and its alternates tend to be more semantically specific. This specificity in turn means that not all of these lexical items can alternate for all others in all circumstances. Aggregating the data for all alternates therefore adds some noise to the calculations (see also 6.2.8 for a thorough discussion of this issue regarding *make*, and a description of the method adopted in response to this issue). This noise must be borne in mind, but aggregating the figures also offers a rough and ready reference for conveying where variation between the corpora might be occurring. Following

this rough picture, more specific and more statistically reliable comparisons can be conducted. The aggregated data in Figure 25 and Figure 26 show probabilities for concrete *take* and its alternates in the written and spoken sub-sections, respectively, of each corpus. This can be used as a stepping stone to a closer, more careful analysis of particular alternates.

Figure 25: Verbs with the sense 'Transfer (Concrete)' in the written portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
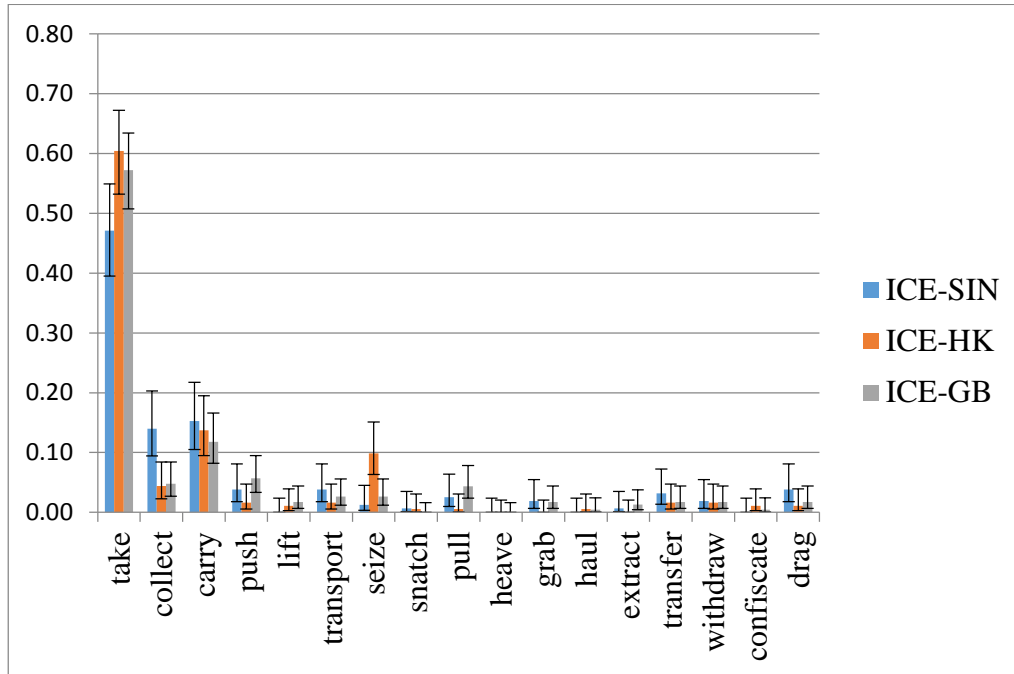
Figure 26: Verbs with the sense 'Transfer (Concrete)' in the spoken portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

It is apparent from Figure 25 and Figure 26 that preferences for *take*, *collect*, and *carry* vary significantly across written and spoken data. In addition, there appears to be a curious preference for *seize* in the spoken portion of ICE-HK and for *transport* in the written section of ICE-GB. Those two anomalies are addressed first, and then the three most frequent, and variable, alternates are discussed below.

The high occurrence of *seize* in speech in ICE-HK does not in fact appear to relate to any kind of linguistic norm for the region. Although the instances are spread across eleven texts, they all occur in descriptions of police *taking* evidence, contraband, drugs, and so on. Similarly, six of the seven instances of *seize* in ICE-GB are related to seizures by police (the seventh instance being literary), and all eight instances of *seize* in ICE-SIN are related to seizures by police. The apparent preference for the alternate *seize* in ICE-HK is actually a reflection of the sampling procedure – ICE-HK happens to contain

more discussions of police activity, particularly in the sampled news broadcasts and courtroom proceedings, than the other two corpora. In this case, there is no regional norm in Hong Kong that prefers *seize* over its alternates, but real-world context and topic within the data, based on the sampling procedure, have affected the picture. In fact, all of the three corpora prefer *seize* when describing police taking evidence, contraband, drugs, and so on, from the scene of a crime, but ICE-HK contains more such descriptions than the other two corpora.

Transport seems to be strongly preferred over a number of other infrequent alternates in ICE-GB, but not in ICE-SIN or ICE-HK. This preference in ICE-GB holds across an array of texts, real-world contexts, and topics. The written portion of ICE-GB does include six instances of *transporting heat*, categorized here as a concrete usage in cases where *heat* is perceived by human senses (as opposed to a phenomenon in physics imperceptible by the five senses). Neither of the other corpora includes references to *transporting heat*. Even if those instances are removed from the ICE-GB data, however, the preference for *transport* in ICE-GB remains. It may be that British English does prefer *transport* in writing over numerous infrequent alternates. I look more closely at *transport* in relation to *collect* and *carry* below.

Figure 27 and Figure 28 isolate the three most frequent and variable alternates, *take*, *collect*, and *carry*, in addition to *transport*, which is preferred in writing in ICE-GB; eliminating the other alternates reduces some of the noise introduced by aggregating those far more semantically specific verbs which do not universally alternate. These four verbs are relatively general semantically, and alternate with each other more consistently in the corpus than with more semantically specific verbs such as *grab* or *haul*, presented in Figure 25 and Figure 26. This also results in a more precise research question: 'Does any variety prefer *take*, *collect, carry*, or *transport* in a unique way, and in what contexts does that preference hold?' We have already seen that British English shows a stronger preference for *transport* in writing than the other two varieties, but the data reveals even more variation.
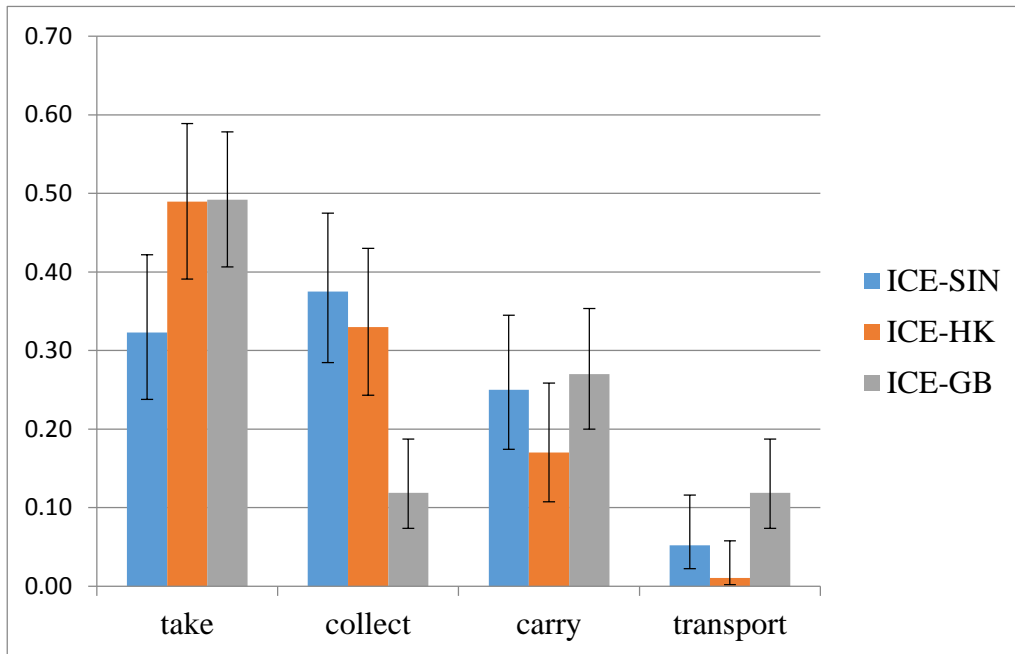
Figure 27: Instances of *take*, *collect*, and *carry* with the sense 'Transfer (Concrete)', in the written portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

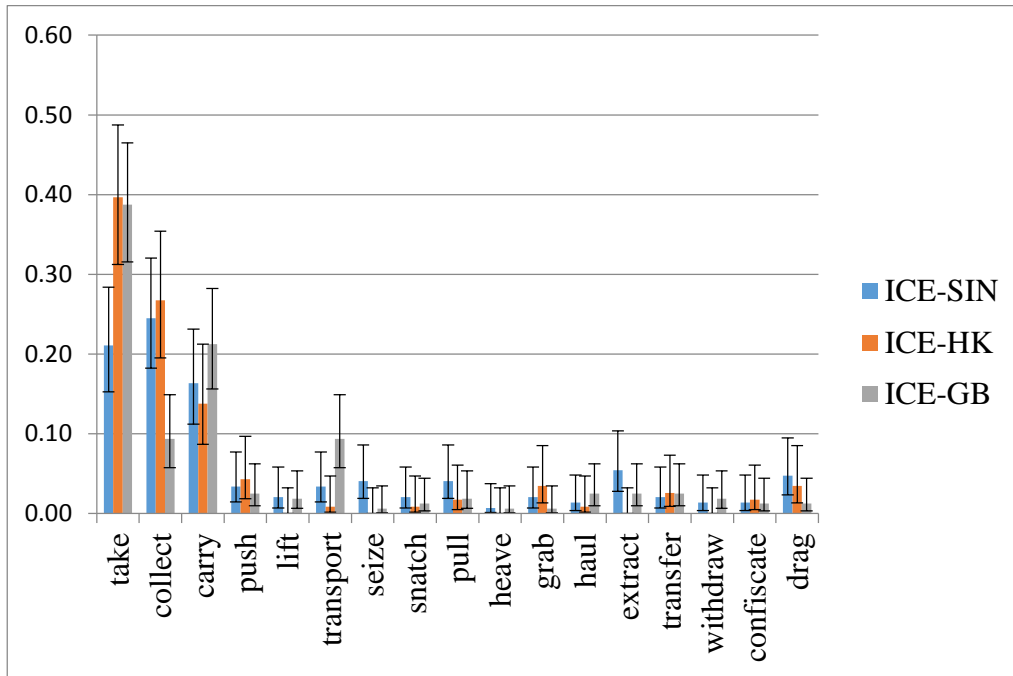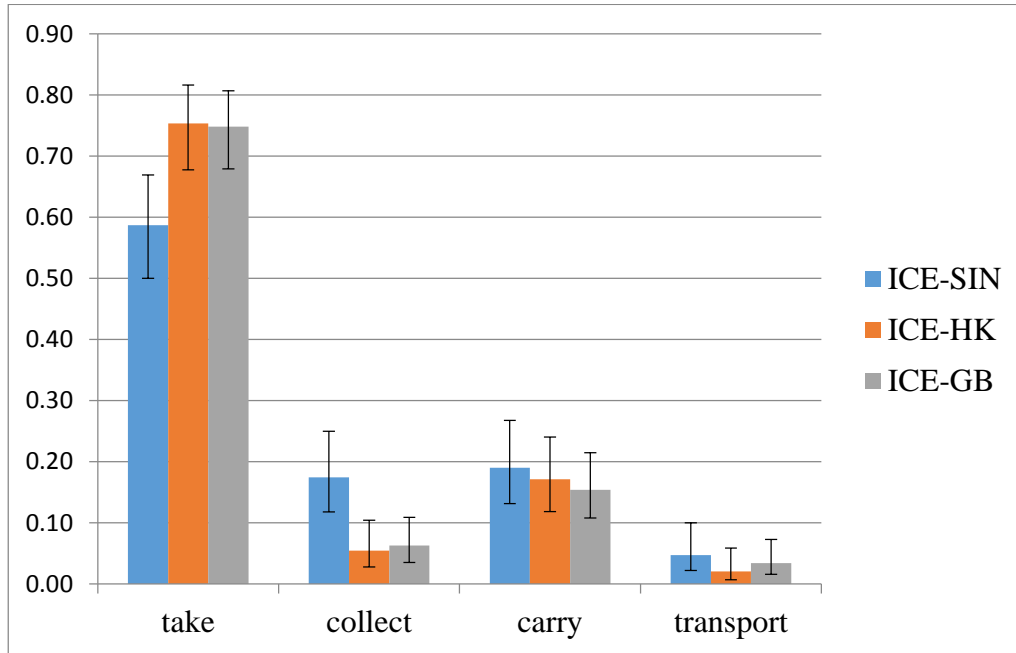Figure 28: Instances of *take*, *collect*, and *carry* with the sense 'Transfer (Concrete)', in the spoken portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

All three corpora exhibit the expected pattern observed for *make* whereby the semantically general, monosyllabic, Germanic alternate *take* is strongly preferred in speech over the more semantically specific, polysyllabic Latinate alternates *collect*, *carry*, and *transport*. In writing, however, the pattern observed for *make* does not hold for *take*. ICE-GB does not exhibit a significantly increased preference for *collect* or *carry* in writing; in both the written and spoken sections of ICE-GB, *take* is significantly preferred over all others, *carry* is second most frequent, and *transport* and *collect* are statistically indistinguishable in third place. ICE-SIN and ICE-HK, on the other hand, do display a significant increase in preference for *carry* and *collect*, but only to the degree that there is an equal preference for *take*, *carry* and *collect* (as evidenced by a Newcombe-Wilson test with continuity correction). As noted above, ICE-SIN and ICE-HK show a significant dispreference in both speech and writing for *transport*. These differences are not due to any particular text with an

unusually high or low number of instances of either alternate: the findings appear to be robust. This is a relatively complex picture that differs greatly from the findings for *make*. The significance of this finding is discussed at much greater length, and in light of all other findings for *make*, *take*, and *give*, in 9.1.

7.2.18. Onomasiological analysis: Transfer (Abstract)

This sense occurs in the corpora with an array of Direct Objects that generally occur only once or twice each. An analysis was conducted of the most frequent Direct Objects, and no consistent alternates were discernible. This finding might suggest that, like abstract *make* (see 6.2.9), the existence of this hypothetical sense (already deemed slippery at best in 7.2.3) would have been refuted. However, the low frequency in use of this sense renders such a conclusion uncertain. Ultimately, the relatively low number of instances of this sense prevents further analysis.

7.2.19. Onomasiological analysis: Adopt

The original analysis of the sense 'Adopt/Assume' included all of the Direct Objects in Table 13.

Direct Objects of *take* with the sense 'Adopt/Assume'	responsibility, opportunity, lead, measure, form, turn, place, step, view, position, precaution, course, prize, job, interest, stance, role, charge, blame, attitude, line, kick, throw

Table 13: Direct Objects of *take* with the sense 'Adopt/Assume' in ICE-SIN, ICE-HK, and ICE-GB

Take with the sense 'Adopt/Assume' is not internally cohesive to any considerable degree either. The two recurring alternates in the corpus are *adopt* and *assume*. The alternate *seize* occurs with the Direct Objects *opportunity* and

chance, but no other Direct Objects, and occurs only six times or fewer in each corpus with this sense. A statistical onomasiological analysis is therefore performed using the three alternates *take*, *adopt*, and *assume* with the sense 'Adopt/Assume', as shown in Figure 29 and Figure 30.
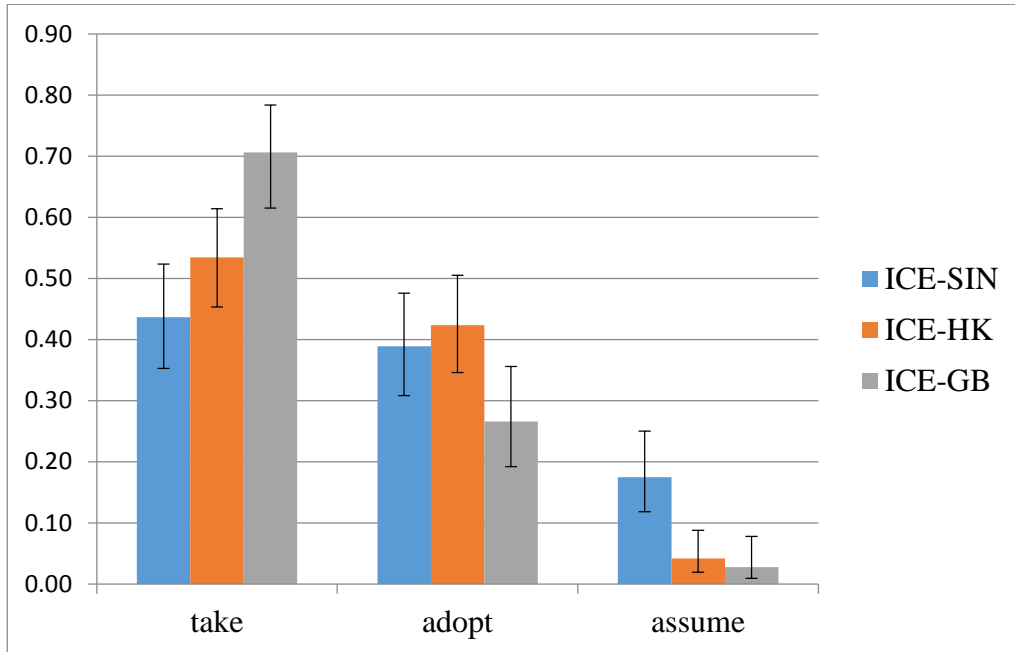
Figure 29: *Take*, *adopt*, and *assume* with the sense 'Adopt/Assume' in the written portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
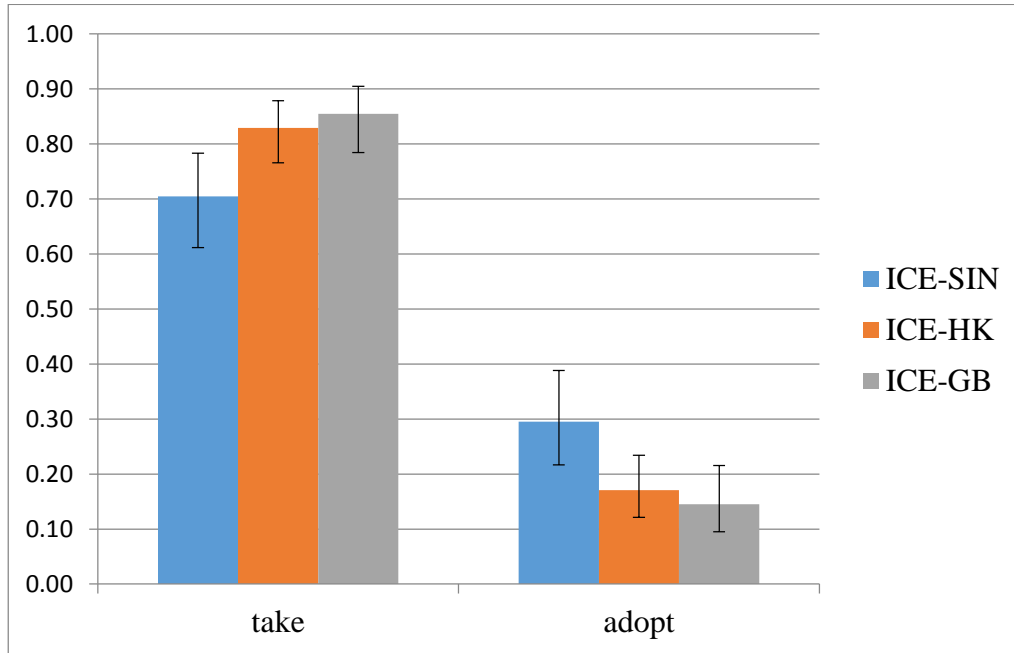
Figure 30: *Take*, *adopt*, and *assume* with the sense 'Adopt/Assume' in the spoken portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

As per the standard expectation, all three corpora show a strong preference in spoken language for the colloquial, monosyllabic, Germanic *take* over its polysyllabic Latinate alternates *adopt* and *assume*. In written language, that preference changes, as with other senses of *take* and with *make*. In this case, ICE-GB continues to prefer *take* over its alternates, while ICE-SIN and ICE-HK show an equal preference for *take* and *adopt*. The higher relative preference for *adopt* in ICE-SIN and ICE-HK is not due to any particular idiosyncratic texts with a high preference for *adopt*, and it is not due to any particularly high number of idiosyncratic Direct Objects with *adopt*. The finding seems to be robust. It is particularly interesting that these measurements for *take* with the sense 'Adopt' are very similar to the measurements for concrete *take*: in both instances, *take* is preferred in all corpora in speech, and by ICE-GB in writing, while ICE-SIN and ICE-HK show an equal preference for *take* and polysyllabic, Latinate alternates in writing.

7.2.20. Onomasiological analysis: Consider/Apprehend

As noted in 7.2.6, this sense appears with various adjuncts and modifiers including *take for granted*, *take into account*, *take into consideration*, *take lightly*, *take seriously*, *take X as Y*, and *take X to be Y*. The prevalence of these constructions complicates the identification of onomasiological alternates. Nevertheless, an analysis of the Direct Objects of *take* with the sense Consider reveals one clear alternate: *consider* itself. *Consider* introduces its own grammatical issues, including a complex transitive construction (e.g. *They consider him a good student*), and, like *take*, a construction with a clausal complement (*consider X to be Y*). Ultimately, these grammatical variants, only a few of which are parallel between *take* and *consider*, are numerous enough, and diverse enough, that reliably identifying instances of alternation was deemed impossible. One solution would perhaps be to compare only bare instances of *take* and *consider*, in which each verb takes a Direct Object with no modifiers, complements, or adjuncts. However, bare instances of *take* occur so infrequently in the corpora that findings cannot be conclusive. A full onomasiological analysis of this sense, therefore, was not performed.

7.2.21. Onomasiological analysis: Delexical

As with *make*, there are no single, consistent semantic alternates for Delexical *take* constructions. Instead, each Delexical construction is equivalent to a unique alternate that the present study has called the related verb, insofar as it is related to Direct Object in the Delexical construction. Delexical *take* constructions consistently alternate with a related verb, such that there is an internally coherent system of alternation.

Because the ICE corpora are relatively small, the present study has consistently aggregated data sets for analysis, particularly for concrete senses of verbs. With Delexical constructions, such aggregation is less feasible – each Delexical construction has an entirely unique alternate unrelated to other Delexical constructions. Four Delexical Direct Objects occur frequently enough in the corpora and have frequent enough alternates in the corpora for statistical

analysis: *take a decision*, *take a look*, *take care*, and *take action*. In a larger corpus, a broader array of Delexical constructions and alternates would appear frequently enough for a full statistical analysis. However, the additional manual analysis of such a corpus would likely be prohibitively costly. Figure 31 and Figure 32 display probabilities of selecting the first two common Delexical constructions and their related verb alternates in each corpus.
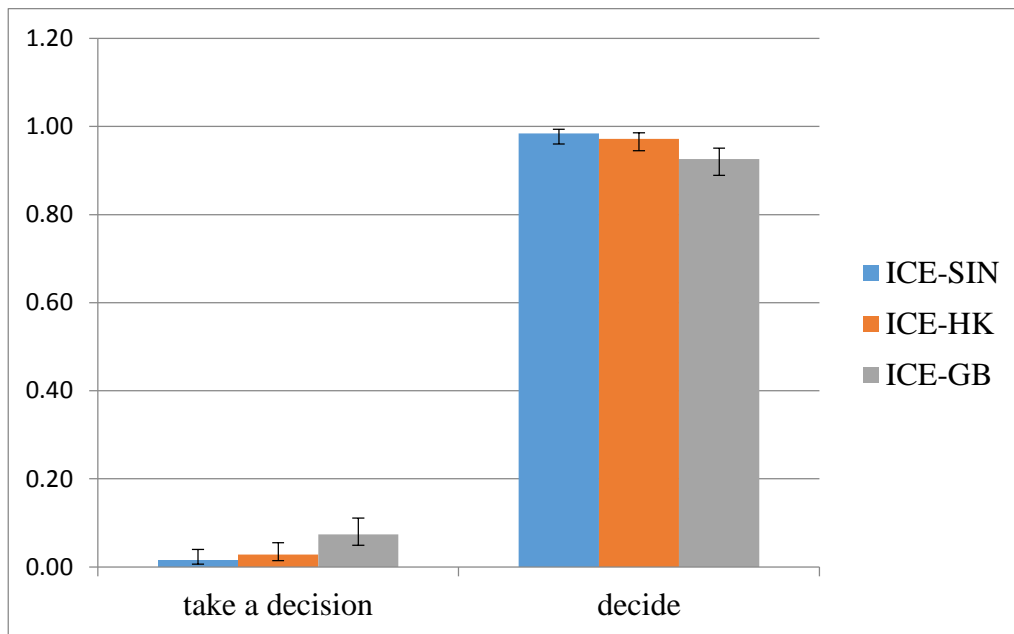
Figure 31: Instances of *take a decision* and *decide* in ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
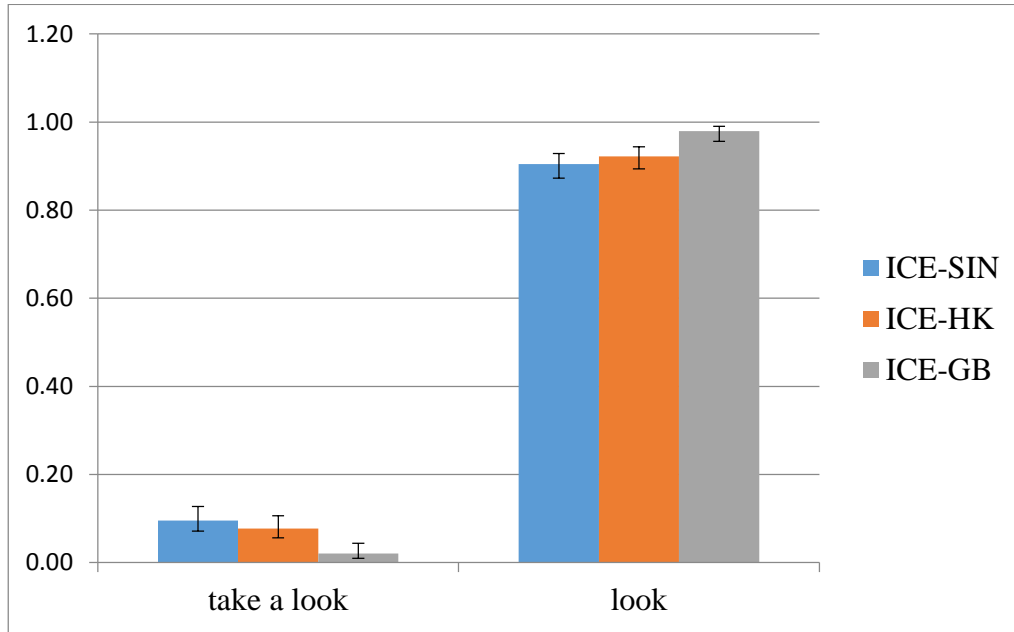
Figure 32: Instances of *take a look* and *look* (*v.*) in ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

In Figure 31 and Figure 32, the related verbs of each Delexical *take* construction are significantly preferred over Delexical *take*. This finding is also true of spoken and written sections of the corpora when analysed separately, and this is consistent with the findings for *make* (see 6.2.10). However, unlike Delexical *make*, this pattern does not hold for all the Delexical *take* constructions analysed here.
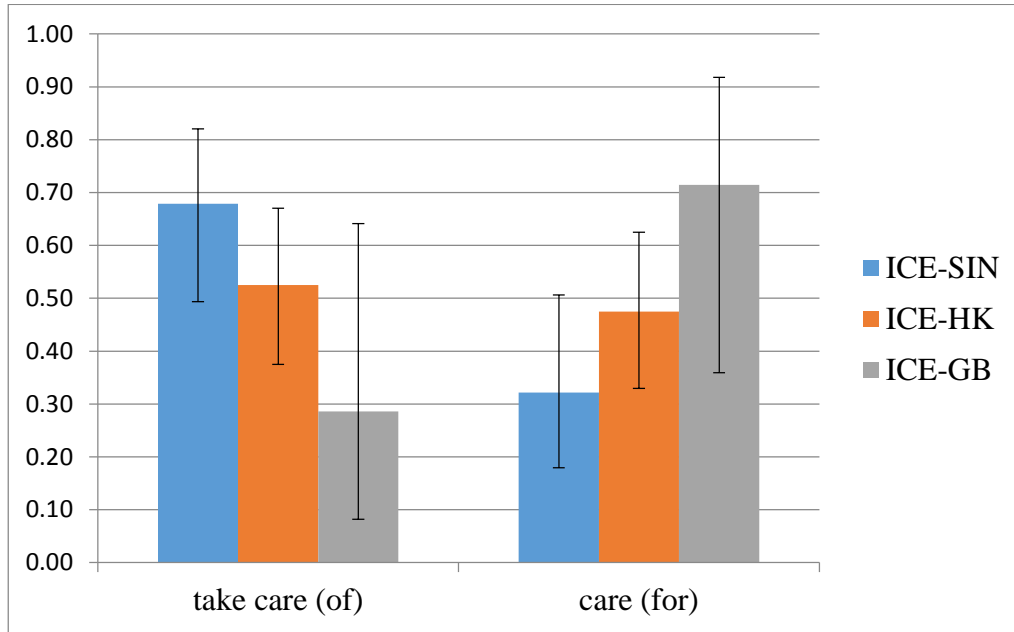
Figure 33: Instances of *take care (of)* and *care (for)* in the written portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
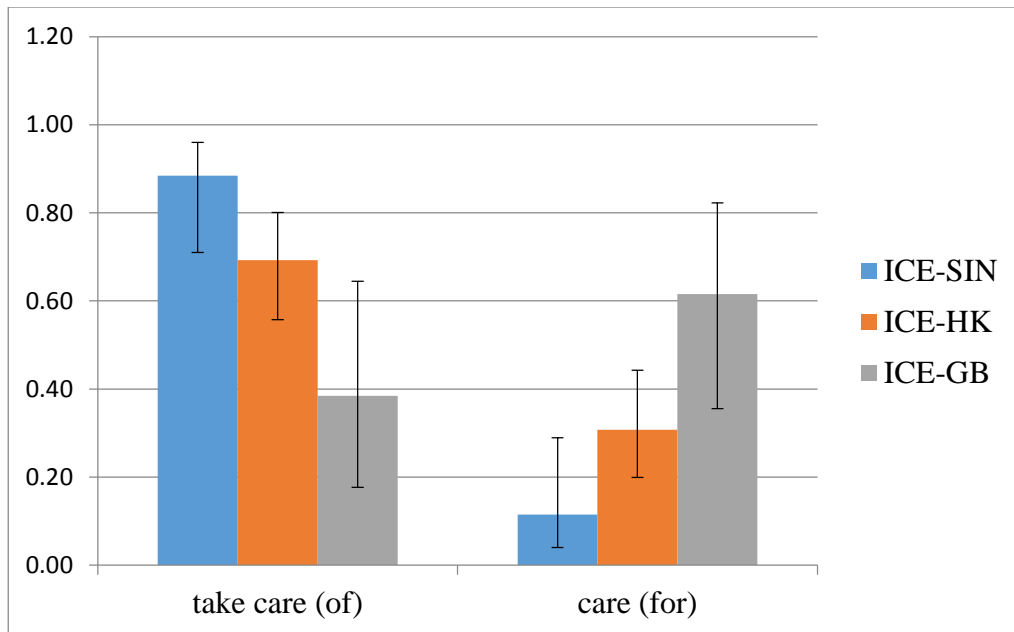
Figure 34: Instances of *take care (of)* and *care (for)* in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

Although it is not immediately apparent in Figure 33 and Figure 34, a
Newcombe-Wilson test with continuity correction reveals that none of the
corpora displays any significant preference for either alternate in written
language, at $p \leq 0.05$. An observer might intuit that the difference in probabilities
between, for example, TAKE CARE (OF) and CARE (FOR) in written ICE-SIN
exhibits a large effect size. The confidence intervals, however, reflect the
relatively low effect size and sample size in the raw tallies (which can be
viewed in the appendix), in turn reflecting a relatively low confidence that the
effect sizes in the raw tallies of occurrences reflect the actual population (see
5.4.3). The Newcombe-Wilson test with continuity correction shows that we
would not expect these effect sizes to be repeated in 19 out of 20 samples from
the same population. Considering significance measures based on 18 out of 20
(or 1 out of 10, i.e. $p \leq 0.1$) is relatively rare in statistics. Such an approach
would accept a considerably lower confidence in research findings than is
generally acceptable. Moving from the written to the spoken corpora, both ICE-
SIN and ICE-HK display a significant preference for the Delexical construction
take care over its alternate. The spoken portion of ICE-GB, on the other hand,
shows no preference. The preferences apparent in speech do not seem to be
attributable to any individual idiosyncratic text, or to any idiosyncrasy in real-
world context or text topic. With this construction, the corpora converge in
writing, and diverge in speech, such that ICE-SIN and ICE-HK resemble each
other and differ from ICE-GB.

Figure 35 and Figure 36 show the probabilities for selecting *take action*
and its related verb *act* in the written and spoken portions of each corpus.
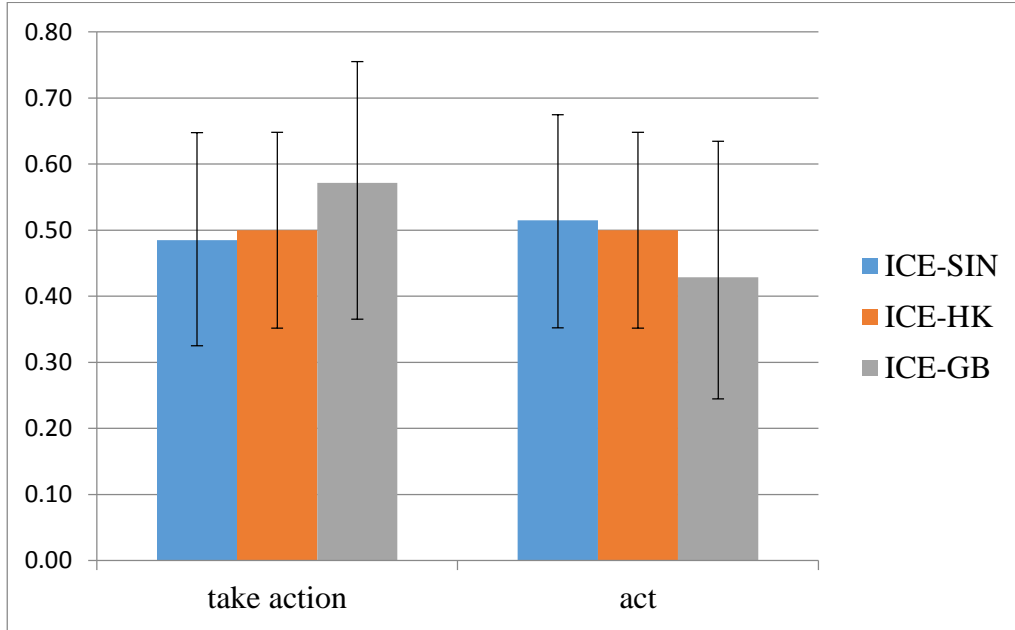
Figure 35: Instances of *take action* and *act* in the written portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
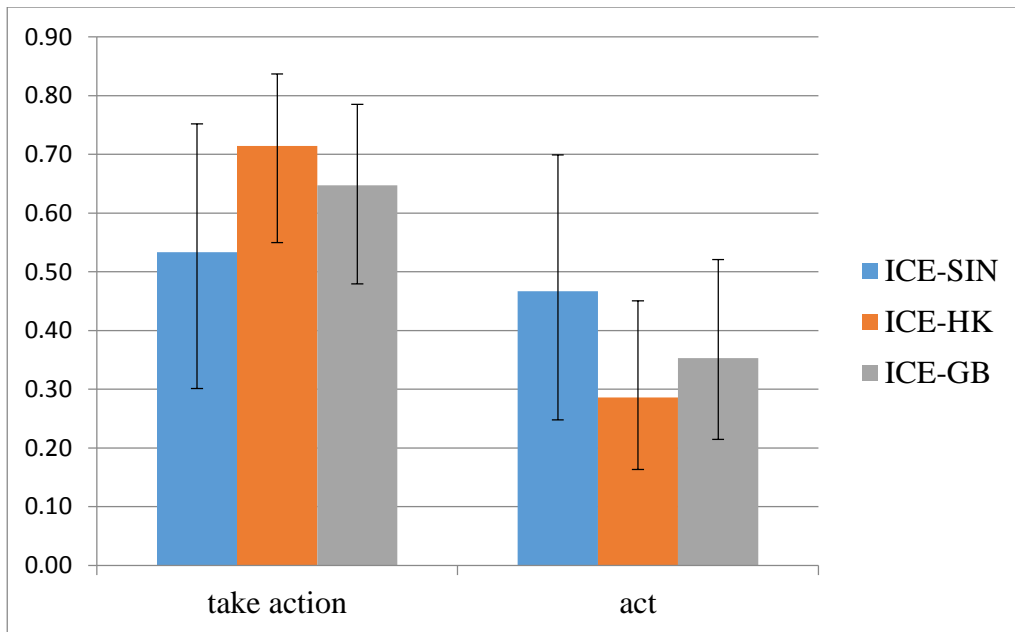
Figure 36: Instances of *take action* and *act* in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

There is no significant preference for either alternate over the other in any of the three written corpora. As with *take care (of)*, however, the spoken portions diverge. Figure 36 shows that ICE-GB and ICE-SIN show no significant preference for either alternate in speech. Again, a casual observer might intuit a large effect size between *take action* and *act* in the spoken portion of ICE-GB. However, the confidence interval based on the raw tallies (see appendix) indicates that we cannot be confident that this effect size would be corroborated in 19 out of 20 samples from the same population (see 5.4.3). That confidence interval reflects the effect size and sample size in the raw tallies of each occurrence in each corpus. ICE-HK shows a significant preference for the Delexical *take* construction in spoken language. With this pair of constructions, the corpora converge in writing but diverge in speech such that ICE-HK is unique in its preference, while ICE-GB and ICE-SIN resemble each other.

7.2.22. Take: Summary

As with *make*, the most striking finding here is the significant similarity in meaning and use of *take* in general across the three corpora: no novel senses are evidenced, and no unique alternates arise, in contrast to expectations in some lexicographical and World Englishes literature (see 4.3). As with *make*, norms for speech tend to be uniform across the three corpora, the sole exceptions arising with *take care (of)* and *take action*, which exhibit differences in spoken language between the corpora. Despite the limited exceptions, this general similarity in speech contradicts the assumption that innovative features should appear first in speech and afterwards spread to writing, and the assumption that there is a common core of usage in written English worldwide.

Concrete and abstract *take* exhibit a remarkable similarity across the three corpora. *Take*, in both concrete and abstract senses, is preferred over its

polysllabic, Latinate alternates (concrete *carry* and *collect* and abstract *adopt*) in speech. In writing, on the other hand, only ICE-GB prefers concrete and abstract *take* over their polysyllabic, Latinate alternates; ICE-SIN and ICE-HK prefer the polysyllabic, Latinate alternates (concrete *carry* and *collect* and abstract *adopt*) equally to *take*. In this case, ICE-SIN and ICE-HK cluster together, whereas for concrete *make*, ICE-HK was the unique dataset. Moreover, in this case it is ICE-GB that, perhaps unexpectedly, prefers the monosyllabic, Germanic alternate in writing rather than the polysyllabic, Latinate alternate; ICE-SIN and ICE-HK prefer the Latinate one. This distinction is discussed further, and compared to the preferences for *make* and *give*, in 9.1.

In Delexical usage, two of the Delexical constructions with *take* resemble Delexical *make*: *take a decision* and *take a look*. In both cases, the Delexical alternates (*decide* and *look* (v.), respectively) are preferred in all corpora, in speech and writing, over the Delexical constructions. The unusual cases, then, are *take care (of)* and *take action*. Each of those constructions displays unique preference patterns in the corpora, with no apparent consistency between datasets or genres. These inconsistencies do not seem to be attributable to variation in text topic or particular idiosyncratic texts. In addition, Delexical *take* seems to be more discrete from other senses of *take* than Delexical *make* is from other senses of *make*. Delexical *take* exhibits less overlap and ambiguity with other senses, and there is less antagonism evidence, autonomy evidence, identity evidence, and truth-condition evidence for allowing overlap between the senses.

8. GIVE

8.1. Hypothesis

Give is the least common of the verbs examined here *pmw*, as evidenced in the BNC, ICE-SIN, ICE-HK, and ICE-GB. Like *make* and *take*, *give* is a highly polysemic verb which is generally considered a good representative of light verbs. In addition, an analysis of *give* can be seen as a complement to an analysis of *take*, already performed in Chapter 5. The high frequency of *give* in all three corpora, as well as its highly polysemic character, facilitates a high degree of nuance in semantic analysis over a relatively large amount of data. As with *make* and *take*, a working hypothesis for the semantics of *give* was first established, influenced by the lexicographical work in the OED and the Collins COBUILD Dictionary, and published academic work. The lexicographical and academic work is summarized first, and the corpus evidence is presented in 8.2.

The OED divides *give* into 14 major senses and 49 sub-senses. Collins COBUILD lists 3 major senses. As with *make* and *take*, Collins COBUILD lists first a sense 'used with nouns describing actions'. This is the light sense, but it also includes other abstract uses. The OED does not include a distinct sense for the Delexical use of *give*, even though *give* is often considered a typical example of a verb used in Delexical constructions (see 3.4). Instead, the OED subsumes Delexical usages under particular types of transfer; many such uses are categorized as transfers 'from oneself', including *give a kick*, *give a kiss*, and *give a look*. The OED entry was published in 1899, and has not been updated, so it is conceivable that the update will incorporate the 20th century notion of light verb usage into its sense distinctions.

> 153.   Humphrey's only reply was **giving** a lash to Billy…
>   (Marryat 1847, qtd. in OED3: *give*, v., 14a)
> 154.   He reached for her hand and **gave** it a reassuring squeeze.
>   (Collins 2012, v., 1)

In Example 153, *give a lash* is equivalent to *lash* (v.), and in Example 154, *give a squeeze* is equivalent to *squeeze* (v.).

Both dictionaries include a set of sub-senses with a concrete Direct Object, in which the concrete Direct Object is transferred from the agent or Subject to a recipient or Indirect Object by a physical act.

155. The letters you **gave** me to deliver at Breme. (Moryson 1617, qtd. in OED3, *give*, v., 6a)

156. They **gave** us t-shirts and stickers. (Collins 2012: *give*, v., 2)

The OED includes a dominant historical sense that could take a concrete or abstract Direct Object, with the meaning 'confer ownership'. In fact, the OED generally delineates senses based on the nature or purpose of transfer (e.g. conferring ownership or not; conferring as a sacrifice or dedication) rather than on the nature of the thing being transferred (e.g. whether it is concrete or abstract). The OED lists multiple examples of abstract sub-senses, delineated based on the nature of the transfer.

157. **Give** my love to Clive. (Thackeray 1853, quoted in OED, *give*, v., 6d)

158. She [Nature] **gave** him [man] alone the power of laughing. (Knox 1794, quoted in OED, *give*, v., 3a)

Example 157 is glossed as 'to deliver' with 'an immaterial object'. Example 158, on the other hand, is glossed as 'to bestow' or 'to grant'. The Direct Objects in both Examples 157 and 158 are abstract, but 157 is seen as delivery without the conferral of ownership, while 158 is seen as distinct for its conferral of ownership. Collins COBUILD does not make this distinction. As discussed in 5.3 and 6.1, the OED and the COBUILD Dictionary are different in many ways. The historical significance of the sense 'confer ownership' and its

ongoing historical relationship with the sense 'Transfer (Concrete)' seems to impact the OED's sense categories in a fundamental and essential way.

Newman (1996) offers perhaps the most thorough academic study on the meaning of *give*. He analyses verbs with the meaning 'give' across numerous languages of the world from a cognitive perspective, taking an 'encyclopedic' approach to *give*'s meanings (*ibid*.: 37). Newman (*ibid*.: 1) acknowledges that *giving* in different settings, languages, and cultures carries additional unique features and rituals, and, while acknowledging that documenting all interconnections between the meanings of *give*, its cultural and social contexts, and even its grammar would be impossible, he nonetheless examines key points about all of those features. He (*ibid*.: 1) identifies a typical sense of *give* whereby a giver 'passes with the hands control over an object… to another person'. This sense of *control* is essential to Newman's definition: *control* may include legal ownership, as in the OED's definition of *give*, but need not be strict legal ownership – indeed, in contrast to the OED, Newman concludes that there is 'no compelling reason' that possession should be considered part of the 'prototypical sense' of *give*, even if *control* is a part of that prototypical sense (*ibid*.: 47). Crucial to understanding *give*, then, are a *giver*, a thing being given, and a *recipient*, as well as the transfer of control. The giver is the grammatical Subject in an active clause, and the movement thing given is moved away from the Subject (*ibid*.: 57). Looking in detail at the context and use of *give*, Newman (*ibid*.: 41) argues for a default sense of *give* in which the thing given is in motion between a giver and receiver, but notes that sometimes the giver may be in motion as well. He claims a 'strong sense of directionality' that corresponds to *give* (*ibid*.: 22) and also a strong 'energy flow' from the giver to the recipient (*ibid*.: 49); it must be noted that 'strong sense' of directionality and the 'energy flow' are not rigourously defined and seem to be untestable in practice. He also notes (*ibid*.: 44) that in the act of concrete *giving*, contact is necessarily maintained first between the giver and the thing, and then between the recipient and the thing (*ibid*.: 44). In part, Newman (*ibid*.: 2) identifies this concrete sense, with all of its detail, as typical

because of the primacy of the concrete sense generally accepted in Cognitive Linguistic frameworks: because cognition derives from embodied experience, concrete senses are primary. Additionally, Newman notes that human beings use concrete things to describe and interpret abstract notions: the concrete sense therefore qualifies as a prototype under Rosch's (1973, 1975a, 1975b) criterion that non-prototypical items are interpreted in terms of prototypical ones.

I. Concrete (Primary): A giver which is the Subject referent passes an object by hand, along with control over that object, to a receiver. Contact is maintained first between the giver and the object, and then between the receiver and the object – *Kim gave Lee a nice birthday present.*

II. Metaphorical Extensions
   a. interpersonal communication, conceptualized as information passed from a giver to a receiver – *give advice*, *give an opinion*
   b. emergence, whereby the thing given to the receiver first emerges from the giver – *give milk*
   c. causative, in which an action is caused to happen, generally with a Direct Object that has a related verb – *give a call*, *give a smile*
   d. permission/enablement, conceptualized as transferring permission or ability from a giver to a recipient – *give someone the right to do something*; *give permission*

It would seem that some element of senses a and d, as in *give an opinion* and *give the right*, could be subsumed under a single 'abstract transfer' sense; I do just that in my hypothesis below. Newman (*ibid.*: 148, 176) notes that *give a yell* or *give a scream* can be seen as examples of 'interpersonal communication', 'emergence', or 'causative' senses, so it is clear that he does not consider these senses necessarily discrete. Newman's causative sense is generally very similar to the category of light verb constructions, and he underlines some of the semantics of light *give*, such that *give* is not devoid of

lexical meaning in such constructions, as claimed by Jespersen (1954) and others (see 3.4), but instead communicates some essential sense of a giver and a receiver. In addition, he notes that 'causative' Direct Objects, or light constructions, must represent acts that are deliberate or controlled, and punctual, similar to the act of concrete *giving* (Newman 1996: 176). It is particularly interesting that Newman happens to have examined the light sense of *give* rather than other verbs. My findings (see 8.2 and 9.1) suggest that some lexical meaning may indeed by retained by *give* in light constructions, but that the other light verbs examined here, *take* and *make*, retain less lexical meaning in their light constructions: it may be that Newman (*ibid.*) happened upon an unusual case.

Gilquin (2008) identifies fifteen senses of *give*, derived from five learner's dictionaries. Like the present study's reference to lexicographical research in its hypothesis formulation, Gilquin's reference to learner's dictionaries constitutes a valid approach. More broadly, her study investigates the cognitive salience of *give*'s various senses and their correlation (or lack thereof) with the frequency *pmw* of these senses in specific corpora. The study and its implications are discussed in greater depth in 9.3. Gilquin's semantic classifications are as follows:

> I. Hand – 'He pulled a handkerchief from his pocked and gave it to him.'
>
> II. Provide – 'Please give your seat to an elderly or disabled.'
>
> III. Communicate – 'Would you like to give me your name?'
>
> IV. Cause – 'All that driving has given me a headache.'
>
> V. Do (delexical sense) – 'He turned to us and gave a big smile.'
>
> VI. Use – 'Give your time to the community.'
>
> VII. Administer – 'Can't you give her anything for the pain?'
>
> VIII. Allow – 'The draft would give him the power to appoint the bank's chairman.'
>
> IX. Job – 'My algebra teacher always gives us a lot of homework.'

X. Organize – 'The ambassador is giving a banquet for the visiting president.'

XI. No longer resist pressure – 'The branch suddenly gave beneath him.'

XII. Decide – 'The judge gave him a nine-month suspended sentence.'

XIII. Idioms – 'You have to give way to traffic coming from the right.'

XIV. Phrasal Verbs – 'We're going to give up our sports club membership after this year.'

XV. Other

(Gilquin 2008: 243)

As with Gilquin's (*ibid*.) sense categories for *take*, there is a large number of senses here, which would seem to result in a relatively cumbersome hypothesis for testing purposes. I avoid that problem in the present study by establishing a hypothesis with fewer sense distinctions. It seems that the example in sense III, 'communicate', could just as easily be glossed as sense II, insofar as communication entails 'providing' information: *Would you like to communicate to me your name* or *Would you like to provide me your name* both seem feasible. Sense II, 'provide', would also seem to be a plausible gloss for the examples in senses VI, VII, and VIII. In the corpora examined here, examples occur that resemble most of the examples and sense distinctions given by Gilquin (*ibid*.). Each of Gilquin's senses can be seen to fit within my hypothesized senses below, with the exception of XI; that sense was not included in my hypothesis, and in fact occurs only extremely rarely in the corpora.

Werner and Mukherjee (2012) build on Gilquin's (2008) study and use the same semantic classifications without elaborating on the categories themselves or the means by which they were derived. They offer no further discussion of those sense categories, but their work is discussed at greater length in 9.3 in relation to corpus frequency and cognitive salience.

The present study forwards three hypothetical sense categories for *give*. These sense categories facilitate comparison with *make* and *take*, as discussed in Chapters 4 and 5. The hypothesized sense categories are also supported to some degree by the OED and COBUILD dictionaries, and are loosely informed by the discussion above concerning academic studies. As with *make* and *take*, these senses are likely to overlap to a limited degree in use; that overlap and the issues surrounding it are discussed in relation to the corpus evidence in 8.2.

I.   Transfer (Concrete) - Senses in which a concrete Direct Object referent is transferred from the Subject referent to an optional Indirect Object referent (or dative alternation)

II.   Transfer (Abstract) – Senses in which an abstract Direct Object referent is transferred from the Subject referent to an optional Indirect Object referent (or dative alternation)

III.   Delexical - Senses in which the Direct Object has a related verb, and the *give* construction is roughly equivalent in meaning to that related verb

Section 6.1 describes details of the establishment of hypothetical sense categories for *make*, and much of that discussion applies here as well. Antagonism evidence, autonomy evidence, identity evidence, and truth-condition evidence was used in the analysis of *give* here, as with *make* and *take*. The analysis in the following section asks whether the hypothesized sense distinctions are supported by the corpora, and whether those sense distinctions vary between the three corpora.

8.2. Findings

8.2.1. Give: The lemma

Give appears in the three corpora a combined total of 3,873 times, and each occurrence was manually analysed in order to determine whether variation exists in frequencies and usage of each sense and its semantic alternates between the three corpora and the populations they represent. First, instances of

give in the corpora were identified and analysed to determine whether they affirmed or refuted the hypothesis presented in 8.1. As with *make* and *take*, an onomasiological analysis was subsequently performed.

The lemma *give*, in the forms *give*, *gives*, *giving*, *gave*, occurs frequently in the corpora; an English listener/reader can expect to encounter *give* quite often, depending on text topic and genre, and, perhaps, regional variety. The number of occurrences of *give* in each corpus appears in Table 14.

	Instances of *give*
ICE-SIN	1,412
ICE-HK	1,321
ICE-GB	1,140

Table 14: Instances of *give* in ICE-Singapore, ICE-HK, and ICE-GB

Figure 37 and Figure 38 display semasiological data on the frequencies of each of the three hypothesized senses. These figures address the question: given that a listener or reader is exposed to *give*, what is the probability of encountering each sense as a proportion of the total number of occurrences of all senses? A lexicographer might find this information useful in arranging the order of entries in a dictionary like the Collins COBUILD dictionary.
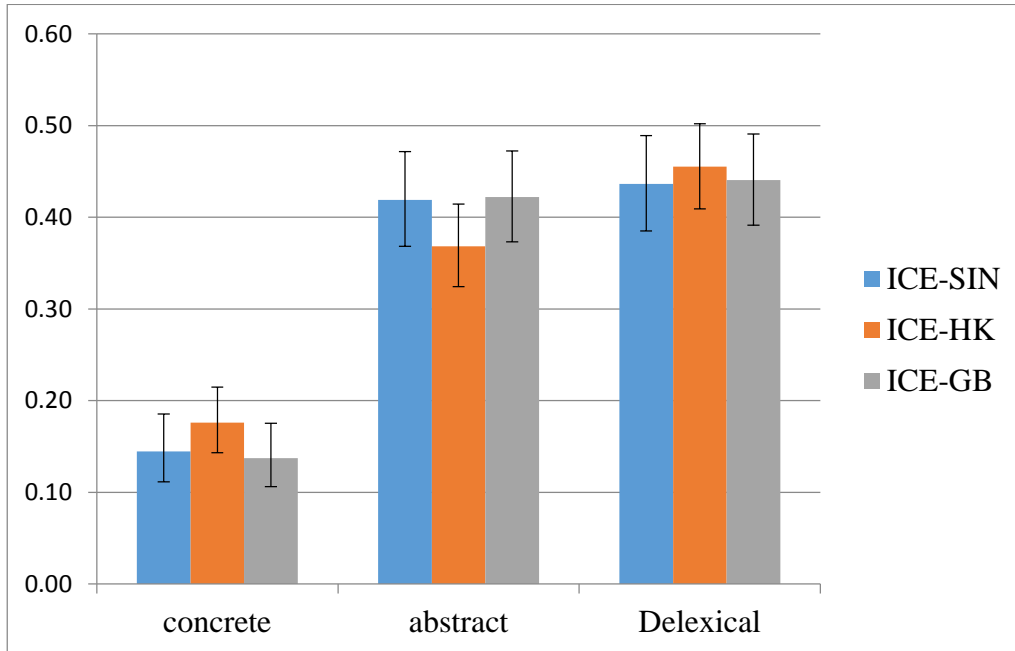
Figure 37: Probabilities of encountering *give* with each of the three hypothesized senses in the written portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
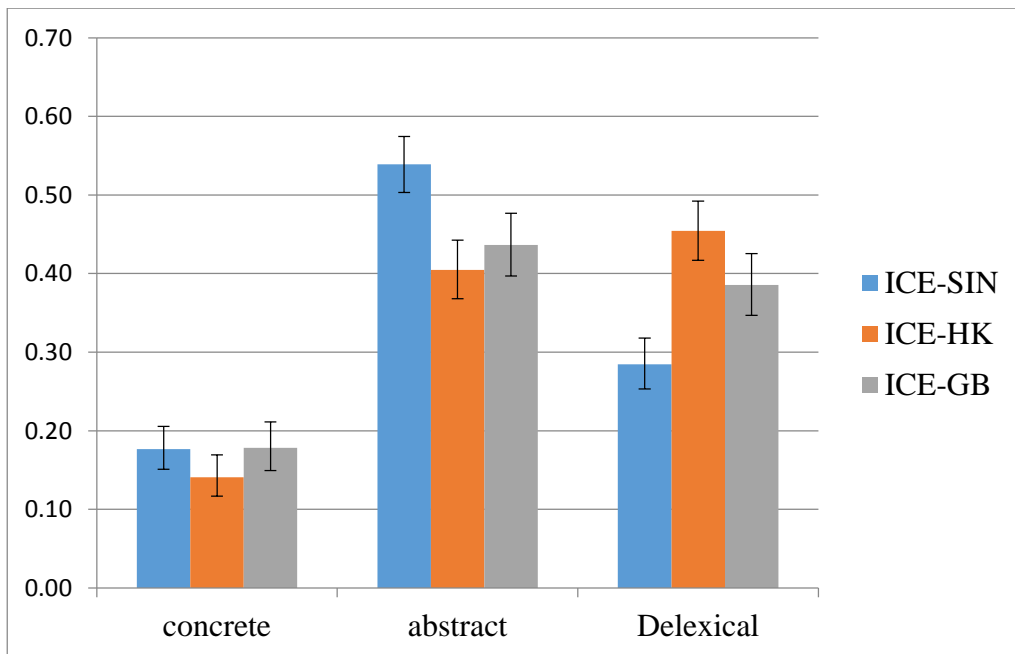
Figure 38: Probabilities of encountering *give* with each of the three hypothesized senses in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

It is clear from the semasiological graphs that exposure rates to Delexical constructions are high, as are exposure rates to the abstract sense. As the subsequent sections will show, however, abstract and Delexical senses of *give* overlap a great deal in all three corpora and, ultimately, may not be distinct at all. This is a remarkable and surprising contrast to *make* and *take*.

8.2.2. Give: Transfer (Concrete)

Categorizing concrete instances of *give* is generally quite straightforward. Clear examples of concrete *give* in the corpora include the following:

> 159.   She always wore her elder sisters' old clothes, and was often **given** the old toys. [ICE-GB W2F-017 #6]
> 160.   The buyer is **given** a slip with the price and a four-digit code. [ICE-HK W2B-036 #33]

In Example 159, the concrete Direct Object is *toys*, and it represents something transferred to a recipient. In Example 160, the concrete patient of the passive *be given* is *slip*, also transferred to a recipient.

In the OED's definition of *give*, the nature of the transfer, particularly as it confers ownership or not, is crucial. In the three corpora studied here, many utterances with concrete *give* do not clearly convey ownership or non-ownership.

> 161.   Meanwhile, a small bit of food was **given** to the subjects when they made the designated response… [ICE-HK W1A-004 #43]

In Example 161, food is provided to subjects participating in an experiment, but whether the subjects are meant to be understood as owning the food is unclear, and indeed perhaps irrelevant. The subjects go on to consume the food (information that is contingent rather than constant to *give*), but ownership does not seem to be emphasized or clarified in any way. In Example 162, conferral of ownership is much more certain.

> 162.   But if there is going to be appreciation you can **give** him a plaque. [ICE-SIN S1B-052 #35]

In Example 162 conferral of ownership is communicated in the utterance via the broader pragmatic context, rather than conveyed strictly via the word *give*. If *giving a plaque* to show appreciation communicates ownership, that is likely because listeners understand that in the real world, when a plaque is transferred from one person to another in order to express appreciation, the recipient probably then owns the plaque.

> 163.   If for some reason he is not in a position to initiate forward play, then he should not be **given** the ball. [ICE-GB W2D-015 #125]

In Example 163, it seems quite certain that *giving the ball* in a football match does not confer ownership of the ball – instead, only a physical transfer is involved. All three corpora are quite consistent in that approximately 40% of instances of concrete *give* communicate clear conferral of ownership; approximately 40% appear quite certainly not to communicate conferral of ownership; and approximately 20% are unclear in this regard, like Example 161. Even in clear cases, like Examples 162 and 163, the conferral of ownership seems to be contingent information rather than information constant to *give*. This breakdown contrasts with concrete *take*, for which only very few

instances convey conferral of ownership. There are no instances of antagonism evidence, autonomy evidence, identity evidence, or truth-conditions evidence to suggest either the unity or discreteness of this ownership/non-ownership pairing with concrete *give*.[37] The possibility that ownership or non-ownership are not distinct elements of the semantics of *give* seems a reasonable proposition, and the present study moves forward with that in mind. In 8.2.3, some evidence for an ownership/non-ownership distinction for abstract *give* is provided.

It is also worth noting that examples like Example 163, particularly describing the giving of a ball in sport, are quite common. These examples would seem to contradict Newman's (1996) claim that in the act of giving, contact must constantly be maintained first between the giver the thing given, and then between the recipient and the thing given. That particular criterion in Newman's (*ibid*.) argument is not supported by the evidence in any of the three corpora, which provides a useful illustration of the value of corpora for testing reflective and subjective understandings of lexical semantics.

8.2.3. Give: Transfer (Abstract)

The abstract sense of *give* involves the transferral of an abstract Direct Object referent from a Subject referent to an optional Indirect Object referent.

164.   You're supposed to **give** me examples. [ICE-SIN S1B-002 #2]

---

[37] Although examples of such evidence do not appear in the corpora, they are conceivable.

i.   She **gave** him that book, but she didn't **give** him that book.

Example i might be interpreted as 'She handed him the book, but she didn't give him ownership of the book.' That such evidence does not occur in the corpora, however, suggests that in actual use, these potential distinctions within concrete *give* do not tend to come into play in any of the regions represented by the corpora.

165.   And here to **give** their perspectives in the studio… [ICE-
         SIN S1B-036 #14]

The Direct Objects *examples* in Example 164, and *perspectives* in Example 165
are abstract, representing information communicated. In these examples, *give*
might be glossed as 'communicate' or 'convey', as it represents the transferral
of abstract information from one person to another. However, it might be
posited that *give* is actually more vague than that here, representing not the
precise meanings 'communicate' and 'convey' but instead a more vague
meaning like 'make available' and also communicating that the thing being
made available is then received by a recipient. If *give* itself conveys this more
vague meaning, then it is only the nature of the Direct Object as representing
information that gives rise to the specific interpretation 'communicate'. In that
case, the constant semantics of *give* would be glossed as 'make available +
receive', with the contingent meaning 'communicate/convey' when the Direct
Object represents information.

166.   At the age of sixty, Alistair MacIntosh was, to his
         surprise, **given** a peerage in the New Year's Honours List. [ICE-
         GB W2F-017 #76]

In Example 166, the Direct Object *peerage* is abstract, but the nature of the act
represented by *give* seems different from that of Examples 164 and 165. A
peerage is not communicated as information but assigned or attributed. Once
again, however, we might interpret *give* slightly more vaguely, such that a
peerage is 'made available and received'. The interpretation 'assign', as
opposed to 'communicate', arises from the fact that a peerage which is made
available and then received is a position that is assigned, rather than
information that is communicated.

        Many examples allow either the 'communicate' reading or the 'assign'
reading.

167.   If we don't **give** them incentives, they're not likely to
continue. [ICE-SIN S1B-056 #68]

In Example 167, it is possible to conceive of incentives *communicated* or
assigned, or perhaps both. The utterance and the context do not disambiguate
the two readings. In instances of *give* with an abstract Direct Object in all
corpora, about 40% quite clearly relate to the communication of information;
about 40% clearly relate to assignation or attribution; and about 20% are
ambiguous between the former and latter senses. It seems reasonable to argue
that the difference between 'communicate' and 'assign/attribute' are in fact
contingent. Moreover, this considerable overlap suggests that in use, *give* does
not reliably distinguish between these senses in the three regions represented.

While abstract *give* does not seem to distinguish reliably between those
two senses, it does in some cases seem to distinguish other specific features.

168.   You have to earn those four marks, they are not going to
give it to you, okay? [ICE-SIN S1B-017 #34]
169.   These people haven't been **given** life, they've only been
lent it. [ICE-GB S2B-033 #19]

Examples 168 and 169 provide truth-condition evidence for *give*, as they
explicitly deny a particular reading of *give*. In Example 168, *be given* is
opposed to *earn*; in Example 169, *give* is opposed to *lend*. These meanings
relate to ownership and the nature of acquiring ownership. As with other truth-
condition evidence, the question arises whether the semantic oppositions are
made explicit in the language because those oppositions are not inherent in the
word *give* and they therefore require disambiguation; or whether the explicit
opposition in the language affirms the constant opposition between the two
words, for example *give* and *lend*. Section 8.2.2 has already argued that
conferral of ownership is not a distinct element of the semantics of concrete

give. These instances of truth-condition evidence for abstract *give* suggest the possibility that with abstract *give*, conferral of ownership might be a distinct element of the semantics; however, these two instances in the corpora do not allow for firm conclusions in that regard. Further inquiry into this semantic distinction, perhaps via psycholinguistic tests with native speakers from each region, would be useful.

8.2.4. Give: Delexical

Many instances of Delexical *give* are quite clear.

> 170.   He demonstrated that in… you know… **gave** an indication of what that should sound like. [ICE-GB S1A-060 #133]
>
> 171.   Unconsciously, she had reproduced the very look her mother **gave** her father… [ICE-SIN W2F-020 #85]
>
> 172.   …it will help us to **give** you more accurate information. [ICE-HK W1B-024 #31]

In Example 170, *give an indication* is equivalent to *indicate*; in Example 171, *give a look* is equivalent to *look* (v.); and in Example 172, *give information* is equivalent to *inform*.

As discussed for Delexical *make* in 6.2.4 and Delexical *take* in 7.2.6, the presence or absence of a patient in a Delexical construction, corresponding to the Direct Object of the related verb, might be relevant to the selection process between alternates, but that question, a primarily grammatical one, is beyond the scope of the present study. Likewise, as discussed in Chapters 6 and 7 for *make* and *take*, passivization or grammatical modification of a Delexical construction might also influence how readily it is perceived as an alternate of a related verb. In the present study, passivized Delexical constructions are analysed as Delexical constructions. Finally, as with *make* and *take*, *give* with a Direct Object that relates to an obsolete verb are not considered Delexical

constructions in the present study. Again, the particular additional issues with light verbs, including these specific grammatical questions regarding alternations, are not the focus of the present study, but these variables must be borne in mind.

One interesting example, which is apparently Delexical, arises only in ICE-HK, twice.

173. British foreign secretary Douglas Hurd has **given** his undertaking that China won't be allowed to interfere in the running of Hong Kong. [ICE-HK S2B-008 #13]

174. …the reluctance on the part of Her Majesty's Government to **give** an undertaking for paying for the repatriation costs… [ICE-HK S1B-059 #75]

A gloss of the precise meaning might require a native speaker informant, but it seems that *give an undertaking* is more or less parallel to *undertake*, and I have categorized it as Delexical. There is no evidence for varied usage or semantics for *undertake* in ICE-HK, but these examples illustrate an innovative flexibility that seems to be allowed in some cases with Delexical verbs in use in different regions.

8.2.5. Give: Transfer (Concrete)/Transfer (Abstract) ambiguities

Examples of ambiguity between concrete and abstract *give* are restricted to the construction *give evidence* (which can be seen not only as abstract or concrete, but also Delexical). In some cases, broader context disambiguates the construction as distinctly concrete. In most cases, however, no disambiguation is possible. Following the general approach of this study (see 6.1), *give evidence* is categorized as Delexical unless clearly disambiguated as concrete. There is no antagonism evidence, autonomy evidence, identity evidence, or truth-condition evidence to raise additional concerns about the discreteness of the concrete and abstract senses, and they seem to be quite distinct.

8.2.6. Give: Transfer (Concrete)/Delexical ambiguities

A few Direct Objects give rise to potential ambiguity between the concrete and Delexical senses: for example, *give a tip* is equivalent to *tip* (v.), but can also suggest the transferral of a concrete object. Similar multiple interpretations are possible with *give a report*, *give evidence*, *give a dose*, and *give an application*.

175.   He **gave** me a ten dollar tip. [ICE-HK S1A-039 #9]

176.   You **give** your report only when the parent come along, isn't it? [ICE-SIN S1B-010 #113]

177.   A dose of activated charcoal was **given** in the emergency department. [ICE-HK W2A-030 #117]

Example 175 might be glossed either as 'tip (v.)' or as 'transfer a concrete thing (a tip) from a giver to a receiver'. Examples 176 and 177 are similar. Following the norms already established in this study, these examples are categorized as Delexical, unless there is clear disambiguation to suggest concreteness.

The discreteness of concrete *give* in the corpora is challenged somewhat by three instances of identity evidence between concrete *give* and Delexical *give*.

178.   You mainly **give** us our technical support and informations, uh, information brochure or some kinds of service support. [ICE-HK S2A-059 #16]

179.   **Give** him a long run and lots of steak from now too. [ICE-GB W2F-001 #142]

180.   He **gave** the young couple his blessing and a rather elegant house to live in. [ICE-GB W2F-001 #052]

In Example 178, *give support* is equivalent to *support* and *give information(s)*[38] is equivalent to *inform* or glossable as 'transfer information', while *brochure* is concrete; in Example 179, the Delexical *give the dog a run* seems to be equivalent to *run the dog*, while *steak* is concrete; in Example 180, *give a blessing* is equivalent to *bless*, while *house* is concrete. There is no comparable evidence whatsoever for coordinated Delexical and concrete Direct Objects in the data for *make* or *take*. Although three instances in all three corpora constitute only limited evidence, the possibility that Delexical and concrete *give* are less distinct from each other than is the case with Delexical and concrete *make* and *take* is remarkable. This suggests the possibility that Delexicality has varying degrees of discreteness in various verbs – that very important implication of the present findings is further corroborated by the ambiguities that arise between abstract and Delexical *give* discussed in the next section, and is summarized and analysed further in 9.1.

8.2.7. Give: Transfer (Abstract)/Delexical ambiguities

Like *make* but unlike *take*, abstract and Delexical instances are often difficult to distinguish from each other.

> 181.   And you know **give give** all the information we can
> about it. [ICE-HK S1A-053 #201]
> 182.   I didn't **give** specific instructions. [ICE-GB S1B-067
> #172]

When a Direct Object refers to information that can be communicated, but also has a related verb form, the distinction between abstract and Delexical senses is unclear. Example 181 can suggest the transfer of abstract *information*, but *give information* is also equivalent to *inform*. Example 182 can indicate the transfer

---

[38] The plural form *informations* for *information* is not uncommon in ICE-SIN and ICE-HK, and has been attested as a common feature of Tanzanian English as well (https://ewave-atlas.org).

of abstract *instructions* from one speaker to another, but *give instructions* is also equivalent to *instruct*. A similar issue arose with Delexical *make*. As with *make*, I categorize such examples as Delexical by default, allowing me to engage more fully with existing literature on Delexical or light verbs. Nonetheless, I believe that the evidence presented thus far seems to suggest that Delexical *give* overlaps more with other senses of *give* than Delexical *make* and *take* overlap with other senses of those verbs.

Unlike *make*, in many cases, *give* takes a coordinated Direct Object that includes both abstract and Delexical elements.

183.   What they really need is to be **given** the technical assistance and guideline... [ICE-HK S1B-047 #90]

184.   Are there any preparatory courses or uhm supports **given** to foreign students? [ICE-SIN S1B-049 #80]

185.   …I would appreciate it if you can **give** us your comments and any ideas… [ICE-SIN W1B-016 #105]

In Example 183, *guideline* is abstract while *give assistance* is equivalent to *assist*; in Example 184, *course* is abstract while *give support* is equivalent to *support* (v.); and in Example 185, *idea* is abstract while *give comments* is equivalent to *comment* (v.). The fact that such coordinated Direct Objects are commonly found in all corpora suggests that the abstract and Delexical senses of *give* are not entirely discrete. There is very little comparable evidence whatsoever for abstract and Delexical *make* and *take*; in fact, the lack of such coordinated Direct Objects was accepted as evidence that Delexical *make* and *take* are relatively discrete from abstract *make* and *take*. Table 15 displays the total number of coordinated Direct Objects including Abstract and Delexical elements in all three corpora.

	Instances of *give* with a coordinated Direct Object: Multiple abstract referents	Instances of *give* with a coordinated Direct Object: Abstract and Delexical referents	Instances of *give* with a coordinated Direct Object: Multiple Delexical referents
ICE-SIN	16	4	5
ICE-HK	10	12	11
ICE-GB	10	2	8

Table 15: Instances of *give* with three types of coordinated Direct Object in ICE-SIN, ICE-HK, and ICE-GB

As in so many other cases already discussed, ICE-HK is the exceptional corpus, with the highest number of coordinated Direct Objects that combine Delexical and abstract referents. In all corpora, however, coordinated abstract and Delexical Direct Objects are remarkably numerous – more so than for *make* or *take*.

With all of the above evidence in mind, it seems clear that Delexical *give* is not entirely straightforward to identify, and there is certainly evidence that Delexical *give* is not as discrete as Delexical *make* and *take*, particularly in ICE-HK. That finding suggests a recognition of something like *degrees of Delexicality*, which is discussed further vis-à-vis a comparison of all three verbs in 9.1. The present study moves forward with an exploratory comparison of light usage in all three of the verbs investigated here, even though the Delexical sense is much more clear and certain in the former two verbs than in *give*.

8.2.8. Onomasiological analysis: Transfer (Concrete)

As with concrete *make* and *take*, in order to measure preferences for semantic alternates in use in the corpora, the present study identified Direct Objects of concrete *give* and then selected the Direct Objects that appear at least twice in any corpus. Alternate verbs were then identified which co-occur with those

Direct Objects, and which convey generally the same meaning as *give*. This list of alternates facilitates a reasonably thorough, evidence-based onomasiological comparison for concrete *give*. Concrete *give* in the three corpora has the alternates shown in Table 16.

Alternates for concrete *give*:	*provide*, *hand*, *issue*, *submit*, *pass*, *donate*, *transfer*, *contribute*, *supply*, *grant*

Table 16: Alternates for concrete *give*, as evidenced in ICE-SIN, ICE-HK, and ICE-GB

Like *make* and *take*, *give* is a quite general term semantically, and its alternates tend to be semantically much more specific. That specificity in turn means that not all of these lexical items can alternate with all others in all circumstances. Comparing all alternates to each other can be seen as a rough first step to give an impression of the types of phenomena that might be occurring. The necessary next step is to isolate individual alternates that do in fact alternate more universally and to compare them one by one: by comparing each alternate to *give* individually, statistical noise is minimized.

Figure 39 and Figure 40 show probabilities for concrete *take* and its alternates in the written and spoken sub-sections, respectively, of each corpus.

Figure 39: Verbs with the sense Transfer (Concrete) in the written portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
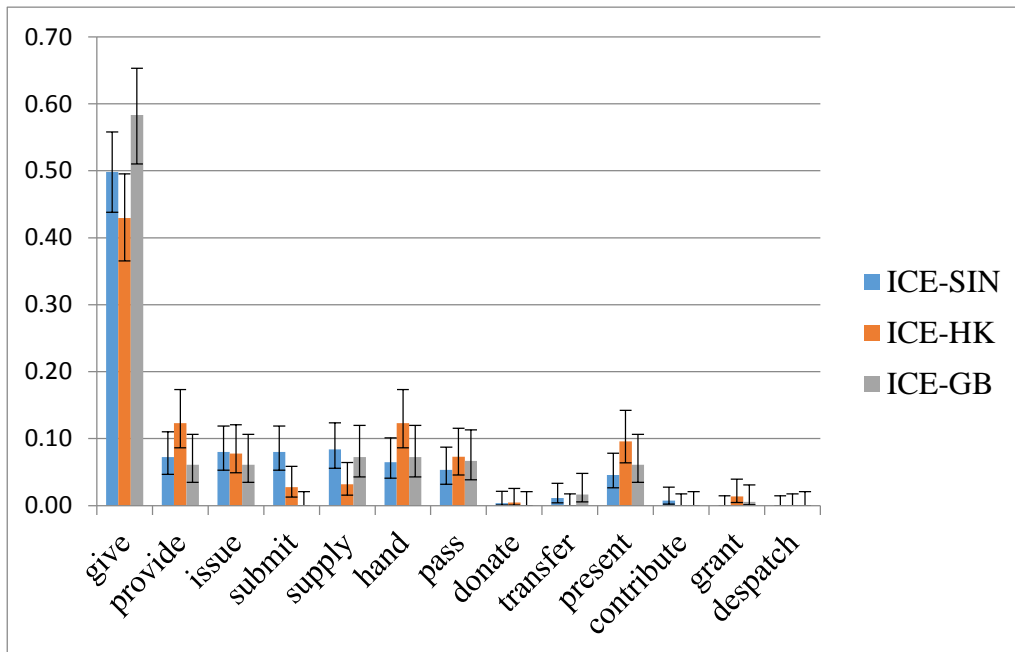
Figure 40: Verbs with the sense Transfer (Concrete) in the spoken portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

It is apparent from Figure 40 that, as with *make* and *take*, preferences in speech for the semantically general, monosyllabic Germanic *give* are stronger than its semantically specific, polysyllabic Latinate alternates. Figure 39 shows that in writing, the three corpora differ significantly in their preferences for several alternates. Figure 41 eliminates some noise from the data by comparing only *give* and those alternates which are preferred in writing as much as or greater than *give*: *provide*, *issue*, and *submit*.
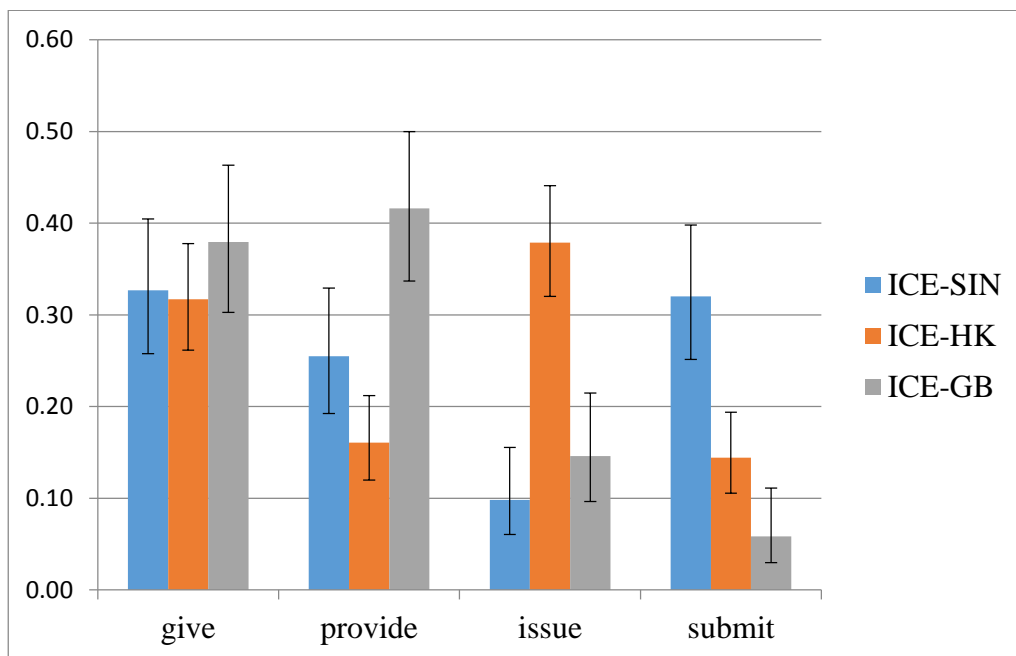
Figure 41: Probability of selecting verbs with the sense Transfer (Concrete) in the written portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

First, it is apparent from investigating individual instances of *issue* and *submit* in each corpus that the fluctuation apparent in Figure 41 for those alternates is

an artefact of the corpus sampling rather than the reflection of a regional norm, as with *seize* in ICE-HK (see 7.2.17). The ICE corpora include a section of instructional writing. This section in ICE-HK is composed almost entirely of instructions for 'issuing' licenses, permits, and certificates: 74 of the 92 instances of *issue* in ICE-HK are from those instructional documents, and occur with those three Direct Objects. ICE-SIN and ICE-GB do not include this phenomenon. It is reasonable to hypothesize that when writers in Singapore and Great Britain do discuss the issuing of licenses, permits, and certificates, they may tend to use the verb *issue*. Similarly, the instructional writing section of ICE-SIN is composed almost entirely of instructions for 'submitting' applications and forms: 32 of the 49 instances of *submit* in ICE-SIN are from those instructional documents and occur with those two Direct Objects. Again, ICE-HK and ICE-GB are not affected by topic in this way. It is reasonable to hypothesize that when writers in Hong Kong and Great Britain do discuss the submission of applications and forms, they may tend to use the verb *submit*. If those instances of *issue* are removed from the data for ICE-HK; and those issues of *submit* removed from the data for ICE-SIN; then no significant difference in usage preferences between the three corpora is apparent. Far from reflecting regional norms, these numbers in fact reflect randomly skewed text topics in the sample, precisely because the ICE corpora do not control for text topic.

The data for concrete *provide* show no skew due to topic or real-world context. Equally importantly, *provide* is semantically very general, like *give*, and *provide* alternates nearly universally with *give*, each occurring with an array of identical Direct Objects.
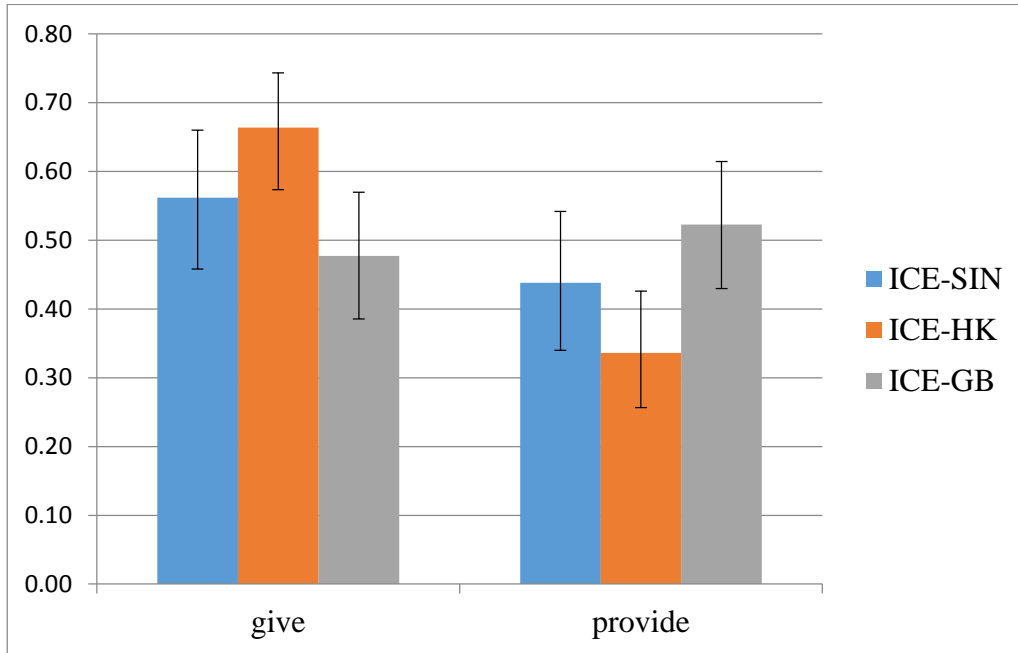
Figure 42: Probability of selecting *give* and *provide* with the sense Transfer (Concrete) in the written portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

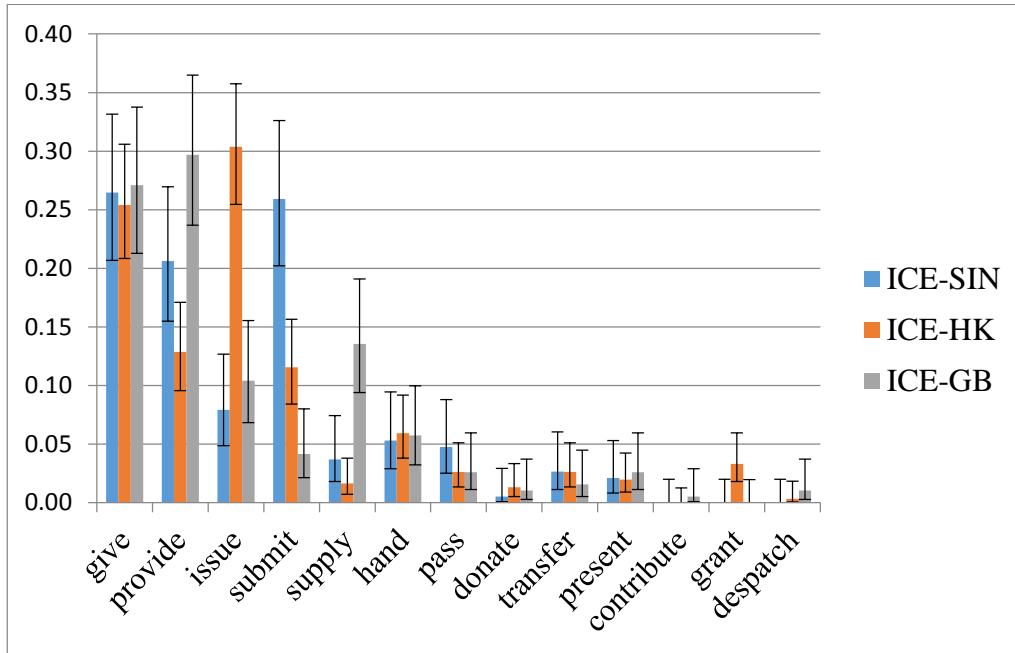Figure 43: Probability of selecting *give* and *provide* with the sense 'Transfer (Concrete)' in the spoken portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

Figure 43 shows that in speech, *provide* is significantly dispreferred against *give* in all corpora. This finding is parallel to the preference for concrete *make* over *provide* and concrete *take* over *collect*, *carry*, and *transport* in speech. Figure 42, perhaps as expected, shows that the preference for *provide* is higher in writing than in speech in ICE-GB and in ICE-SIN; in ICE-HK, however, *provide* is still significantly dispreferred against *give* in writing. This finding parallels the unique preference against *produce* in favour of concrete *make* in the written portion of ICE-HK.

8.2.9. Onomasiological analysis: Transfer (Abstract)

The original analysis of the sense 'Transfer (Abstract)' included all of the Direct Objects in Table 17.

Direct Objects of *give* with the sense 'Transfer (Abstract)'	*opportunity, example, impression, idea, chance, sense, reason, message, address, number, right, choice, account, view, confidence, access, result, status, insight, guideline, opinion, lesson, freedom, flexibility, feeling, incentive, feedback, meaning, rule, picture, overview, discount*

Table 17: Direct Objects of *give* with the sense 'Transfer (Abstract)' in ICE-SIN, ICE-HK, and ICE-GB

As discussed in 8.2.3, some of these Direct Objects more readily suggest an interpretation of *give* as 'communicate' (*e.g. example*, *impression*, *idea*), while

others suggest a reading of *give* as 'assign' (*e.g. opportunity*, *chance*). However, the corpus evidence does not suggest that those two senses are discrete. Instead, abstract *give* is interpreted as 'make available + receive'. The only consistent onomasiological alternate across these constructions is *provide*, which is also the most consistent onomasiological alternate for concrete *give*. Data for abstract *give* and *provide* in the written and spoken sections of each corpus are provided in Figure 44 and Figure 45.
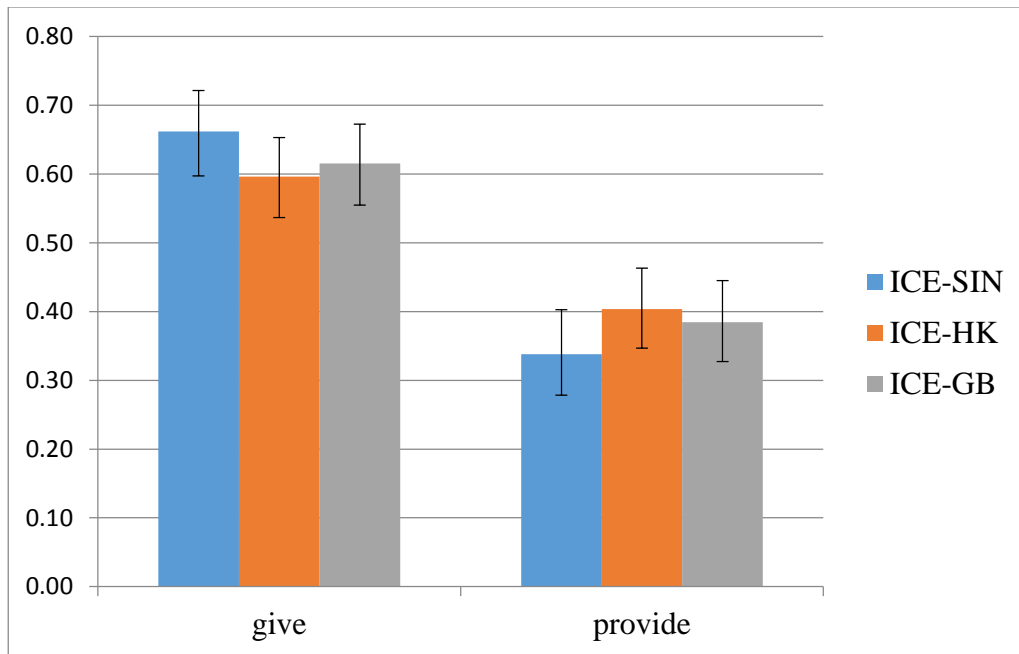
Figure 44: Abstract *give* and *provide* with the sense in the written portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
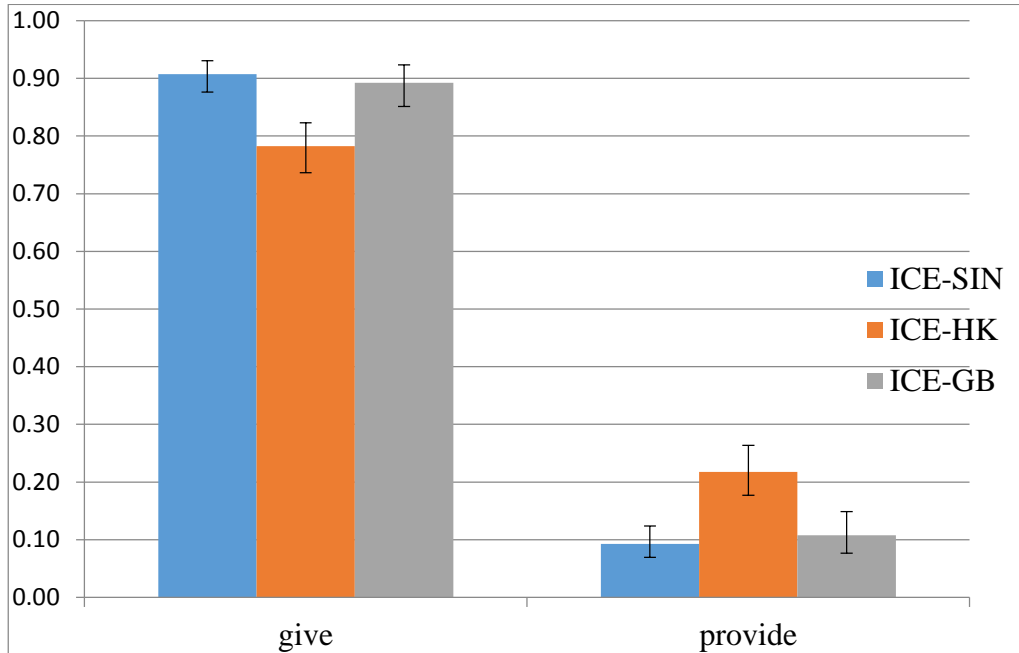
Figure 45: Abstract *give* in the spoken portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

Once again, all three corpora show a strong preference in spoken language for the colloquial, monosyllabic, Germanic *give* over its polysllabic Latinate alternate *provide*. That preference persists in written language: *provide* is more likely to occur in written language than in spoken language, but it is still significantly less likely than *give* in both speech and writing. In this case, all three corpora and the regions they represent seem to exhibit similar usage patterns for this semantic field.

8.2.10. Onomasiological analysis: Delexical

Each Delexical *give* construction alternates with a verb that relates to its Direct Object. However, unlike *make* and *take*, Delexical *give* also alternates consistently with another verb: *provide*. It may be that *provide* also has a Delexical sense, and Delexical *provide* overlaps significantly with Delexical *give*. Alternatively, it may be that *give* and/or *provide* are similarly vague across these two senses, as evidenced by the replaceability of *give* with *provide* for

both senses. This problem, in addition to the discussion of Delexical/abstract overlap in 8.2.7, might be seen as a challenge to viewing *give* as Delexical in the way that *make* and *take* seem to be Delexical. Similarly, Newman (1996) argued that *give* does not lose its lexical meaning in these apparently 'light' constructions. This section analyses Delexical *give* constructions alongside their related verb alternates, and alongside their related *provide* constructions. Existing studies of light verbs have not engaged with *give* in this way, and this approach allows me to move towards a fuller understanding of *give*.

Direct Objects that occur frequently enough, and whose alternates occur frequently enough, for statistical analysis include only *give support* and *give information*. A larger corpus would certainly provide more data for a broader analysis of such constructions, but a larger corpus would also require considerable additional manual analysis in recognizing and categorizing each individual instance of Delexical *give*, which could be prohibitively costly. *Give support*, *provide support*, and *support* (v.) are considered first.
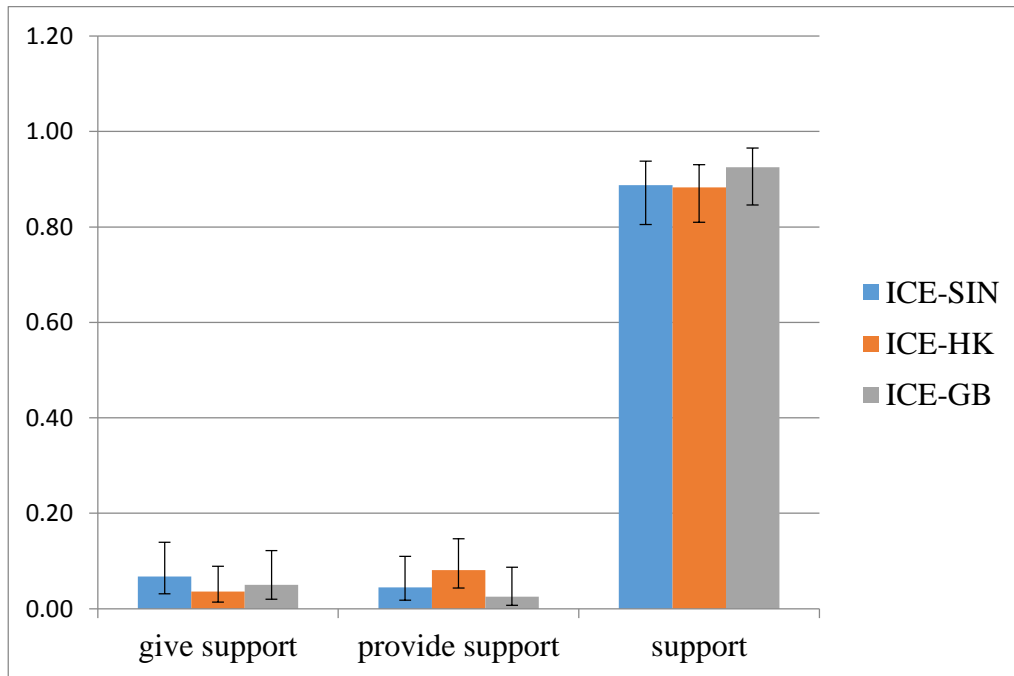
Figure 46: Instances of *give support*, *provide support*, and *support* in the written portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents

probabilities for each term in each corpus, and error bars represent Wilson intervals.
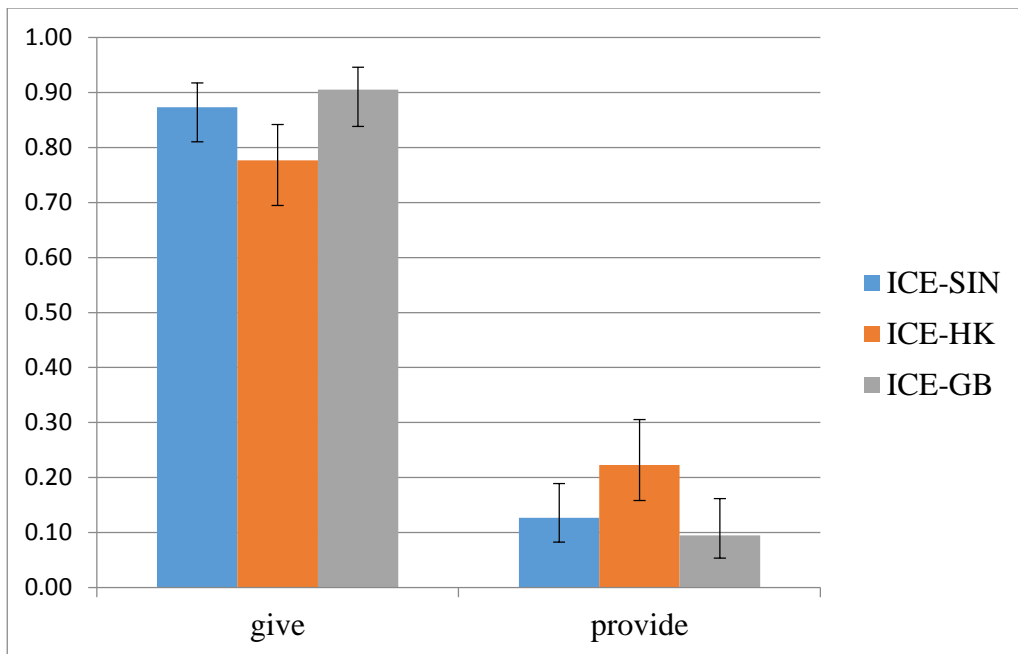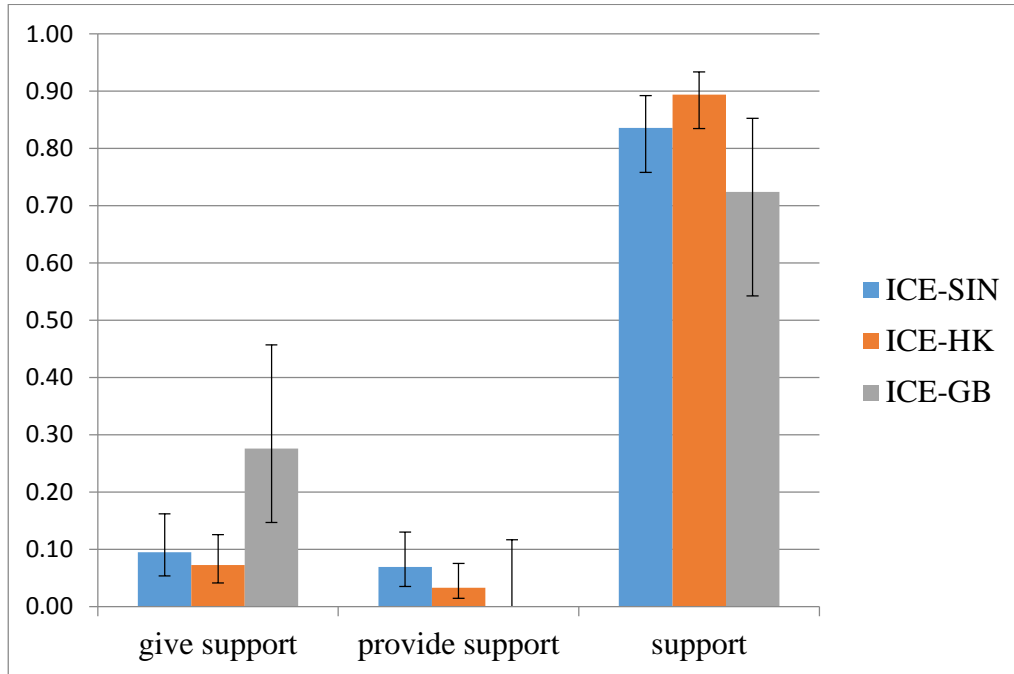
Figure 47: Instances of *give support*, *provide support*, and *support* in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

As shown in Figure 46 and Figure 47, *support* is significantly preferred over *give support* and *provide support* in all corpora in both speech and writing. *Provide support* and *give support* generally occur with similar frequency in speech and writing, with the exception that *give support* is significantly preferred over *provide support* in the spoken section of ICE-GB. In fact, as shown in Table 18, occurrences of *give support* in the spoken portion of ICE-GB are not particularly high, but *provide support* does not occur at all.

	give support	*provide support*	*support*
ICE-SIN	11	8	97
ICE-HK	11	5	135
ICE-GB	8	0	59

Table 18: Instances of *give support*, *provide support*, and *support* in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB.

It seems that ICE-GB significantly disprefers *provide support* in spoken language, in contrast to the other two corpora.

In speech, ICE-GB prefers *inform* significantly more than both *give information* and *provide information*, which are both preferred equally.
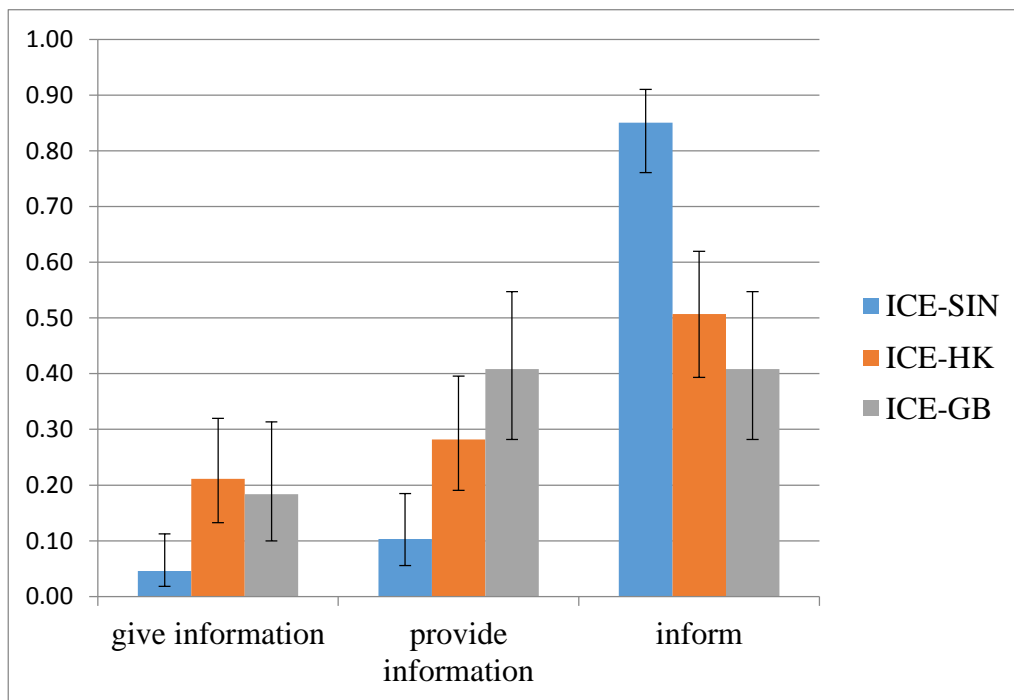
Figure 48: Instances of *give information*, *provide information*, and *inform* in the written portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
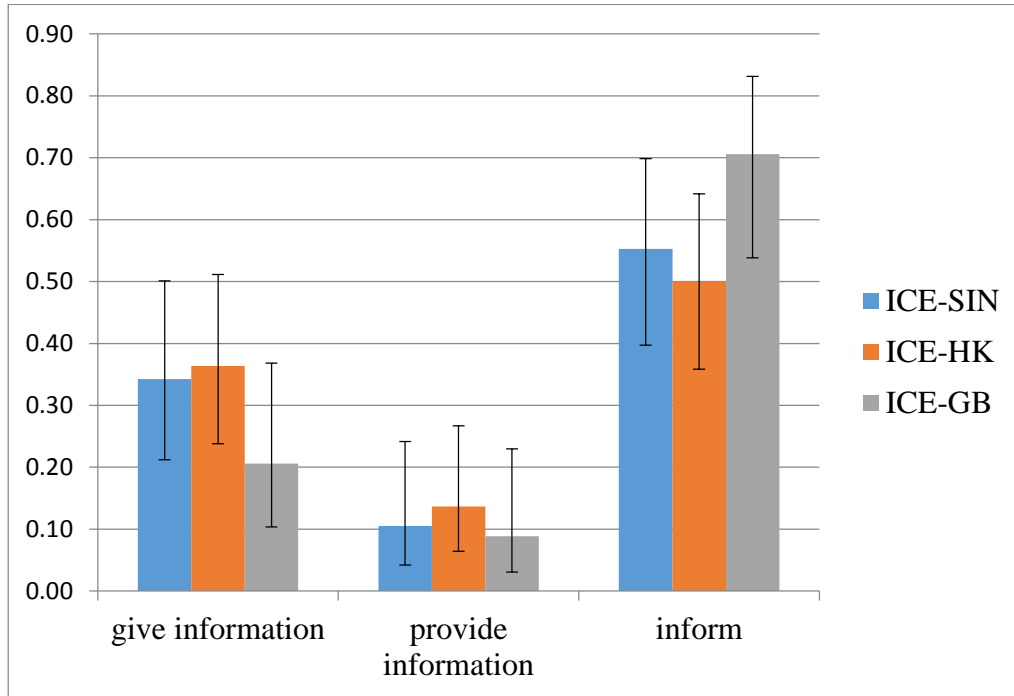
Figure 49: Instances of *give information*, *provide information*, and *inform* in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

In speech, ICE-SIN and ICE-HK, however, prefer *inform* and *give information* statistically equally, and *give information* and *provide information* statistically equally, with a significant difference between *provide information* on the low end and *inform* on the high end. In writing, ICE-GB prefers *provide information* and *inform* equally, and prefers both significantly more than *give information*. ICE-HK and ICE-SIN, however, significantly prefer *inform* to both *provide information* and *give information*, which are preferred equally. In this case, the distinctions are subtle and complex, but ICE-SIN and ICE-HK are more similar to each other than to ICE-GB.

8.2.11. Give: Summary

As with *make* and *take*, *give* does not exhibit any unique senses or sense distinctions in any individual corpus. Examples of *give* fit into the hypothesized

sense categories, with the exception that Delexical *give* does not seem to represent a discrete sense category to the degree that Delexical *make* and *take* do. Although *give* is often considered a typical Delexical verb, it does not seem to be Delexical in the ways that *make* and *take* can be. In particular, *give* readily accepts a coordinated Direct Object with both Delexical and non-Delexical elements; and Delexical *give* readily alternates with *provide*. *Make* and *take*, on the other hand, do not readily coordinate Delexical and non-Delexical elements in Direct Object position. Likewise, Delexical *make* and *take* do not alternate with any other verb. *Provide* is an alternate for both concrete and abstract *give*, as well as Delexical *give*. There are a few possible conclusions to be drawn from this: either *provide* is discretely polysemic in precisely the ways that *give* is discretely polysemic, both lexical items exhibiting a concrete, an abstract, and a Delexical sense; or both *give* and *provide* are in fact vague across two or more of these categories. If the latter is true, then Delexical *give* could be seen to alternate with *provide* precisely because some lexical meaning related to the concrete or abstract sense 'provide' is retained in Delexical *give*. At the very least, it can be concluded that Delexical *give* is different from the other two verbs examined here.

Usage preferences for concrete *give* are nearly identical to concrete *make*. In speech, all varieties prefer concrete *give* over its polysyllabic, Latinate alternates. In writing, preferences differ; this would seem to challenge the assumption that there is a common core to written English worldwide. In writing, as with *make*, it is ICE-HK that differs from the expected norm. Both ICE-GB and ICE-SIN prefer the polysyllabic, Latinate alternate *provide* in writing, while ICE-HK strongly prefers concrete *give* over *provide*, even in writing. This suggests an internally coherent genre-related preference in ICE-HK that differs from the other datasets.

As discussed above, it is clear that Delexical *give* differs semantically from Delexical *make* and *take*. That difference is explored further in 9.1.

9. CONCLUSIONS

9.1. The lexis: make, take, and give

Findings for *make*, *take*, and *give* indicate that there are no unique senses or sub-senses of these verbs in regular use in any individual region represented by the corpora. The lack of unique senses or sub-senses is remarkable given the numerous hypotheses in lexicological and lexicographical research proposing that very subtle semantic variation in use is likely to be found across World Englishes (Lambert 2012: 307; Brutt-Griffler 2002: 153-4; Hymes 1996: 9; Görlach 1995 [1990]: 127; Platt *et al*. 1984: 105; see also 4.3).

Some variation in lexical meaning hypothesized in existing literature (*cf*. Adejare and Afolayan 1982, quoted in Platt *et al*. 1984) was discussed in 4.3 and excluded from the type of semantic variation examined here. Those proposals suggest that variation in encyclopedic knowledge related to individual lexical items should be treated as semantic variation. The present study has adopted a stricter definition of lexical semantics. For example, the present study has not investigated real-world variation in the use of *take* in *taking tea* insofar as *taking tea* can typically involve different concrete objects and actions, customs, and social and cultural variables in different locales. Such variation is not considered lexical semantic variation in the present study.

Make, *take*, and *give* were selected for this study because they are highly frequent and highly polysemic, attributes which, hypothetically, should constitute a rich ground for subtle but observable semantic variation in use. The three verbs were also selected in part because they are considered typical light verbs, with both concrete and light or Delexical senses. It was hypothesized that semantic variation would be observed primarily in abstract or light uses of these three verbs: these non-concrete uses might be expected to be more prone to subtle and slight semantic change than concrete vocabulary, given the lack of directly observable concrete referents for abstract and light constructions. Such directly observable concrete referents might have been expected to 'ground' concrete usage in a more stable worldwide norm. In fact, perhaps surprisingly,

light usage is very uniform across the three regions: in most cases of light usage, the related verb is strongly preferred over the light construction in both speech and writing. Exceptions occur, but not with any internal coherence or consistency – it is possible that the exceptions are the result of some unidentified variable.

Onomasiological variation is most pronounced in concrete senses of the three verbs. With concrete *make* and *give*, when differences appear between the three corpora, it is ICE-HK that is unique, in that it does not follow the general, expected, traditional trend of preferences for monosyllabic, Germanic alternates in speech and for polysyllabic, Latinate alternates in writing. Concrete *take* differs from concrete *make* and *give*. In fact, concrete and abstract *take* are quite similar to each other. In both cases, as expected, all corpora prefer *take* over its alternates in speech. In writing, ICE-GB prefers *take* over its alternates, while ICE-SIN and ICE-HK show an equal preference for *take* and for the polysyllabic, Latinate alternates *collect* (concrete) and *adopt* (abstract). It might be maintained that the findings for *take* are not actually dissimilar from the findings for concrete *make* and *give*: for *take*, ICE-HK, this time along with ICE-SIN, does not exhibit the stylistic, register-based difference between the monosyllabic, Germanic alternates and the polysyllabic, Latinate alternates that is exhibited in ICE-GB. In this interpretation, ICE-GB exhibits sharp differences in selection processes between speech and writing; ICE-HK is consistently different from ICE-GB in its selection preferences; and ICE-SIN is sometimes more similar to ICE-GB, sometimes more similar to ICE-HK. While that might be a satisfactory conclusion in a general sense, it fails to recognize that for *take*, ICE-GB significantly prefers the Germanic *take* in both speech and writing, whereas for *make* and *give*, ICE-GB significantly prefers the Germanic alternate in speech, the Latinate alternate in writing. That is, the preference in ICE-GB, representing British English, seems not to be actively influencing usage in ICE-SIN and ICE-HK. It might have been concluded from the general evidence on usage preferences that ICE-HK tends to prefer the 'simple' Germanic, monosyllabic alternate at least as much as the 'complex'

Latinate, polysyllabic alternate in all situations (preferring *make* at least as much as *produce*, for example, even in written language), and that that preference pushes back, as it were, against the stylistic norms of written British English, such that the varieties are engaged in a sort of dynamic interrelationship. For concrete *make* and *give*, that hypothesis could be maintained. But *take* complicates the picture: why does ICE-HK not prefer the 'simple', monosyllabic *take* over its alternates in writing? If the preference for the simple option in ICE-HK were pushing back against the stylistic norms of British English, then the ICE-HK would mimic or even exaggerate the ICE-GB preference for *take* in writing. The ICE-GB preference for *take* in writing would not allow for an increased preference for *collect* (concrete) and *adopt* (abstract) in ICE-HK writing – but that is exactly what we find. This finding may suggest that the preferences for the monosyllabic, Germanic alternate and the polysyllabic, Latinate alternate in writing are an independent feature of ICE-HK (and, sometimes, ICE-SIN), rather than any sort of push-back against British English norms. This argument might constitute a useful working hypothesis for future research. To sum up, one of the crucial and surprising findings of the present study is that variation in onomasiological preferences tends to arise in concrete senses rather than abstract or Delexical ones.

In addition to the question of variation between the corpora, the present study also reached findings on variation between the three words. To reiterate, the three verbs were selected for this study in part because they are considered typical light verbs. However, each corpus displays differences between Delexical *make*, *take*, and *give* in remarkably consistent ways. Rather than finding a standard Delexical sense across *make*, *take*, and *give*, the present study instead uncovered notable differences, suggesting something like *degrees of Delexicality*. Those differences were only discernible via the three key methods employed in this study: manual semasiological analysis; a consideration of traditional polysemy tests and their corresponding forms of corpus evidence; and onomasiological analysis. To summarize:

a. *Give* is highly ambiguous between its Delexical and abstract senses in all three corpora. Identity evidence suggests relatively common overlap between the Delexical, abstract, and concrete senses (e.g. *give comments and ideas*; *give the dog steak and a good run*; see 8.2.4 and 8.2.5). Moreover, Delexical *give* alternates not only with a related verb (as *give support* alternates with *support* (v.), for example), but also with *provide* (as *give support* also alternates with *provide support*). *Provide* also alternates with *give* in abstract and concrete senses. *Give* does not seem so light, in Jespersen's (1954) sense, after all.

b. *Make* is hypothetically ambiguous between Delexical and abstract senses insofar as most Delexical constructions can be conceived as producing some abstract result (as *make a decision* alternates with *decide*, but can also be conceptualized as *produce a decision*). There is very little antagonism evidence, autonomy evidence, identity evidence, or truth-condition evidence for the overlap in practice, however. Delexical *make* also overlaps not too rarely with concrete *make* (e.g. *make a mark*).

c. *Take* is not ambiguous between Delexical and abstract senses at all. There is rarely any conceivable conceptual overlap, and there is no overlap in actual use in the corpora. There is also no overlap between Delexical and concrete *take*. Delexical *take* seems to be more discrete from other senses of *take* than is the case for *make* or *give*.

As *give* appears to be the least Delexical of the three verbs in question, it is also worth recalling that *make* and *take* appear to alternate with each other in Example 76, reprinted below as Example 186.

186. The staff member shall not: 1) Take or permit to be **made** any alterations in the internal construction or arrangements or in the external appearance or in the present

scheme of decoration of the premises. [ICE-SIN W2D-003 #130-1]

It is also worth noting that *make* and *take* both occur Delexically with the Direct Object *decision*, and it seems that there is rough equivalency and the general ability to alternate between those constructions. Delexical *give*, on the other hand, shows no overlap in use with either Delexical *make* or Delexical *take*. Even if *take* appears to be more discretely Delexical than *make*, the two seem to overlap with each other more than with *give*.

It might also have been the case that the three corpora showed varying degrees of Delexicality for the same verb: one corpus might have shown unique ambiguity or vagueness between the abstract and Delexical senses of a particular verb. However, these degrees of Delexicality are remarkably uniform. That such uniformity should occur with such fine nuances of meaning is noteworthy – such subtle semantics could have been expected to vary between the regions. In fact, this raises questions as to the mechanisms that might allow for such consistency: how are such fine nuances maintained across such different geographies, and across such different sociolinguistic circumstances? Much World Englishes research focuses on differences between regions, and there are theoretical models for explaining many differences that are observed, but an explanation of consistency and similarity in fine distinctions such as these, which operate well below the level of consciousness, is not available in existing theoretical frameworks.

The idea of degrees of Delexicality challenges traditional notions of light or Delexical verbs (though Poutsma 1926: 394-400 does in a rather unclear way refer to degrees of vagueness in these constructions; see 3.4). In the more traditional accounts of light verbs, Poutsma's (1926: 394-400) and Jespersen's (1954: 117) early identification of light constructions was largely reflective. Wierzbicka's (1982) detailed analysis of light constructions with *have* showed that a great deal of meaning was retained by *have* in light constructions, and she argues that light verbs are in fact not so semantically

light. Similarly, Newman (1996) argues that the prototypical meaning of *give* is retained in light constructions. The present findings corroborate Newman's finding, but it is not at all clear that all light verbs retain such lexical meaning, as Newman (1996) and Wierzbicka (1982) suggest – in fact, it seems that *take* conveys very little meaning in these light constructions. It may be that a spectrum exists between verbs like *take* that seem to retain very little meaning in light constructions, and verbs like *give* that retain quite a bit of their lexical meaning in light constructions. In that case, the notion of degrees of Delexicality should be useful for linguists moving forward with this kind of research. The present study might set a useful template for measuring degrees of Delexicality: corpus work will be crucial, and the tools derived from traditional polysemy tests in the present study should prove useful as well. Perhaps, building on the present template, even more reliable tools might be developed and adopted for identifying and establishing this category of verbs in the future, including experimental psycholinguistic tools. In addition, a careful cognitive approach such as that taken by Newman (1996) and Wierzbicka (1982) for *give* and *have*, respectively, could be applied to *take* and *make* as well; that would be a valuable complement to the present findings. Moving forward, then, I propose that the idea of degrees of Delexicality will be a valuable one.

9.2. Impact: World Englishes

Chapter 4 discussed multiple theoretical frameworks for World Englishes, and the place of lexical semantics within World Englishes discourse. One of the key research questions in the present study is: to what extent do the present findings corroborate or refute established frameworks of World Englishes? Throughout chapters 6, 7, and 8, I have been careful to discuss *corpora* rather than *varieties*, referring to data in ICE-SIN, for example, rather than to Singapore English. As discussed in 5.2, the corpus is not the variety – the sample is not the population. At this stage, it becomes useful to carefully abstract from the sample to the population from which the sample was drawn, keeping in mind precisely what

each sample represents, and maintaining nuance in the conclusions drawn about each variety. In 9.1, I proposed that the semantic variation that does exist between the corpora tends to be variation in usage preferences between monosyllabic, Germanic verb forms and polysyllabic, Latinate verb alternates – this is most pronounced in concrete senses, but there is also evidence for this variation in abstract senses. I also proposed in 9.1 a possible interpretation of this evidence: that the equal preference for each alternate in ICE-HK, and occasionally ICE-SIN, might be an independent feature of Hong Kong English, and occasionally Singapore English, rather than a dynamic response to the norms of British English.

To what degree can broad conclusions about each variety be drawn from the present findings? There are several issues to be borne in mind here. First, and most generally, the findings do not incorporate the contributions of English language learners, with the exception that ICE-HK may include English language learners in inconsistent ways, due to the impossibility of reliably identifying English language schools in Hong Kong. The probable presence of English language learners, unique to ICE-HK, is discussed at greater length below. In addition, numerous variables are uncontrolled in the corpora: gender identification, age, education, social class or socio-economic status, and racial identification.[39] Because those details are uncontrolled, it is impossible to determine whether variation that appears to correlate with each region might in fact correlate with any of those variables. For example, it might be that the unique preferences for concrete usage in ICE-HK are not related to Hong Kong English as such, but instead to unique variation in some uncontrolled and unreported variable, such as, for example, a particularly high number of university-educated speakers in ICE-HK. Moreover, it should be noted that other types of variation might be found within each region represented (see 5.1)

---

[39] As noted in 5.2, the corpus compilers of ICE-GB have reported gender, age, and education of language users represented, but the corpus is not designed to balance those factors in any way; the compilers of ICE-SIN and ICE-HK have not reported those details.

– for example: between Singlish and standard Singapore English; between the English of Tamil speakers and others in Singapore; between northern and southern speakers in Great Britain, and so on. That potential variation would be indiscernible in the present study.

Just as importantly, the contributions of foreign speakers and writers living in each locale are not reflected in the present findings, with the exception of ICE-HK. As discussed in 5.2, ICE-HK includes a high number of non-local speakers in its face-to-face conversations. Speakers of Hong Kong English can be expected to converge towards foreign norms to some degree when speaking with foreign interlocutors. The foreign speakers in ICE-HK have been described as generally representing inner-circle varieties (p.c. Kingsley Bolton 2012). If foreign speakers have influenced the sample in that way, the result might be a similarity between spoken data in ICE-HK and spoken data in ICE-GB. In that case, a Hong Kong corpus without foreign interlocutors would perhaps have differed even more from ICE-GB. Likewise, without foreign interlocutors in ICE-HK, it might have been the case that ICE-HK speech differed from ICE-GB speech just as ICE-HK writing differs from ICE-GB writing – but we cannot be sure, given the present datasets.

The present study, thorough as it is, represents only a sliver of the evidence for lexical semantic variation in World Englishes. That said, it is still worth considering whether existing frameworks for World Englishes can account for the present findings on lexical semantic variation, and whether the present lexical semantic evidence corroborates existing theories. As discussed in 4.3, existing theories of World Englishes do not tend to focus on lexical semantics as a defining feature of their frameworks. In the present study, the lexical semantic data does not consistently match or support any of the existing models of World Englishes. In a way, this should not be surprising – the theories are generalizations and cannot be expected to describe every detail of every feature of World Englishes. In general, there is no semantic variation in *make*, *take*, or *give* in the spoken portion of the corpora: that finding is discussed further below. In writing, where variation in onomasiological

preferences does occur, ICE-HK tends to differ significantly from ICE-GB, while ICE-SIN tends to be similar to ICE-GB, with occasional greater similarities to ICE-HK. This variation might suggest a continuum with ICE-GB on one end and ICE-HK on the other, and ICE-SIN in the middle. Schneider's (2007; see 4.1) model, however, would have predicted something quite different: his framework places a relatively endo-normative Singapore English furthest on a continuum from British English, and a relatively exo-normative Hong Kong English close to British English. His model predicts that Hong Kong English will follow the external norms set by British English, while Singapore English will set its own internal norms which may be quite different from British English. Kachru's (1985; see 4.1) model groups Singapore English and Hong Kong English together as outer circle varieties, with British English an inner circle variety. Kachru (1992), however, acknowledges that the role of English in outer-circle varieties changes over time. Singapore English may certainly be seen as developing from a non-native exonormative variety to a native endonormative variety soon after Kachru's (1985) model was proposed, with a switch towards universal English-language schooling initiated in 1983 (Deterding 2007: 86) and finalized in 1987 (Ling 2010: 232). In Hong Kong, on the other hand, English was made available via a limited number of English-language schools following the educational reforms of 1978 (Bolton 2006: 2). In 5.2, I discussed the broad variation in actual use of English within nominally English language schools in Hong Kong. By 1997, roughly when ICE-HK was compiled, 100 out of 460 secondary schools in Hong Kong were allowed to teach in English (Bolton 2006: 9), with an unknown number of those schools adhering consistently to English for all purposes. In addition, the 1990 Singapore census reported that more than 20% of Singaporeans used English in every aspect of life (Tickoo 1996: 431); in 1987, Fu (quoted in Bolton 2002: 41) reported that 98% of Hong Kong residents speak Cantonese at home, and that English was only used as a non-native language by the vast majority of residents. Nonetheless, in 1993, Hong Kong residents claiming a 'reasonable' command of English reached 33% (Bacon-Shone and Bolton 1998). Measuring

the number of competent speakers of a language, or even native speakers, is notoriously difficult (*cf*. Bolt and Bolton 1996: 200), and these figures reflect that difficulty. Nevertheless, what is certain is that ICE-HK is more a corpus of English as a second or foreign language than ICE-GB and ICE-SIN are. Unlike ICE-GB and ICE-SIN, the compilers of ICE-HK explicitly required that language users in the corpus be users whose first language is Cantonese, and who would normally use Cantonese beyond the home environment (Bolt and Bolton 1996: 199). In that case, it may be that the primary/non-primary language distinction is the most important one in relation to lexical semantics and usage variation here. Schneider (2007) predicts that in the early stages of the development of World English varieties, the external norm of British English will set the standard for the newly developing variety of World English. But in these early stages, which Schneider (2007) sees as exo-normative, processes of acquisition of English as a foreign language must play a role as well. The framework proposed by Brutt-Griffler (2002; see 4.1) is particularly relevant here. Interestingly, Brutt-Griffler is also one of few theorists to have expounded on the significance of lexical semantic variation in studying World Englishes (see 4.3). She predicts that in the non-native stage of development of a variety of World English, there is little sense of the nuances of an exo-normative standard. In such a model, a non-native variety like Hong Kong English would not be expected to conform to the subtle usage norms of written British English, while a native variety like Singapore English would be expected to at least have a sense of those norms, even if those norms are sometimes broken. Brutt-Griffler's framework would seem to support the discussion in 9.1 on the lack of a 'push-back' from Hong Kong English against British English written norms; according to her theory, language users in Hong Kong would not push back against those norms because, as language learners, they would not be fully aware of those norms. Thus, the preference in ICE-HK against the polysyllabic, Latinate alternates in writing would be an independent development in Hong Kong rather than a dynamic response to British English standards. This perspective contrasts considerably with Schneider's (2007)

perspective on Hong Kong as an exo-normative variety dependent on British English norms as a standard.

Both Schneider (2007) and Brutt-Griffler (2002), and indeed most World Englishes theorists, refer to the interaction between the local language and English as a critical influence: in that case, it would be worth investigating the potential influence of Cantonese on the usage preferences exhibited in ICE-HK. Of course, Cantonese does not offer a choice between monosyllabic, Germanic alternates and polysyllabic, Latinate alternates, but a consideration of semantic alternation in relation to genre, and in relation to semantic concreteness and abstractness, could prove fruitful.

The differences observed between the corpora also relate to the difference between speech and writing. Writing is generally viewed as a conservative force, and standards of written English worldwide are generally expected to be stricter than norms for spoken English. However, that assumption does not generally consider the process of acquiring English in a speech community, as Brutt-Griffler's (2002) model does, and the process of learning genre-related norms and standards, which can be seen as quite subtle. The present findings suggest that lexical semantic variation between regions, at least in high-frequency verbs, is more pronounced in writing than in speech. In this case, writing seems to be the locus of innovation, in the form of innovative genre norms, rather than of conservatism.

On the differences between speech and writing in World Englishes, Hundt (2006) has noted less variation between writing and speech in ICE-SIN and ICE-Philippines than in ICE-GB in terms of collective noun concord. Hundt (*ibid*.: 233) interprets these findings as evidence of the exo-normative force of written British English: collective noun concord in both the written and spoken portions of ICE-SIN and ICE-Philippines resembles the written portion of ICE-GB. Hundt attributes this phenomenon to the practice of learning English from written sources, and to the role of written sources in establishing an external norm. However, it is important to note that a similarity between two corpora or sub-corpora is not necessarily an indication of an influence: Hundt

(*ibid.*) does not prove an influence, only a similarity, and similar features may develop in different varieties for different and independent reasons. Hundt (2009) notes a major difference between written and spoken portions of ICE-GB in use of *get*-passives, and limited difference between written and spoken portions of ICE-Philippines, with ICE-SIN falling in between the other two corpora, but that study correctly notes the limited conclusions that can be drawn from semasiological data calculated *per million words*. The present study observes a larger difference between speech and writing in ICE-GB, and occasionally ICE-SIN, than in ICE-HK, but the present study provides clear evidence that this is not because ICE-HK follows written British English norms − quite the opposite. ICE-HK shows less difference between speech and writing precisely because it differs from the apparent norms of written British English. The present study contrasts with Hundt's (*ibid.*) findings. In the present findings, ICE-HK shows unique, possibly innovative features in writing but not in speech, and certainly does not adhere to the norms of written British English.

As discussed in Chapter 4, some hypotheses about semantic variation in World Englishes have suggested that essential cultural differences could give rise to subtly differing semantics that would go under the radar of everyday use. The actual semantic variation that is evidenced here seems to be onomasiological variation related to usage preferences, genre norms, and the selection process for different lexical units that can fill a given semantic slot. One interesting feature of these usage preferences is that preferences for Delexical senses of polysemic verbs vary so little between corpora and regions. This is a remarkably robust 'common core' based on usage and practice that is observable in corpora, but that language users would likely be unable to express or codify. Baker (2010: 83) has commented that similarities between language varieties are crucial, and must be explained; otherwise, researchers risk 'bottom drawer syndrome', whereby a majority of studies evidences similarity between varieties, but only a minority of studies that evidences differences is published. Owusu-Ansah (1994: 341) provides a strong example of a reflective study that aims 'to look at non-native varieties from the viewpoint of how they are similar

to native varieties'. He asserts that 'a model of NNEs [Non-Native Englishes] that highlights only the differences between them and NEs [Native Englishes] to the neglect of similarities is a distortion of facts'. Owusu-Ansah's (*ibid.*) study may be an exception to the norm: bottom drawer syndrome may very well be a reality. The present study, then, can also be seen as acknowledging critical similarities between varieties – these three varieties seem to represent a common core of semantic senses for these three verbs, as well as remarkable similarities in usage of light verb constructions and abstract senses.

In a potential future theory of World Englishes built on data representing lexical semantic variation, it might be hypothesized first that a common core does indeed exist in general, even if noteworthy exceptions such as Singapore English *get* and the usage preferences found here for concrete, high-frequency verbs arise as well. In addition, it might be that the native/non-native speaker distinction or the primary/non-primary language distinction could correlate most consistently with what general, broad semantic variation does exist in use. It would be exciting to see a semantics-based model arise in the way that Schneider (2008) established a phonetics-based model. Such a model would allow for far greater precision than hypotheses about native/non-native distinctions, instead showing clusters of varieties that display similar semantic preferences, regardless of social, political, or geographical distinctions. The present findings on usage preferences for concrete, high-frequency verbs might then constitute just a few data points towards that future model. Further study into concrete, high-frequency verbs and other word classes is warranted, and particularly into preferences for monosyllabic, Germanic alternates in speech and polysyllabic, Latinate alternates in writing.

9.3. Impact: Cognitive Linguistics

A key discussion in Cognitive Linguistics has revolved around the relationship between corpus frequencies and cognitive salience. Various researchers have hypothesized a link between corpus frequencies and cognitive salience, but the nature of that link remains unclear, and Gilquin (2008) showed that there may

in fact not be evidence for such a link (see 3.4). As discussed in 3.4, one of the fundamental, but under-discussed, issues in the literature is the nature of corpus frequency – the definition of frequency has tended to be unstated, with an assumption that *frequency* is *semasiological frequency* and/or *frequency per million words*. Geeraerts (2010) has offered an alternative in the form of his hyopthesis of *onomasiological salience*, which suggests that relative onomasiological frequencies should correlate with cognitive salience. In this section, I consider the present data in light of these questions and issues.

Gilquin (2008) refers to spoken data in the Switchboard Corpus and found that light senses of *give* were approximately four times more common than concrete senses, and that light senses of *take* were approximately twice as common as concrete senses. Gilquin has not published the raw data from her analysis, so a more precise report of frequencies is not possible. Werner and Mukherjee (2012) analyse senses of *take* and *give* in ICE-GB, ICE-India and ICE-Sri Lanka, and find that light senses of *take* and *give* are two to four times more common than concrete senses. The ICE-GB data presented by that study cannot be reliably compared to the present analysis of *take* and *give* in ICE-GB: Werner and Mukherjee (*ibid*.) draw instances from a 230,000-word sub-section of writing in ICE-GB, but they do not report raw numbers. Instead, they report instances *per 230,000 words*: reported frequencies should therefore be expected to be integers, representing actual instances from the corpus. However, their reported frequencies are given with two decimal places. It may be that the subsections of each corpus were not precisely composed of 230,000 words, but only approximately 230,000 words, and that raw numbers were therefore adjusted or normalized *per 230,000 words*, but this is not stated clearly. As a result, it is impossible to compare their findings with the present findings. Werner and Mukherjee (*ibid*.) rely on the sense categories established by Gilquin (2008), with only one small exception. Those sense categories are not thoroughly explained or defined in either study. In the present study, semasiological measures vary slightly between words and between regions, specifically for the verb *take*. The semasiological graphs in 6.2.1, 7.2.1, and

8.2.1 are redrawn below to isolate the frequencies of concrete and light senses; these can be compared with Gilquin's (2008) and Werner and Mukherjee's (2012) conclusions. Although Gilquin (2008) and Werner and Mukherjee (2012) do not precisely state a research question in this way, these graphs can be seen to answer the following question: given that a listener or reader is exposed to the concrete or Delexical sense of these verbs, at what rate is the listener or reader exposed to one or the other?
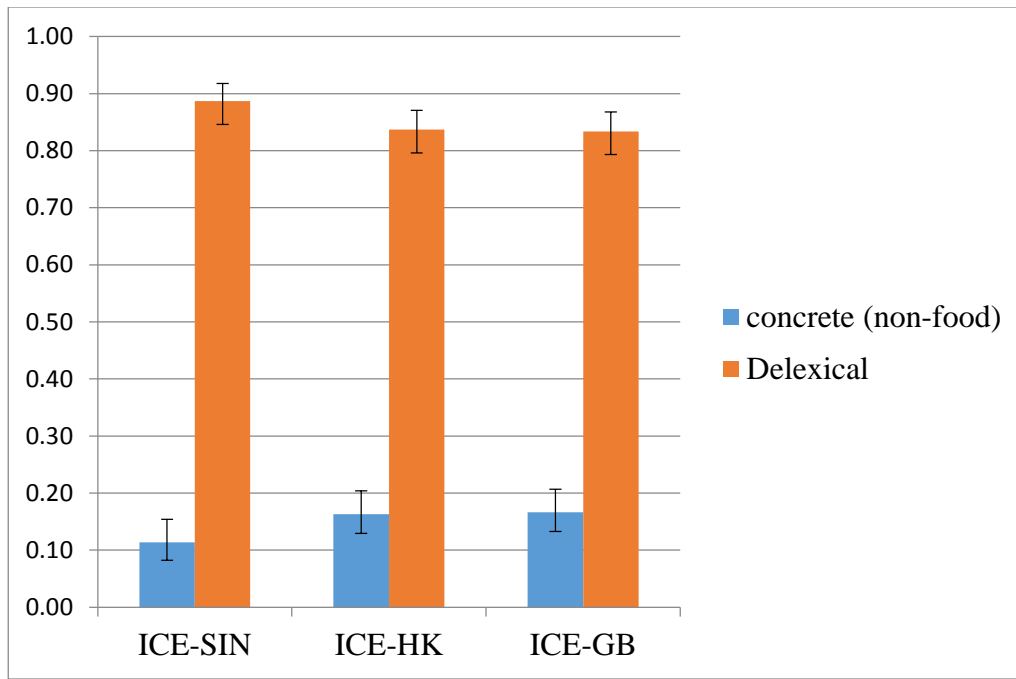
Figure 50: Probability of being exposed to *make* with the Concrete (non-food) sense or the Delexical sense in the written portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

Figure 51: Probability of being exposed to *make* with the Concrete (non-food) sense or the Delexical sense in the spoken portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

Figure 52: Probability of being exposed to *take* with the Concrete (non-food) sense or the Delexical sense in the written portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

Figure 53: Probability of being exposed to *take* with the Concrete (non-food) sense or the Delexical sense in the spoken portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

Figure 54: Probability of being exposed to *give* with the Concrete (non-food) sense or the Delexical sense in the written portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
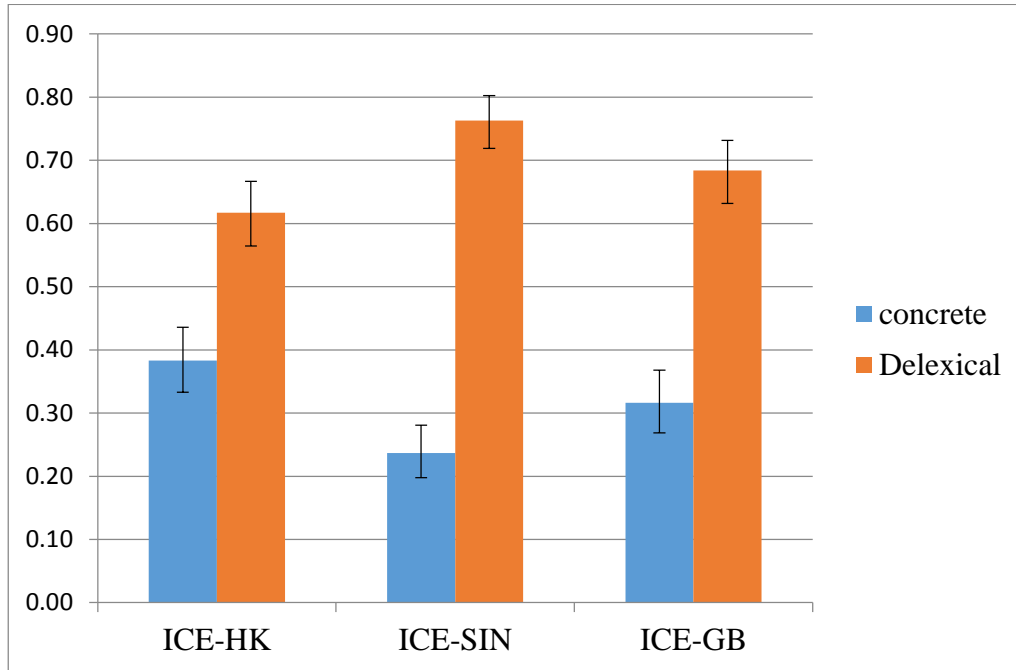
Figure 55: Probability of being exposed to *give* with the Concrete (non-food) sense or the Delexical sense in the spoken portion of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

Figure 50 through Figure 55 show that light senses are usually significantly more frequent than concrete senses of each verb. This is true of both *make* and *give*, in speech and writing, and it corroborates Gilquin's (2008) findings. The interesting exception to this trend is visible in Figure 52 and Figure 53, representing *take* in ICE-GB. In writing, shown in Figure 52, the Delexical sense is statistically indistinguishable from the frequency of the concrete sense. The error bars representing those senses overlap, and a Newcombe-Wilson test with continuity correction affirms the lack of a significant difference between the numbers. In speech, shown in Figure 53, the concrete sense is significantly more frequent than the Delexical sense. Those error bars do not overlap. This finding is robust – it does not appear to be the result of any idiosyncratic texts or any unusually common Direct Objects. Werner and Mukherjee (2012) found that concrete *take* was uniquely highly frequent in their sub-section of writing

from ICE-GB, so the present findings seem to corroborate their findings. Despite the unique case of *take* in ICE-GB, concrete senses in the ICE corpora examined here generally do tend to be among the least common senses semasiologically – concrete senses certainly tend to be less common than Delexical senses. However, as discussed in 2.2 and 2.3, this semasiological baseline measures probability in terms of exposure rates; the probability for producing each verb with each sense, given the actual opportunity for each sense to arise, might be very different. That probability can only be expressed with an onomasiological baseline, as presented below.

One key question highlighted in 3.4 was: Given that language users are exposed to light senses of verbs far more often than concrete senses, why should concrete senses come to mind first in elicitation tests? The first point in addressing this question must be methodological: if salience does not seem to correlate with frequency, then perhaps our measurements of salience or frequency are not being taken appropriately. Considering *salience* first, it is possible that laboratory settings are not ideal for measuring salience: the apparent high salience of concrete senses in elicitation tests may be an artefact of the artificial laboratory environment. In addition, it may be that semasiological elicitation tests like Gilquin's (2008) are less effective than onomasiological measures of salience – or that onomasiological measures would at least be a valuable complement to semasiological ones (see 3.4). Those experimental issues cannot be fully addressed within the scope of the present study. However, questions of the nature of *frequency* are very much the focus of the present study. It is possible that onomasiological frequencies are in fact more useful measures than semasiological frequencies. Geeraerts's hypothesis of *onomasiological salience* (2010; see 3.4) is extremely relevant here. Onomasiologically, concrete *make*, *take*, and *give* are significantly more common than their alternates in speech.

Figure 56: Probability of selecting concrete *make* and *produce* in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
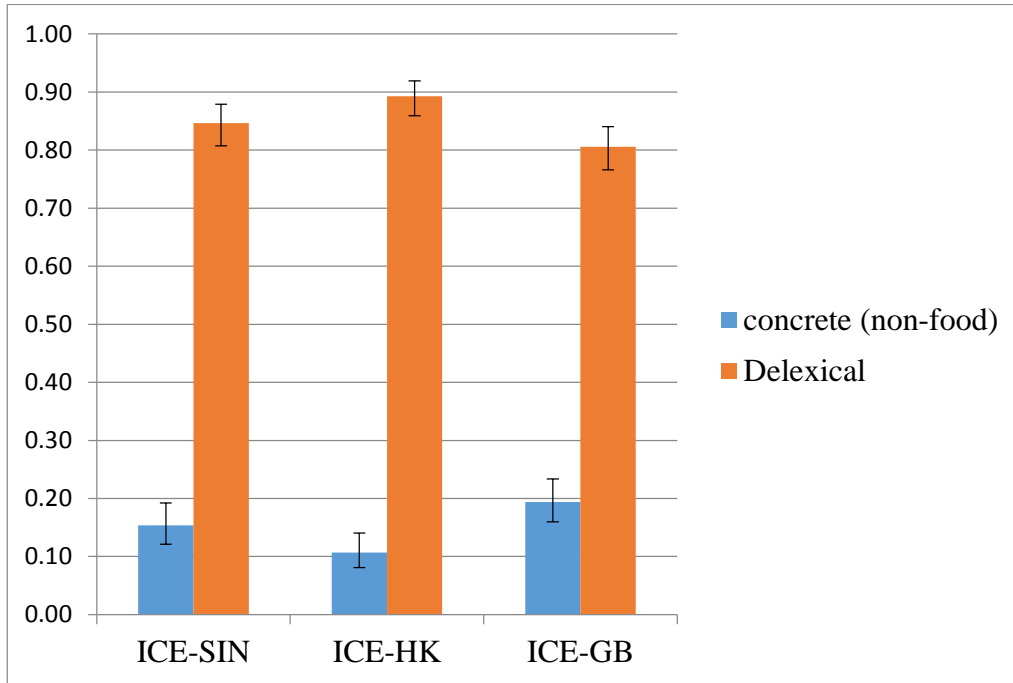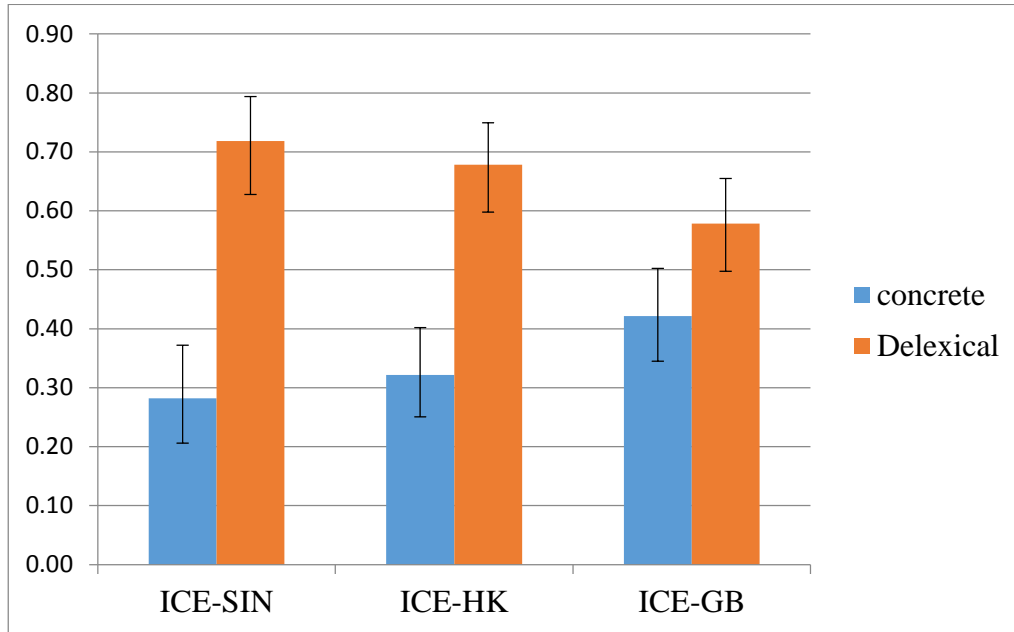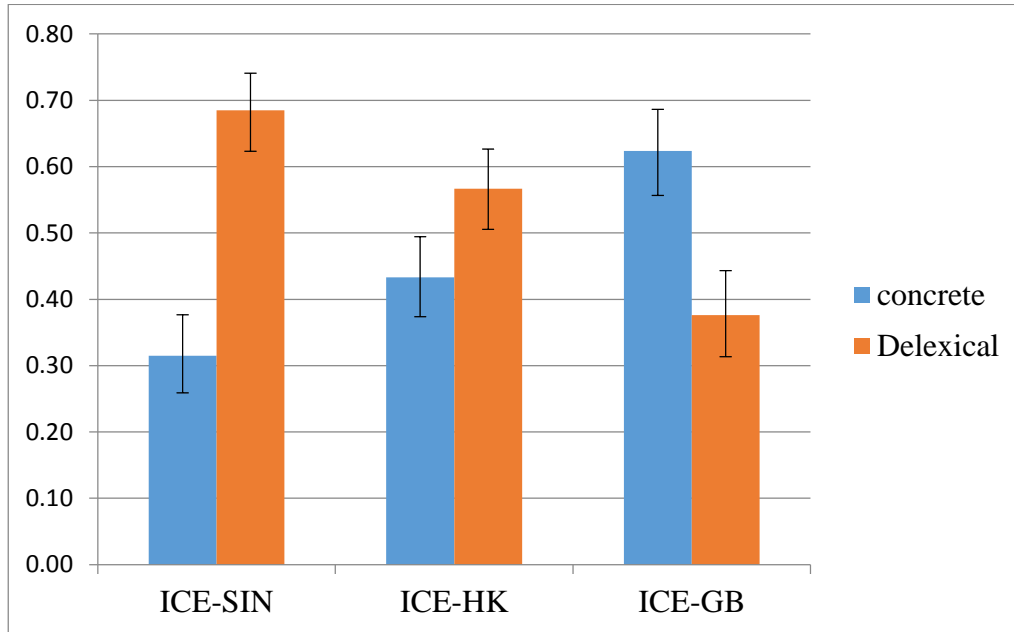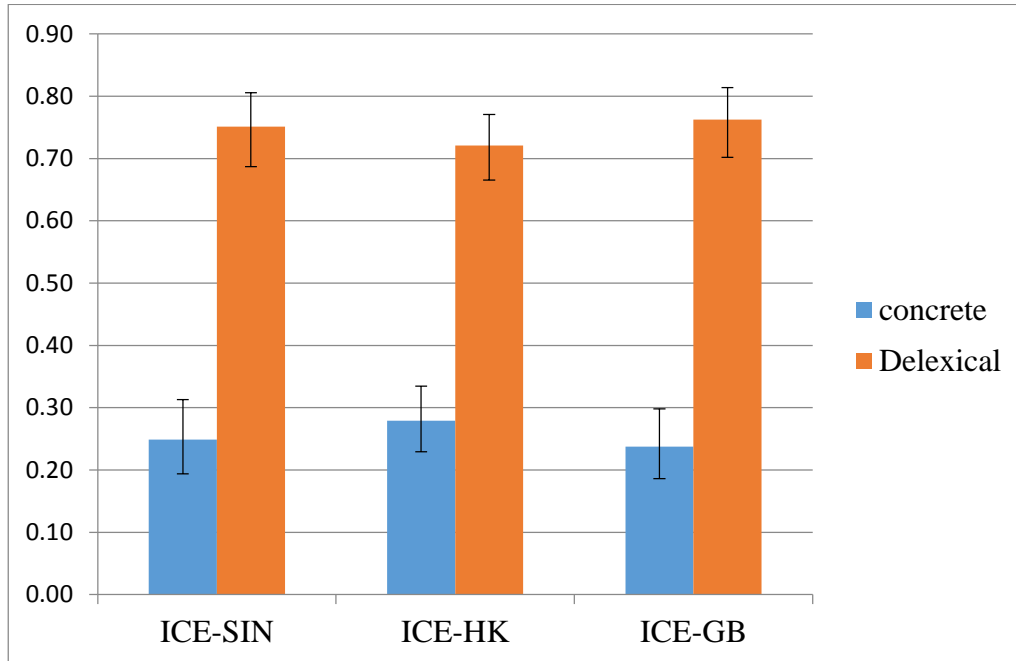
Figure 57: Probability of selecting concrete *take*, *collect*, *carry*, and *transport* in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

Figure 58: Probability of selecting concrete *give* and *provide* in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

Figure 56 through Figure 58 show that in speech, the verbs *make*, *take*, and *give* are significantly more common than their semantic alternates in the concrete sense. This seems to correlate with the high cognitive salience of the concrete sense of verbs like *take* and *give*, as measured by Gilquin (2008). That is, cognitive salience does in fact correlate with corpus frequency when frequency is measured onomasiologically in spoken language. There are two important caveats to that claim: first, Gilquin's findings on the salience of the concrete sense were determined semasiologically, and it is unclear what onomasiological elicitation tests might indicate; second, Gilquin's experiments were conducted on speakers of American English, and elicitation tests would need to be conducted on speakers of Singapore English, Hong Kong English, and British English in order to confirm a real correlation. Nonetheless, this finding is remarkable, as it suggests that it is at least possible to find a correlation between cognitive salience and corpus frequency, if the nature of corpus frequency is considered closely and measured in a particular way. This finding also suggests

that Geeraerts's hypothesis of onomasiological salience is in fact valid: onomasiological frequency (rather than semasiological frequency) can be seen to correlate with cognitive salience. In that case, speakers of all three varieties might be expected to exhibit similar cognitive salience patterns in laboratory settings for the concrete senses of these three verbs. That expectation could be tested and corroborated or refuted.

The present written data, however, does not correlate so neatly with Gilquin's (2008) findings on the high salience of concrete senses.
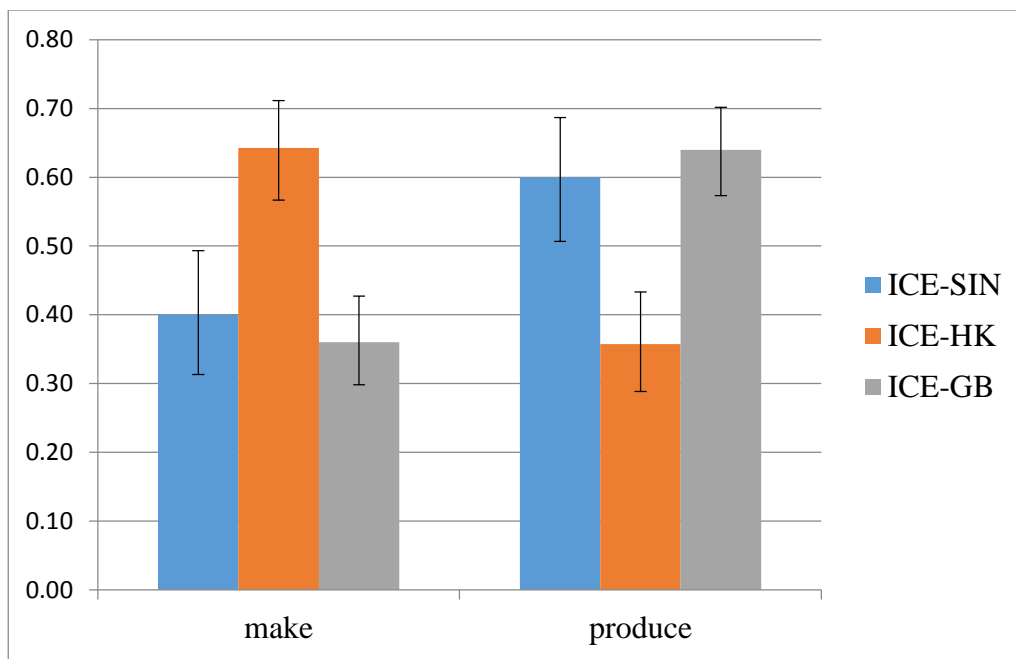
Figure 59: Probability of selecting concrete *make* and *produce* in the written portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

Figure 60: Probability of selecting concrete *take*, *collect*, *carry*, and *transport* in the written portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
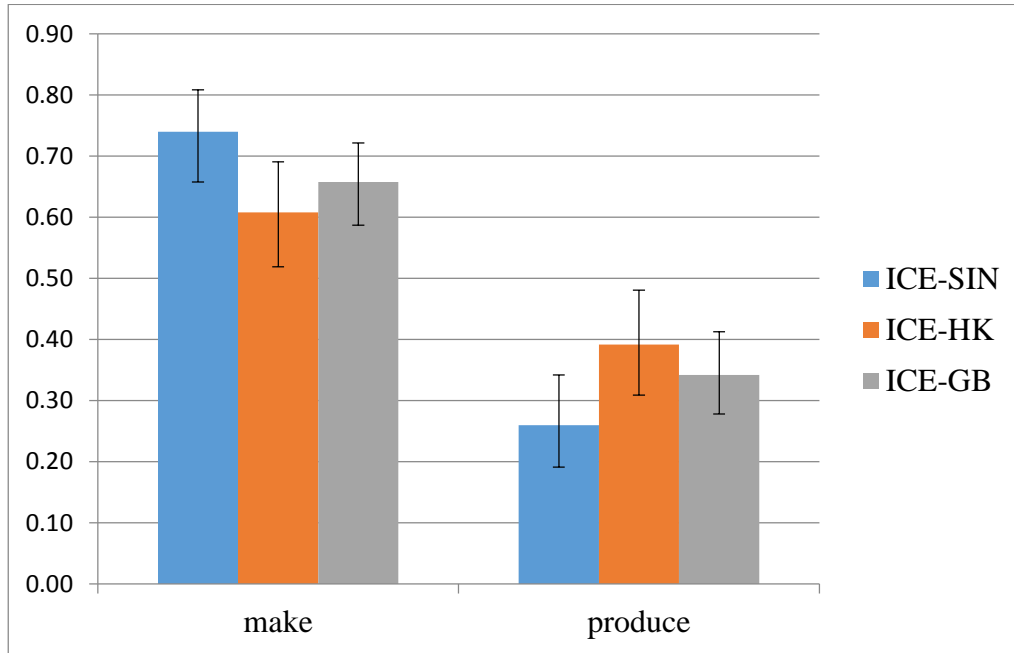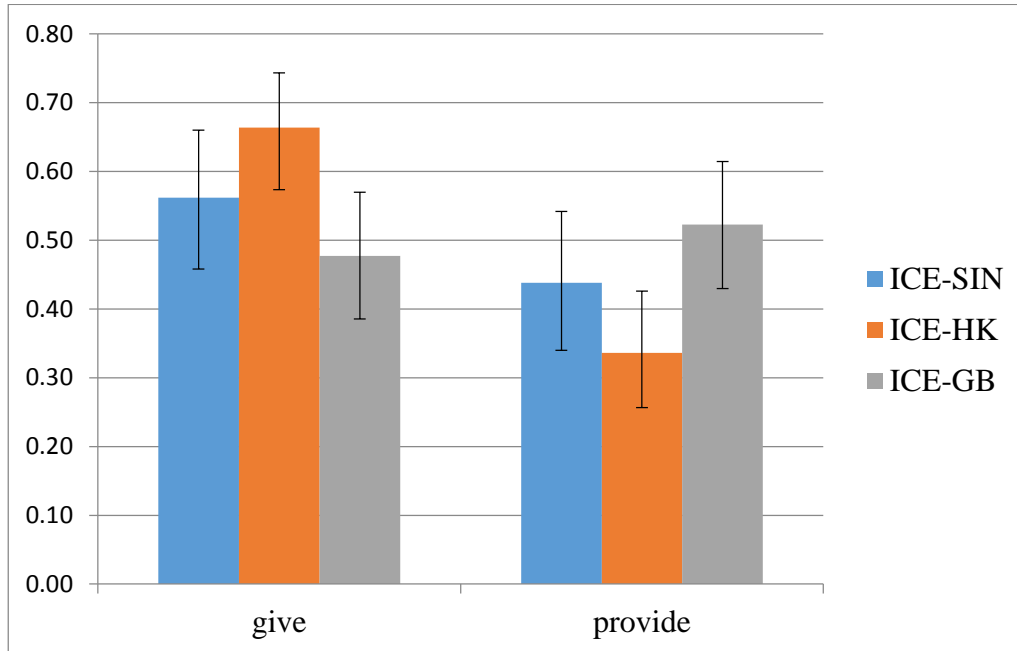
Figure 61: Probability of selecting concrete *give* and *provide* in the written portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

As Figure 59 through Figure 61 show, writers in each corpus do not tend to prefer the concrete senses of *make*, *take*, and *give* over their alternates. These findings therefore do not correlate with Gilquin's finding on the high cognitive salience of the concrete sense. Thus, data from written language do not seem to be able to usefully inform our understanding of cognitive salience as measured by Gilquin (2008). It may be that Geeraerts's hypothesis of onomasiological salience does not apply to written language. Written language can be seen as more mediated than spoken language, and differing stylistic standards influence psycholinguistic selection processes in speech and writing. Spoken language might therefore be expected to resemble cognitive salience patterns more strongly than written language, and Geeraerts's hypothesis of onomasiological salience could perhaps be modified to state that spoken onomasiological frequencies will correlate with cognitive salience.

Light senses tend to be the most common semasiologically, according to present findings (see Figure 50 through Figure 55, above), and to Gilquin (2008), Werner and Mukherjee (2012), and as suggested by the Collins COBUILD Dictionary. However, Gilquin (2008), via elicitation tests, has shown that light senses are extremely non-salient. In light of Geeraerts's hypothesis of onomasiological salience, we can ask whether onomasiological frequencies of light senses correlate with cognitive salience. Written data are not reprinted here, but can be referenced in chapters 6, 7, and 8. As with concrete uses, the data from writing is less consistent than the data from speech, and the data from writing does not apparently correlate with cognitive salience as measured by Gilquin (2008). In speech, Delexical *make*, *take*, and *give* tend to be less common than their onomasiological alternates.
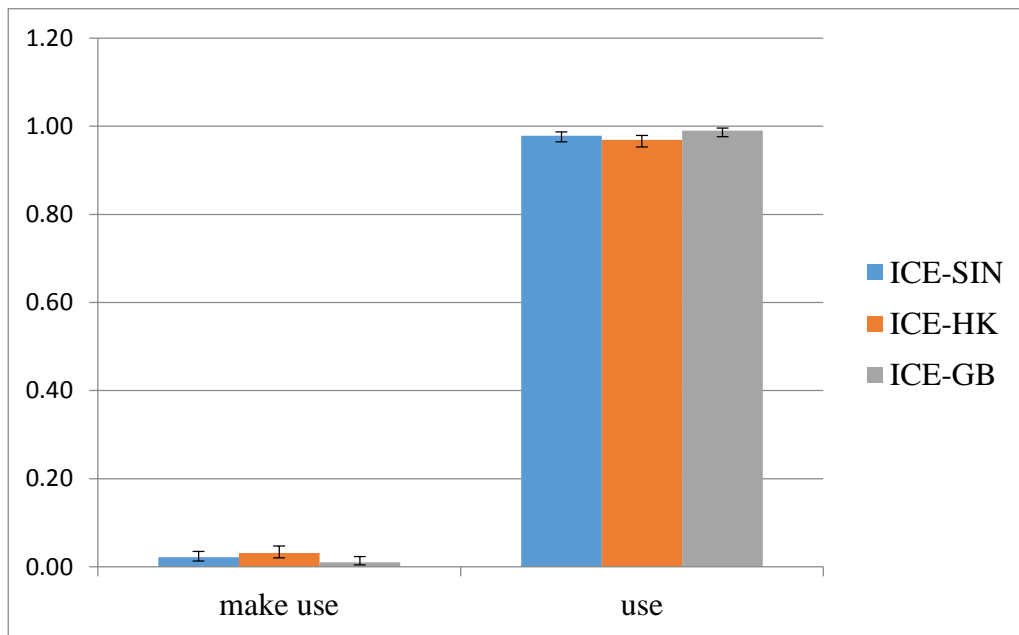
Figure 62: Probability of selecting *make use* and *use* (v.) in ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
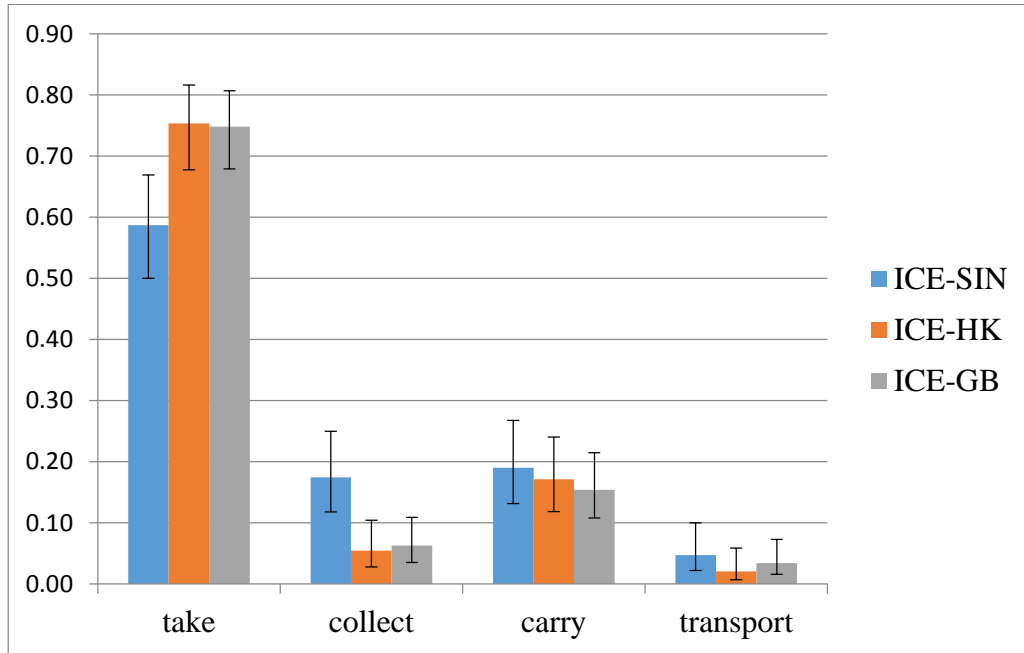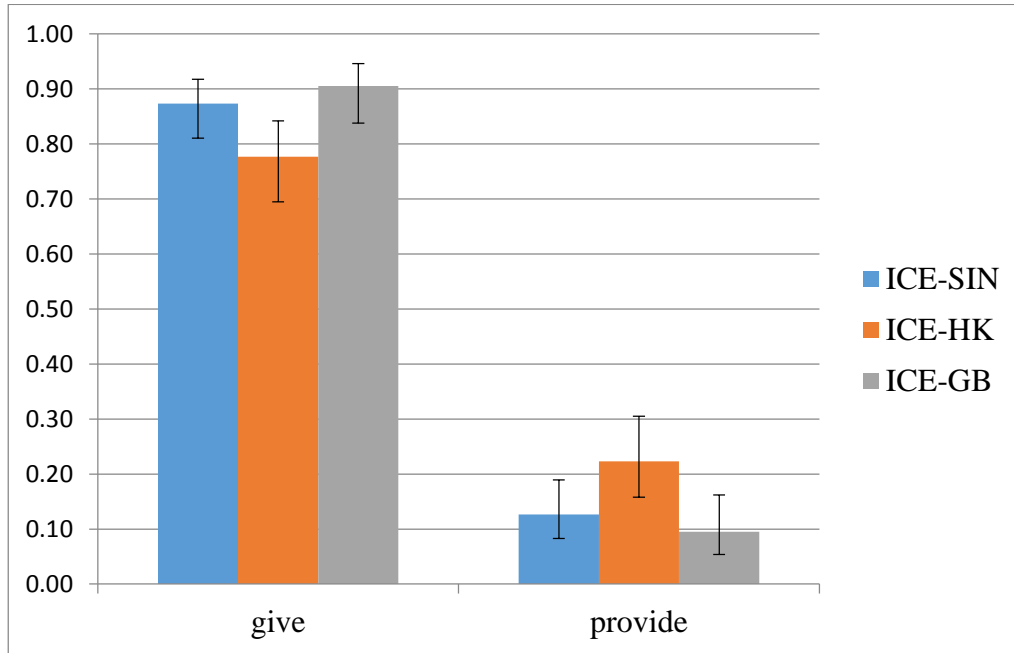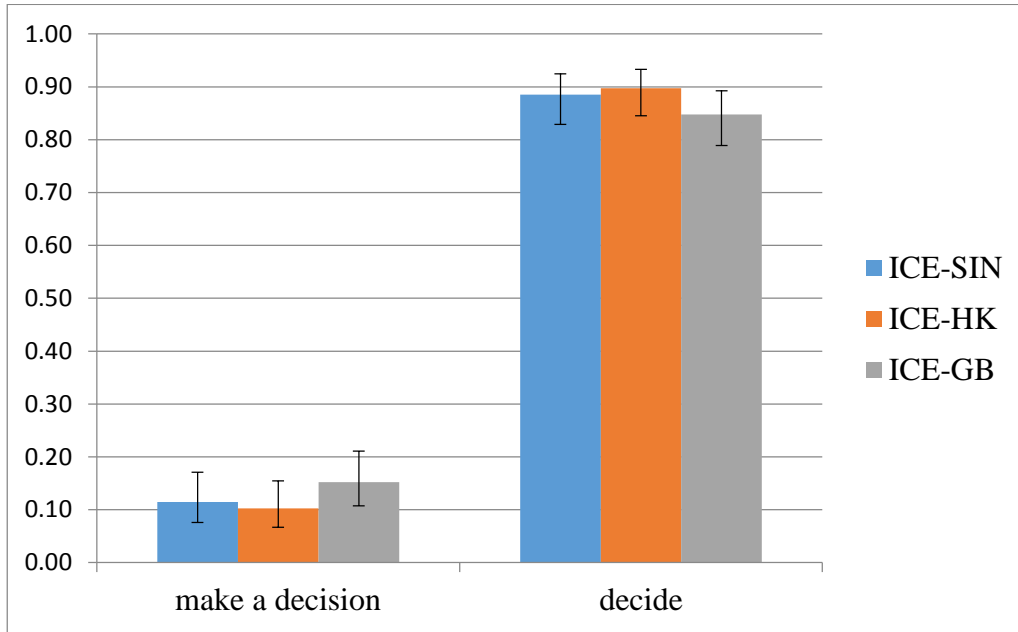
Figure 63: Probability of selecting *make a decision* and *decide* in ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
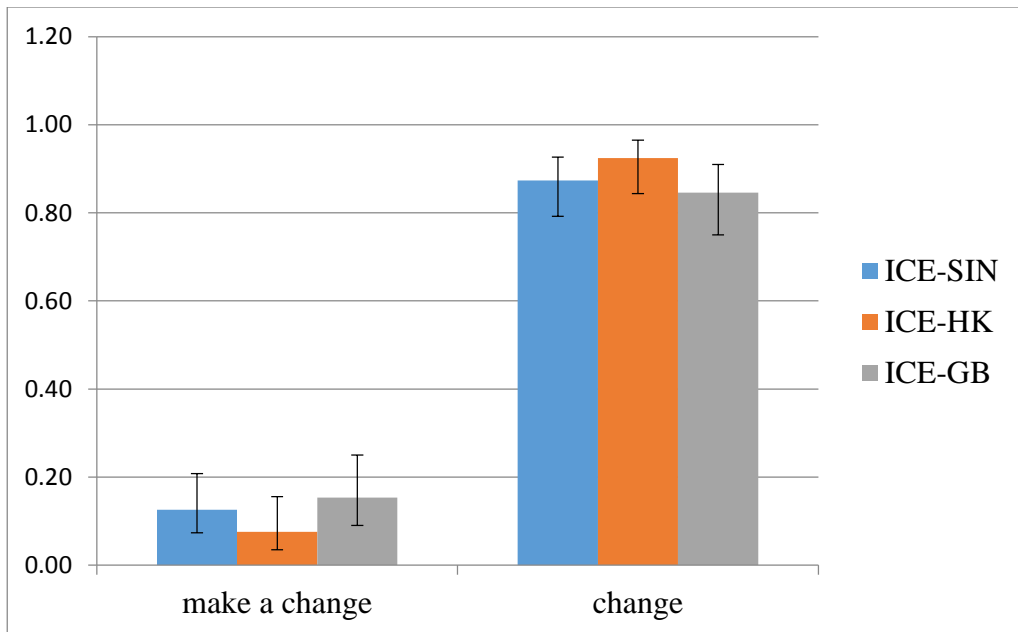
Figure 64: Probability of selecting *make a change* and *change* (v.) in ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
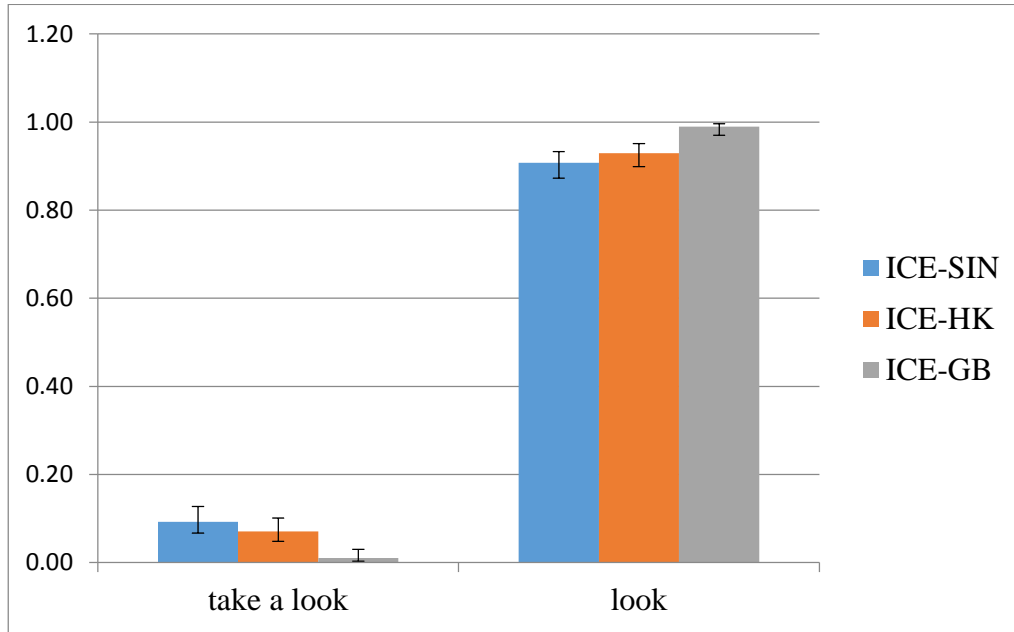
Figure 65: Probability of selecting *take a look* and *look* (v.) in ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
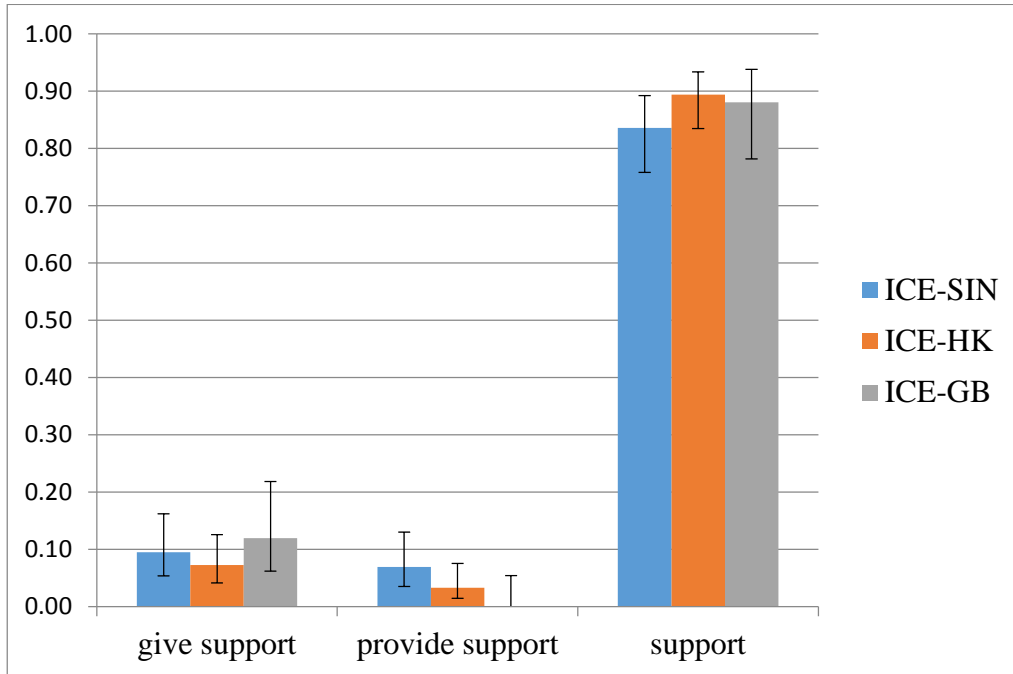
Figure 66: Probability of selecting *give support*, *provide support*, and *support* (v.) in ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

All of the above examples evince the low onomasiological frequency of light verbs and the high onomasiological frequency of the related verb. These examples seem to be very much in line with Gilquin's (2008) findings on the low cognitive salience of the light sense. There are, however, important exceptions to this onomasiological trend, as displayed in the following figures.
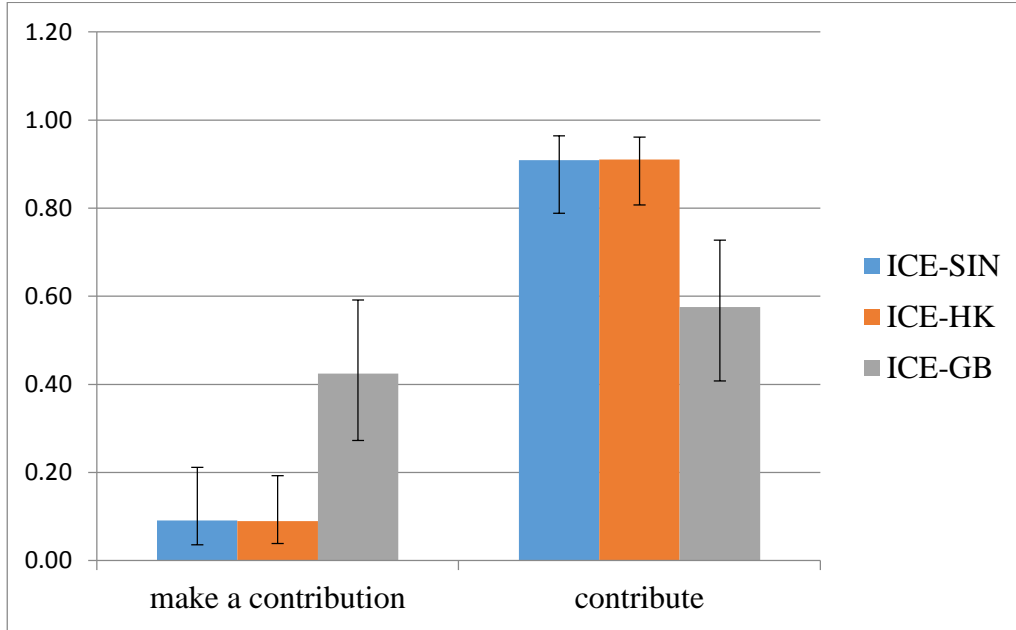
Figure 67: Probability of selecting *make a contribution* and *contribute* in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
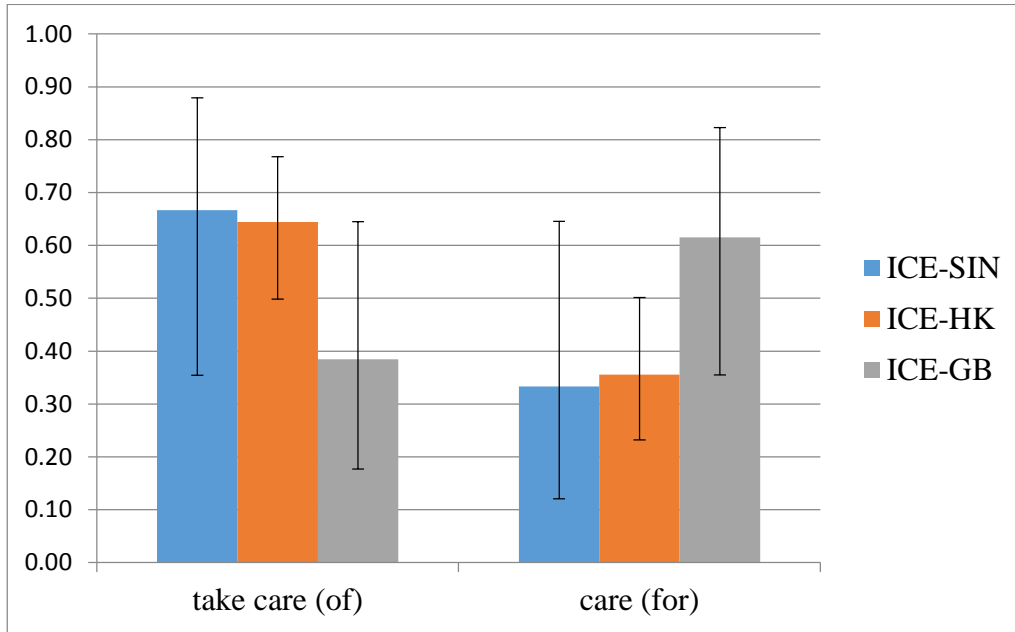
Figure 68: Probability of selecting *take care (of)* and *care (for)* in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
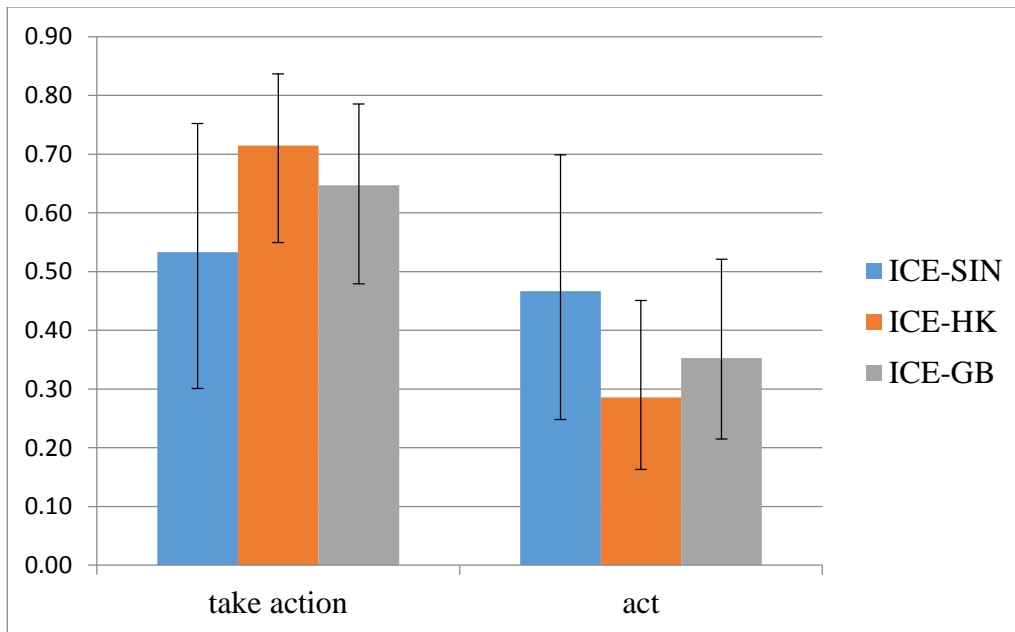
Figure 69: Probability of selecting *take action* and *act* in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.
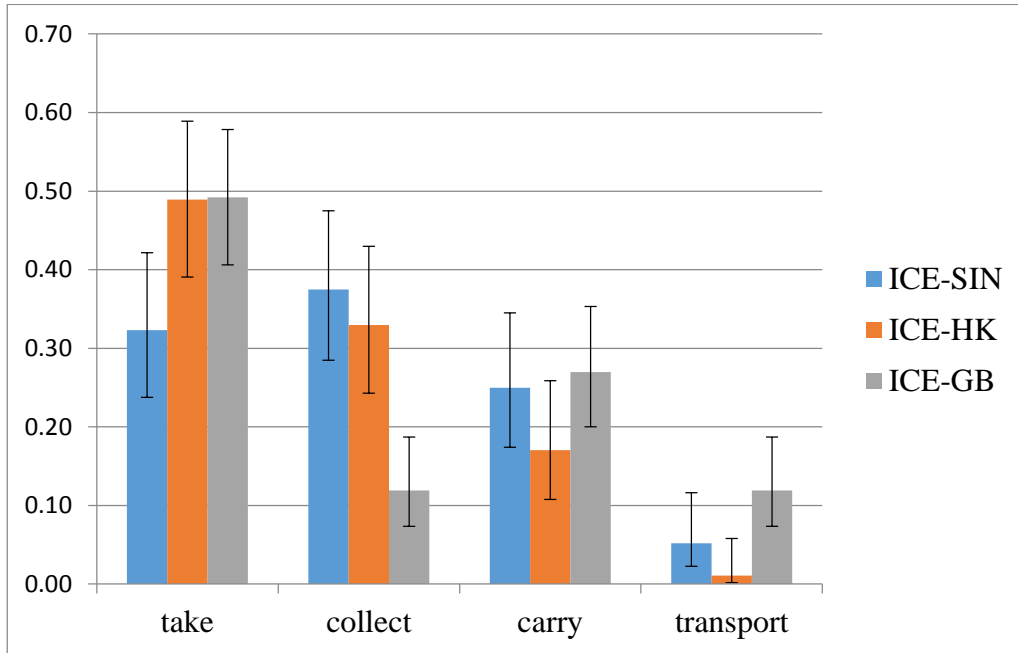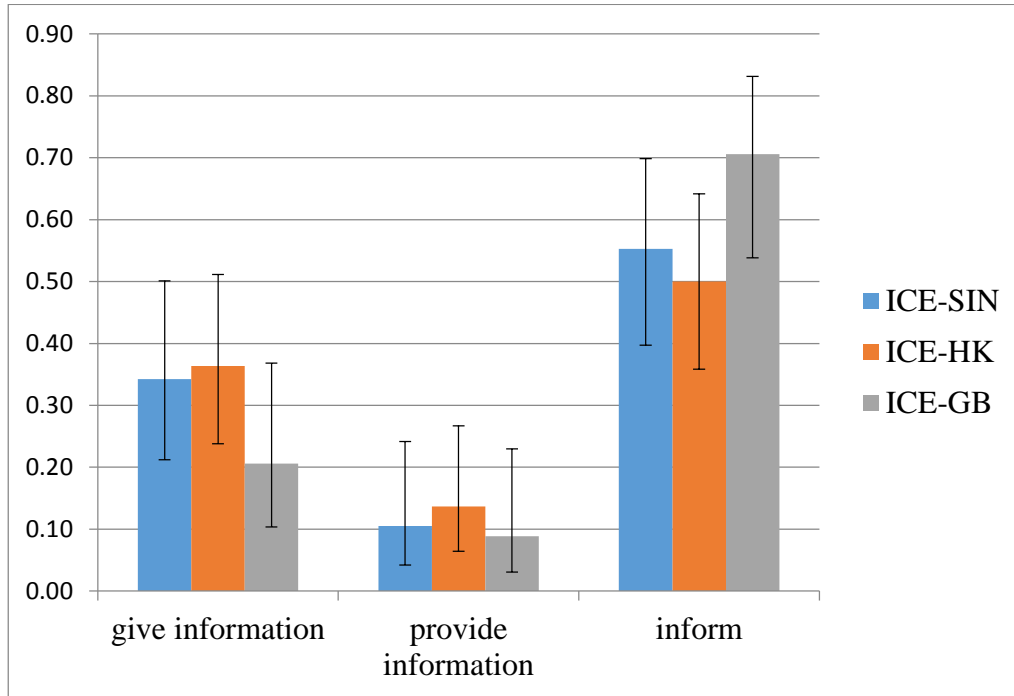
Figure 70: Probability of selecting *give information*, *provide information*, and *inform* in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB. The *y*-axis represents probabilities for each term in each corpus, and error bars represent Wilson intervals.

These figures might suggest that speakers in particular regions would exhibit unique cognitive salience patterns for the Delexical senses of these verbs. For example, speakers in Great Britain might tend to exhibit high cognitive salience for *make a contribution* in relation to *contribute*, while that might differ in Singapore and Hong Kong. Future laboratory tests of salience, perhaps both semasiological and onomasiological, in each region would be useful in further corroborating or refuting the onomasiological salience hypothesis.

In sum, semasiologically, although Gilquin (2008) found light senses of *give* and *take* to be more common than concrete senses in the Switchboard corpus, in the ICE corpora, light senses of *make*, *take*, and *give* are not consistently more common than concrete senses. Werner and Mukherjee (2012) found similar semasiological relations. Nonetheless, light senses do indeed tend to be more common than concrete senses. Onomasiologically, in speech,

concrete senses tend to be more frequent than their alternates while light senses tend to be less frequent than their alternates. Variation appears and exceptions occur for each lexical item, genre, and regional variety. The high cognitive salience observed by Gilquin (2008) generally tends to correlate with strong onomasiological preferences, and that correlation tends to be strongest in spoken language. Geeraerts's (2010) hypothesis might be revised, therefore, to propose specifically that relatively high onomasiological frequencies specifically in spoken language correlate with cognitive salience.

If onomasiological corpus frequencies do in fact correlate with cognitive salience more broadly, and if that correlation is corroborated in future investigations, then the ramifications for research in corpus linguistics and Cognitive Linguistics are immense. It would conceivably be possible to conduct experiments into cognitive salience without relying on laboratory settings and on-site psycholinguistic testing, and instead to rely on corpora. If the present conclusions are correct, then spoken corpora in particular would be appropriate. This potential shift in data collection methods would facilitate cognitive investigations into a wide array of linguistic features, focused on any region of the world for which a corpus was available, and conducted from anywhere in the world. In that case, compiling spoken corpora would be an extremely important task for cognitive researchers in the future.

9.4. In closing

The present study is an investigation of the semantics of *make*, *take*, and *give* in three varieties of World Englishes, but, equally, it is an experimental test for a particular set of methods. In the first section of this work, I have made the following methodological arguments:

a. Lexical semantic investigation must address language in use, and corpora are a powerful tool to that end. Corpora uniquely allow investigation of what language users do with language in the real world, and corpora can provide evidence of phenomena that occur far below

consciousness awareness. In addition, corpora can reliably corroborate or refute intuitions about language.

b. Corpus studies in general, and corpus studies of lexical semantics in particular, must include a careful consideration of the nature of *frequency*. Frequencies must be defined in terms of baselines, and an appropriate baseline must be determined and applied in relation to a particular research question. An onomasiological baseline has been presented as a strong methodological choice, especially for lexical semantic studies, because it accords with a logically and mathematically sound approach to frequency.

c. Evidence and examples derived from polysemy tests are a useful tool in corpus studies of lexical semantics. Antagonism evidence, autonomy evidence, identity evidence, and truth-condition evidence, in the form of corpus examples that resemble traditional polysemy tests, can inform our understanding of meaning in corpora.

d. Semantic research is essential to World Englishes scholarship, in order for basic knowledge and theoretical frameworks of World Englishes to be complete.

The present findings could not have been obtained without the above methodological arguments and approaches.

Summarized findings include the following:

a. Semantic variation in *make*, *take*, and *give* arises in onomasiological preferences, specifically in concrete senses and in written language. Semantics of these terms in spoken language is remarkably similar across the three corpora. Likewise, Delexical usage is remarkably similar between the regions in extremely consistent and nuanced ways. The present findings also suggest *degrees of Delexicality*, such that not all Delexical verbs have an equally discrete Delexical sense.

b. Semantic variation seems to reflect independent developments in written Hong Kong English, and sometimes in written Singapore English, rather than a dynamic relationship between those varieties and British English norms. This finding does not corroborate Schneider's (2007) dynamic model, but does seem to accord with Brutt-Griffler's (2002) model of language acquisition by speech communities.

c. Corpus frequencies do indeed correlate with cognitive salience if corpus frequencies are measured onomasiologically in speech, but not in writing. This deserves further investigation, including further experimental tests of cognitive salience and further corpus studies of onomasiological frequencies.

The present findings on usage preferences and semantic variation for *make*, *take*, and *give* and their alternates were not predicted, and could not have been uncovered without data in the form of large samples of language in use, like the ICE corpora. The present findings move far beyond the lexical exotica that are well documented for World Englishes, beyond the reflective, subjective observations on differences in meaning between regions, and even beyond predictable semantic variation. The ICE corpora have proven absolutely crucial in reaching such findings. Onomasiological analysis was performed on each sense of each word in the corpora, and issues in onomasiological analysis were addressed and resolved. For example, only alternate verbs that exhibited nearly universal alternation were selected for final analysis. A larger corpus would certainly allow a more complete analysis of lower-frequency alternates, but would likely require so many hours of manual semantic analysis as to be prohibitive. Indeed, the cost of the onomasiological approach is clear: in the present study, a total of 15,369 instances of *make*, *take*, and *give* were manually semantically analysed, one by one. Many thousands of instances of alternates of each sense were manually analysed as well, including instances of word forms such as *produce* and *provide* that were not ultimately semantic alternates at all. The cost of the approach is certainly the reason that many semantic researchers

opt against onomasiological analysis. However, the benefits are also clear. The entirely novel findings on usage preferences in relation to genre and abstract/concrete semantics could not have been uncovered in any other way.

Evidence for polysemy and distinct senses in the present study included methods derived from traditional polysemy tests. Examples of antagonism evidence, autonomy evidence, identity evidence, and truth-condition evidence (resembling the polysemy tests with the same names) were identified in each corpus and considered as evidence for or against polysemy. These polysemy tests were not applied in a simple way, for an absolute, binary answer, but were instead applied as heuristic tools for reflecting on, and analysing, language in use as evidenced by corpora. This proved to be a useful tool and an interesting reflective aid, and this sort of evidence deserves to be applied in future studies to continue to test its efficacy.

The present findings on lexical semantic variation do not clearly corroborate Schneider's (2007) dynamic model of World Englishes, and in fact seem to refute the predictions of that model. Instead, the present findings seem to align much more clearly with Brutt-Griffler's (2002) theory that World Englishes varieties emerge via processes of language acquisition, without a fully formed sense of an exonormative standard at all. Brutt-Griffler's (*ibid*.) model has not been a dominant model in the field, and it should perhaps be given closer attention. In addition, it is hoped that the present findings on *make*, *take*, and *give* could contribute just a few data points towards a data-driven model of semantic similarities and differences between varieties of World Englishes, not based on social, political, or geographical factors, but built up from semantic facts.

Finally, a relationship between corpus frequency and cognitive salience was identified, but, as I asserted in my methodological arguments at the outset, it is absolutely necessary that corpus linguists actively reflect on, and engage with, the nature of frequencies. It is inadequate for corpus linguists to accept raw numbers as meaningful frequencies. Moreover, while a *pmw* baseline is generally the simplest approach, and sometimes a useful approach for specific

research questions, an onomasiological baseline, which reflects selection processes underlying language production, is an extremely valuable tool, particularly for lexical semantic studies. An onomasiological baseline can uncover unique findings for lexical semantic studies, like the present findings on onomasiological variation for concrete senses of *make*, *take*, and *give*, and an onomasiological baseline, when measuring spoken language, correlates uniquely with cognitive salience for senses of polysemic words.

APPENDIX

	concrete	abstract	Delexical	Causal	Complex Transitive
ICE-SIN	41	52	266	58	196
ICE-HK	101	93	318	72	219
ICE-GB	72	70	321	56	214

Table 19: Instances of *make* in five hypothetical sense categories in the written portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 3 and Figure 50).

	concrete	abstract	Delexical	Causal	Complex Transitive
ICE-SIN	85	115	331	63	243
ICE-HK	63	104	367	69	240
ICE-GB	112	131	353	86	253

Table 20: Instances of *make* in five hypothetical sense categories in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 4 and Figure 51).

	make	bake	prepare	cook
ICE-SIN	7	0	1	2
ICE-HK	39	3	9	7
ICE-GB	8	0	0	0

Table 21: Instances of verbs with the sense 'Produce (Concrete, Food)' in the written portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 5 and Figure 7).

	make	bake	prepare	cook
ICE-SIN	25	8	5	21
ICE-HK	19	0	10	12
ICE-GB	27	1	0	17

Table 22: Instances of verbs with the sense 'Produce (Concrete, Food)' in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 6 and Figure 8).

	make	produce	create	manufacture	generate	build
ICE-SIN	8	3	2	1	0	0
ICE-HK	2	6	1	0	1	1
ICE-GB	5	6	2	1	1	0

Table 23: Instances of verbs with the sense 'Produce (Concrete, Non-Food)' and the Direct Object *product* in ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 9 and Figure 10).

	ICE-SIN	ICE-HK	ICE-GB
make	37	121	163
produce	66	106	199
create	28	56	36
prepare	0	17	0
manufacture	17	11	23
generate	16	24	12
build	74	105	43
emit	4	5	2
construct	12	17	15
develop	36	48	35
draw	17	25	25
yield	11	5	1
erect	5	6	2
compile	5	4	2
dig	3	1	1

Table 24: Instances of verbs with the sense 'Poduce (Concrete, Non-food)' in the written portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 11, Figure 13, Figure 15, and Figure 59).

	ICE-SIN	ICE-HK	ICE-GB
make	69	54	96
produce	33	46	64
create	22	14	10
prepare	0	10	0
manufacture	8	8	5
generate	6	12	3
build	40	60	17
emit	0	0	1
construct	4	4	2
develop	16	25	25
draw	13	24	15
yield	0	0	0
erect	1	0	2
compile	1	2	1
dig	1	0	1

Table 25: Instances of verbs with the sense 'Produce (Concrete, Non-food)' in the spoken portion of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 12, Figure 14, Figure 16, and Figure 56).

	make	put in	put	give	take	spend	expend	pay
ICE-SIN	15	6	3	2	1	2	1	3
ICE-HK	17	5	0	0	0	4	0	0
ICE-GB	18	2	0	0	0	0	1	0

Table 26: Instances of *make* and alternates for *make* in the construction *make an effort*, in ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 17).

	make use	use
ICE-SIN	51	1250
ICE-HK	36	1259
ICE-GB	11	1185

Table 27: Instances of *make use* and *make* in ICE-SIN, ICE-HK, and ICE-GB
(to accompany Figure 18 and Figure 62).

	make a decision	decide
ICE-SIN	34	253
ICE-HK	103	273
ICE-GB	59	262

Table 28: Instances of *make a decision* and *decide* in ICE-SIN, ICE-HK, and
ICE-GB (to accompany Figure 19 and Figure 63).

	make a change	change
ICE-SIN	18	107
ICE-HK	13	103
ICE-GB	20	91

Table 29: Instances of *make a change* and *change* (v.) in ICE-SIN, ICE-HK,
and ICE-GB (to accompany Figure 20 and Figure 64).

	make contact	contact
ICE-SIN	2	46
ICE-HK	7	113
ICE-GB	10	47

Table 30: Instances of *make contact* and *contact* (v.) in ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 21).

	make a contribution	contribute
ICE-SIN	13	88
ICE-HK	17	88
ICE-GB	19	48

Table 31: Instances of *make a contribution* and *contribute* in ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 22 and Figure 67).

	concrete	abstract	adopt	consider	Delexical
ICE-SIN	31	17	65	42	79
ICE-HK	46	25	82	59	97
ICE-GB	62	31	84	45	85

Table 32: Instances of *take* in five hypothetical sense categories in the written portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 23 and Figure 52).

	concrete	abstract	adopt	consider	Delexical
ICE-SIN	74	38	80	85	161
ICE-HK	110	34	145	89	144
ICE-GB	131	62	116	99	79

Table 33: Instances of *take* in five hypothetical sense categories in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 24 and Figure 53).

	ICE-SIN	ICE-HK	ICE-GB
take	31	46	62
collect	36	31	15
carry	24	16	34
push	5	5	4
lift	3	0	3
transport	5	1	15
seize	6	0	1
snatch	3	1	2
pull	6	2	3
heave	1	0	1
grab	3	4	1
haul	2	1	4
extract	8	0	4
transfer	3	3	4
withdraw	2	0	3
confiscate	2	2	2
drag	7	4	2

Table 34: Instances of verbs that alternate with *take* with the sense 'Transfer (Concrete)' in the written portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 25, Figure 27, and Figure 60).

	ICE-SIN	ICE-HK	ICE-GB
take	74	110	131
collect	22	8	11
carry	24	25	27
push	6	3	13
lift	0	2	4
transport	6	3	6
seize	2	18	6
snatch	1	1	0
pull	4	1	10
heave	0	0	0
grab	3	0	4
haul	0	1	1
extract	1	0	3
transfer	5	3	4
withdraw	3	3	4
confiscate	0	2	1
drag	6	2	4

Table 35: Instances of verbs that alternate with *take* with the sense 'Transfer (Concrete)' in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 26, Figure 28, and Figure 57).

	take	adopt	assume
ICE-SIN	55	49	22
ICE-HK	78	61	6
ICE-GB	77	29	3

Table 36: Instances of verbs with the sense 'Adopt/Assume' in the written portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 29).

	take	adopt
ICE-SIN	74	31
ICE-HK	141	29
ICE-GB	112	19

Table 37: Instances of verbs with the sense 'Adopt/Assume' in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 30).

	take a decision	decide
ICE-SIN	4	252
ICE-HK	8	273
ICE-GB	21	262

Table 38: Instances of *take a decision* and *decide* in ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 31).

	take a look	look
ICE-SIN	41	388
ICE-HK	35	416
ICE-GB	6	288

Table 39: Instances of *take a look* and *look* (v.) in ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 32 and Figure 65).

	take care (of)	care (for)
ICE-SIN	14	9
ICE-HK	21	19
ICE-GB	2	5

Table 40: Instances of *take care (of)* and *care (for)* in the written portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 33).

	take care (of)	care (for)
ICE-SIN	6	3
ICE-HK	29	16
ICE-GB	5	8

Table 41: Instances of *take care (of)* and *care (for)* in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 34 and Figure 68).

	take action	act
ICE-SIN	16	17
ICE-HK	20	20
ICE-GB	12	9

Table 42: Instances of *take action* and *act* (v.) in the written portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 35).

	take action	act
ICE-SIN	8	7
ICE-HK	25	10
ICE-GB	22	12

Table 43: Instances of *take action* and *act* (v.) in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 36 and Figure 69).

	concrete	abstract	Delexical
ICE-SIN	50	145	151
ICE-HK	77	161	199
ICE-GB	52	160	167

Table 44: Instances of *give* in three hypothetical sense categories in the written portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 37 and Figure 54).

	concrete	abstract	Delexical
ICE-SIN	131	400	211
ICE-HK	94	270	303
ICE-GB	105	257	227

Table 45: Instances of *give* in three hypothetical sense categories in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 38 and Figure 55).

	ICE-SIN	ICE-HK	ICE-GB
give	50	77	52
provide	39	39	57
issue	15	92	20
submit	49	35	8
supply	7	5	26
hand	10	18	11
pass	9	8	5
donate	1	4	2
transfer	5	8	3
present	4	6	5
contribute	0	0	1
grant	0	10	0
despatch	0	1	2

Table 46: Instances of verbs that alternate with *give* with the sense 'Transfer (Concrete)' in the written portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 39, Figure 42, and Figure 61).

	ICE-SIN	ICE-HK	ICE-GB
give	131	94	105
provide	19	27	11
issue	21	17	11
submit	21	6	0
supply	22	7	13
hand	17	27	13
pass	14	16	12
donate	1	1	0
transfer	3	0	3
present	12	21	11
contribute	2	0	0
grant	0	3	1
despatch	0	0	0

Table 47: Instances of verbs that alternate with *give* with the sense 'Transfer (Concrete)' in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 43, and Figure 58).

	give	provide
ICE-SIN	145	74
ICE-HK	161	109
ICE-GB	160	100

Table 48: Instances of verbs that alternate with *give* with the sense 'Transfer (Abstract)' in the written portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 44).

	give	provide
ICE-SIN	400	41
ICE-HK	270	75
ICE-GB	257	31

Table 49: Instances of verbs that alternate with *give* with the sense 'Transfer (Abstract)' in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 45).

	give support	provide support	support
ICE-SIN	6	4	79
ICE-HK	4	9	98
ICE-GB	4	2	74

Table 50: Instances of *give support*, *provide support*, and *support* (v.) in the written portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 46 and Figure 66).

	give support	provide support	support
ICE-SIN	11	8	97
ICE-HK	11	5	135
ICE-GB	8	0	59

Table 51: Instances of *give support*, *provide support*, and *support* (v.) in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 47).

	give information	provide information	inform
ICE-SIN	4	9	74
ICE-HK	15	20	36
ICE-GB	9	20	20

Table 52: Instances of *give information*, *provide information*, and *inform* in the written portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 48).

	give information	provide information	inform
ICE-SIN	13	4	21
ICE-HK	16	6	22
ICE-GB	7	3	24

Table 53: Instances of *give information*, *provide information*, and *inform* in the spoken portions of ICE-SIN, ICE-HK, and ICE-GB (to accompany Figure 49 and Figure 70).

BIBLIOGRAPHY

Primary Sources

Davies, Mark. 2004. *BYU-BNC*. (Based on the British National Corpus from
        Oxford University Press). Available online at http://corpus.byu.edu/bnc/

Nelson, Gerald, Sean Wallis and Bas Aarts. 2002. *Exploring natural language:*
        Working with the British component of the International Corpus of
        English. Amsterdam: John Benjamins.

International Corpus of English. 2014. www.ice-corpora.net/ice

References

Aarts, Bas. 2011. *Oxford Modern English grammar*. Oxford: Oxford University
        Press.

Aarts, Bas, Joanne Close and Sean Wallis. 2014. Choices over time:
        Methodological issues in investigating current change. In Bas Aarts,
        Joanne Close, Geoffrey Leech and Sean Wallis (eds.), *The verb phrase*
        in English: Investigating recent language change with corpora.
        Cambridge: Cambridge University Press. 14-45.

Aarts, Jan. 1991. Intuition-based and observation-based grammars. In Karin
        Aijmer and Bengt Altenberg (eds.), *English corpus linguistics: Studies*
        in honour of Jan Svartvik. Harlow, England: Longman. 44-63.

Adejare, O. and A. Afolayan. 1982. Semiotics: The unexplored level of
        variation in second language use. Paper presented at the 17th regional
        seminar, SEAMEO Regional Language Centre, Singapore.

Algeo, John. 1995. Having a look at the expanded predicate. In Bas Aarts and
        Charles Meyer (eds.), *The verb in contemporary English: Theory and*
        description. Cambridge: Cambridge University Press.

Arppe, Antti, Gaetanelle Gilquin, Dylan Glynn, Martin Hilpert and Arne
        Zeschel. 2010. Cognitive corpus linguistics: Five points of debate on
        current theory and methodology. *Corpora* 5 (1), 1-27.

Atkins, Sue and Michael Rundell. 2008. *The Oxford guide to practical lexicography*. Oxford: Oxford University Press.

Bacon-Shone, John and Kingsley Bolton. 1998. Charting Multilingualism: Language censuses and language surveys in Hong Kong. In Martha C. Pennington (ed.), *Language in Hong Kong at Century's End*. Hong Kong: Hong Kong University Press. 43-90.

Baker, Alan. 2010. Simplicity. In Edward M. Zalta (ed.), *Stanford Encyclopedia of Philosophy*. Palo Alto: Stanford University.

Balasubramanian, Chandrika. 2009. Circumstance adverbials in registers of Indian English. *World Englishes* 28 (4), 485-508.

Bao, Zhiming and Lionel Wee. 1999. The passive in Singapore English. *World Englishes* 18 (1), 1-11.

Begum, Rizwana and Thiru Kandiah. 1997. Misrecognitions of variability in new varieties of English: Tamil minority usage in Singapore English. In Edgar W. Schneider (ed.), *Englishes around the world*, vol.2. Amsterdam: Benjamins. 189-204.

Bell, Alan. 2006. Speech accommodation theory and audience design. In Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd edn. Amsterdam: Elsevier. 648-51.

Bolt, Philip and Kingsley Bolton. 1996. The International Corpus of English in Hong Kong. In Sidney Greenbaum (ed.), *Comparing English worldwide: The International Corpus of English*. Oxford: Clarendon Press. 197-214.

Bolton, Kingsley. 2006. World Englishes Today. In Braj B. Kachru, Yamuna Kachru and Cecil L. Nelson (eds.) *The Handbook of World Englishes*. Malden, MA: Blackwell. 240-69.

Brown, Adam. 1999. *Singapore English in a nutshell: An alphabetical description of its features*. Singapore: Federal Publications.

Brugman, Claudia and George Lakoff. 1988. Cognitive typology and lexical networks. University of California, Berkeley.

http://georgelakoff.files.wordpress.com/2011/04/cognitive-topology-and-lexical-networks-lakoff-and-brugman-1988.pdf

Brutt-Griffler, Janina. 2002. *World English: A study of its development*. Clevedon: Multilingual Matters Ltd.

Butterfield, J. and R. Krishnamurthy. 2000. Beyond the dictionary: On-line learning in the classroom. *TESOL Spain Newsletter* 23, 3-5.

Chisanga, T. and N. M. Kamwangam Alu. 1997. Owning the other tongue: The English language in southern Africa. *Journal of Multilingual and Multicultural Development* 18 (2), 89-99.

Chomsky, Noam. 1962. Paper given at the University of Texas, 1958, 3rd Texas Conference on Problems of Linguistic Analysis in English. Austin: University of Texas.

Collins, Peter. 2005. The modals and quasi-modals of obligation and necessity in Australian English and other Englishes. *English World-Wide* 26 (3), 249-73.

Collins, Peter. 2009. Modals and quasi-modals in World Englishes. *World Englishes* 28 (3), 281-92.

Cruse, D. Alan. 1982. On lexical ambiguity. *Nottingham Linguistics Circular* 11 (2): 65-80.

Cruse, D. A. 1986. *Lexical semantics*. Cambridge: CUP.

Cruse, Alan. 2004. *Meaning in language: An introduction to semantics and pragmatics*. Oxford: OUP.

Cummings, Patrick J. and Hans-Georg Wolf. 2011. *A dictionary of Hong Kong English: Words from the fragrant harbour*. Hong Kong: Hong Kong University Press.

Derrida, Jacques. 1977. *Signature, event, context*. Trans. Samuel Weber and Jeffrey Mehlman. *Glyph* 1, 172-97.

de Klerk, Vivian. 2005. Expressing levels of intensity in Xhosa English. *English World-Wide* 26 (1), 77-95.

Deterding, David. 2007. *Singapore English*. Edinburgh: Edinburgh University Press.

Durkin, Philip. 2014. *Borrowed words*. Oxford: Oxford University Press.

Enfield, N. J. 2002. *Linguistic epidemiology: Semantics and grammar of language contact in mainland Southeast Asia*. London: Routledge Curzon.

Evison, Jane. 2010. What are the basics of analysing a corpus? In Anne O'Keefe and Michael McCarthy (eds.), *The Routledge handbook of corpus linguistics*. London: Routledge. 122-35.

Facchinetti, Roberta. 2007. *Theoretical description and practical applications of linguistic corpora*. No location: QuiEdit.

Firth, J. (1957). *Papers in Linguistics, 1934-1951*. Oxford: Oxford University Press.

Francis, W. Nelson and Henry Kučera. 1982. *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

Fuchs, Robert. 2012. Focus marking and semantic transfer in Indian English. *English World-Wide* 33 (1), 27-52.

Fuchs, Robert, Ulrike Gut and Taiwo Soneye. 2013, "We just don't even know": The usage of the pragmatic focus particles *even* and *still* in Nigerian English. *English World-Wide* 34 (2), 123-45.

Geeraerts, Dirk. 1988. Where does prototypicality come from? In Brygida Rudzka-Ostyn (ed.), *Topics in Cognitive Linguistics*. Amsterdam: John Benjamins. 207-29.

Geeraerts, Dirk. 1997. *Diachronic prototype semantics: A contribution to historical lexicology*. Oxford: Clarendon Press.

Geeraerts, Dirk. 2002. The scope of diachronic onomasiology. In Vilmos Agel, Andreas Gardt, Ulrike Hass-Zumkehr & Thorsten Roelcke (ed.), *Das Wort. Seine strukturelle und kulturelle Dimension. Festschrift für Oskar Reichmann zum 65*. Tübingen: Niemeyer. 29-44.

Geeraerts, Dirk. 2006. Onomasiology and lexical variation. In Keith Allan (ed.), *Concise encyclopedia of semantics*. Amsterdam: Elsevier.

Geeraerts, Dirk. 2006 [1989]. Prospects and problems of prototype theory. In Dirk Geeraerts, *Words and other wonders*. Berlin: Mouton de Gruyter. 3-26.

Geeraerts, Dirk. 2006 [1993]. Vagueness's puzzles, polysemy's vagaries. In Dirk Geeraerts, *Words and other wonders*. Berlin: Mouton de Gruyter. 99-148.

Geeraerts, Dirk. 2010. *Theories of lexical semantics*. Oxford: Oxford University Press.

Geeraerts, Dirk, Stefan Grondelaers and Peter Bakema. 1994. *The structure of lexical variation: Meaning, naming, and context*. Berlin: Mouton de Gruyter.

Gibbs, R. W. Jr. and T. Meetlock. 2001. Psycholinguistics and mental representations. *Cognitive Linguistics* 10 (3): 263-9.

Gilquin, Gaetanelle. 2006. The place of prototypicality in corpus linguistics: Causation in the hot seat. In Stefan Gries and Anatol Stefanowitsch (eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis*. Berlin: Mouton de Gruyter. 159-191.

Gilquin, Gaetanelle. 2008. What you think ain't what you get: Highly polysemous verbs in mind and language. In Jean-Remi Lapaire, Guillaume Desagulier and Jean-Baptiste Guignard (eds.), *From gram to mind: Grammar as cognition*. Bordeaux: Presse Universitaires de Bordeaux. 235-255.

Görlach, Manfred. 1995 [1990]. Dictionaries of transplanted Englishes. In Manfred Görlach (ed), *More Englishes: New studies in varieties of English, 1988-1994*. Amsterdam: John Benjamins.124-63.

Görlach, Manfred. 1998. *Even more Englishes: Studies 1996-1997*. Amsterdam: John Benjamins.

Gramley, Stephan. 2001. *The vocabulary of World English*. London: Arnold.

Grant, Lynne E. 2012. Culturally motivated lexis in New Zealand English. *World Englishes* 31, 162-76.

Greenbaum, Sidney. 1991. The development of the International Corpus of English. In Karin Aijmer and Bengt Altenberg (eds.), *English corpus linguistics: Studies in honour of Jan Svartvik*. Harlow, England: Longman. 83-94.

Greenbaum, Sidney. 1996. Introducing ICE. In Sidney Greenbaum (ed.), *Comparing English worldwide: The International Corpus of English*. 3-12.

Gries, Stefan Th. 2006a. Introduction. In Stefan Gries and Anatol Stefanowitsch (eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis.* Berlin: Mouton de Gruyter. 1-17.

Gries, Stefan Th. 2006b. Corpus-based methods and cognitive semantics: The many senses of *to run*. In Stefan Gries and Anatol Stefanowitsch (eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis*. Berlin: Mouton de Gruyter. 57-99.

Gries, Stefan Th. 2012. Behavioral profiles: A fine-grained and quantitative approach in corpus-based lexical semantics. In Gary Libben, Gonia Jarema and Chris Westbury (eds.), *Methodological and analytic frontiers in lexical research*. Amsterdam: John Benjamins Publishing Company. 57-80.

Haase, Christoph. 2004. Conceptual specifics in East African English: Quantitative arguments from the ICE-East Africa corpus. *World Englishes* 23 (2): 261-8.

Hanks, Patrick. 2000. Do word meanings exist? *Computers and the humanities* 34: 205-11.

Hanks, Patrick. 2009. The impact of corpora on dictionaries. In Paul Baker (ed.), *Contemporary corpus linguistics*. London, Continuum. 214-36.

Hanks, Patrick. 2013. *Lexical analysis: Norms and exploitations*. Boston: MIT Press.

Hashim, Azirah and Gerhard Leitner. 2011. Contact expressions in contemporary Malaysian English. *World Englishes* 30 (4): 551-68.

Hempl, George. 1902. Stovepipes and funnels. *Dialect Notes* 2, 250-56.

Heylen, Kris, Jose Tummers and Dirk Geeraerts. 2008. Methodological issues in corpus-based Cognitive Linguistics. In Kristiansen and Dirven (ed.) *Cognitive Sociolinguistics: Language variation, cultural models, social systems*. Berlin: Mouton de Gruyter. 91-128.

Hilpert, Martin. 2008. The English comparative – language structure and language use. *English Language and Linguistics* 12 (3), 395-418.

Hiramoto, Mie and Yosuke Sato. 2012. *Got* interrogatives and answers in Colloquial Singapore English. *World Englishes* 31, 198-207.

Historical Thesaurus of the Oxford English Dictionary.

Hofland, Knut and Stig Johansson. 1982. *Word frequencies in British and American English*. Bergen: The Norwegian Computing Centre for the Humanities.

Huddleston, Rodney. 1984. *Introduction to the grammar of English*. Cambridge: Cambridge University Press.

Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Hundt, Marianne. 2006. The committee has/have decided… On concord patterns with collective nouns in inner and outer circle varieties of English. *Journal of English Linguistics*, 34 (3). 206-32.

Hundt, Marianne. 2007. *English mediopassive constructions: A cognitive, corpus-based study of their origin, spread, and current status*. Amsterdam: Rodopi.

Hundt, Marianne. 2009. How often to things *get V-ed* in Philippine and Singapore English? A case study of the *get* -passive in two outer-circle varieties of English. In Rhonwen Bowen, Mats Mobarg & Solve Ohlander (eds.), *Corpora and discourse - and stuff: Papers in honor of Karin Aijmer*. Gothenburg Studies in English 96, 121-131. Gothenburg: Gothenburg University Press.

Hymes, D. H. 1996. *Ethnography, linguistics, narrative inequality: Toward an understanding of voice*. Bristol, PA: Taylor and Francis.

The ICE Project. 2009. *International Corpus of English*. http://icecorpora.net. Accessed September 2011.

Imm, Tan Siew. 2009. Lexical borrowing from Chinese languages in Malaysian English. *World Englishes* 28 (4), 451-84.

Jenkins, Richard. 1996. Social Identity. London: Routledge.

Jespersen, Otto. 1954. *A Modern English grammar on historical principles, Part VI: Morphology*. London: Bradford and Dickens.

Kachru, Braj. 1965. The Indianness in Indian English, *Word* 21, 391-410.

Kachru, Braj. 1969. English in South Asia. In Thomas A. Sebeok (ed.), *Current trends in linguistics*. The Hague: Mouton. 627-87.

Kachru, Braj. 1975. Lexical innovations in South Asian English. *International Journal of the Sociology of Language* 4: 55-94.

Kachru, Braj. 1983. The *Indianness of English: The English language in India*. New Delhi: Oxford University Press.

Kachru, Braj B. 1985. Standards, codification and sociolinguistic realism: the English language in the outer circle. In Randolph Quirk and H. G. Widdowson (eds.), *English in the world: Teaching and learning the language and literatures*. Cambridge: Cambridge University Press. 11-31.

Kachru, Braj. 1991. Liberation linguistics and the Quirk concern. *English Today* 7, 3-13.

Kachru, Braj B. 1992. *The other tongue: English across cultures*. 2nd edn. Urbana & Chicago: University of Illinois Press.

Kay, Christian J. and Thomas J. Chase. 1990. Semantic approaches to an historical thesaurus. In Jerzy Tomaszczyk and Barbara Lewandowska-Tomaszczyk (eds.), *Meaning and lexicography*. Amsterdam: John Benjamins. 303-12.

Kemp, Gary. 2012. *Quine versus Davidson: Truth, reference and meaning*. Oxford: Oxford University Press.

Kempson, Ruth M. 1977. *Semantic theory*. Cambridge: Cambridge University Press.

Kilgarriff, Adam and Iztok Kosem. 2012. Corpus tools for lexicographers. In Sylviane Granger and Magali Paquot (eds.). *Electronic lexicography*. Oxford: Oxford University Press. 31-55.

Kim, Chonghyuk and Yosuke Sato. 2013. More on *kena*-passives in Singapore English. *World Englishes* 32, 297-307.

Kim, Chonghyuk and Lionel Wee. 2009. Resolving the paradox of Singapore English *hor*. *English World-wide* 30 (3): 241-61.

Kouega, Jean-Paul. 2006. *Aspects of Cameroon English usage: A lexical appraisal*. Munich: Lincom Europe.

Kouega, Jean-Paul. 2007. *A Dictionary of Cameroon English Usage*. Frankfurt: Lang.

Krishnamurthy, R. 2000. Collocation: From *silly ass* to lexical sets. In C. Heffer, H. Sauntson and G. Fox (eds.), *Words in Context: A tribute to John Sinclair on his retirement*. Birmingham: University of Birmingham.

Kurath, Hans, Marcus L. Hansen, Bernard Bloch and Julia Bloch. 1939. *Handbook of the linguistic geography of New England*. Providence, RI: Brown University.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Lakoff, George. 1970. A note on vagueness and ambiguity. *Linguistic Inquiry* 1 (3), 357-9.

Lambert, James. 2012. Beyond Hobson-Jobson: Towards a New Lexicography for Indian English. *English Worldwide* 33, 292–320.

Lange, Claudia. 2007. Focus marking in Indian English. *English Worldwide* 28 (1): 89-118.

Lee, Jackie F. K. and Peter Collins. 2004. On the usage of have, dare, need, ought and used to in Australian English and Hong Kong English. *World Englishes* 23 (4), 501-13.

Lee, Nala Huiying, Ling Ai Ping and Hiroko Nomoto. 2009. Colloquial Singapore English *got*: Functions and substratal influences. *World Englishes*, 28.3. 293-318.

Lee, Sarah & Debra Ziegeler. 2006. Analysing a semantic corpus study across English dialects: Searching for paradigmatic parallels. In Andrew Wilson, Dawn Archer & Paul Rayson (eds.), *Corpus linguistics around the world*. Amsterdam: Rodopi. 121-39.

Leimgruber, Jakob R. E. 2013. *Singapore English: Structure, variation, and usage*. Cambridge: Cambridge University Press.

Levshina, Natalia, Dirk Geeraerts, and Dirk Speelman. 2014. Dutch causative constructions: Quantification of meaning and meaning of quantification. In Dylan Glynn and Justyna A. Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*. Amsterdam: John Benjamins Publishing Company. 205-22.

Libben, Gary, Chris Westbury and Gonia Jarema. 2012. The challenge of embracing complexity. In Gary Libben, Gonia Jarema and Chris Westbury (eds.), *Methodological and analytic frontiers in lexical research*. Amsterdam: John Benjamins Publishing Company. 1-11.

Lindquist, Hans. 2009. *Corpus linguistics and the description of English*. Edinburgh: Edinburgh University Press.

Ling, Low Ee. 2010. English in Singapore and Malaysia: Differences and similarities. In Andy Kirkpatrick (ed.), *The Routledge handbook of World Englishes*. 229-46.

Lyons, John. 1963. *Structural semantics: An analysis of part of the vocabulary of Plato*. Oxford: Basil Blackwell.

McArthur, Tom. 1981. *Longman lexicon of contemporary English*. New York: Longman.

McEnery, Tony and Andrew Wilson. 2001. *Corpus linguistics*. 2nd edn. Edinburgh: Edinburgh University Press.

McEnery, Tony and Costas Gabrielatos. 2006. English corpus linguistics. In Bas Aarts and April McMahon (eds.), *The handbook of English linguistics*. Malden, Massachusetts: Blackwell. 33-71.

McEnery, Tony, Richard Xiao and Yukio Tono. 2006. *Corpus-based language studies: An advanced resource book*. New York: Routledge.

Mehl, Seth. 2013. Thinking linguistically about Keywords: Polysemy, semantic change, and divergent identities. The Keywords Project. http://keywords.pitt.edu/pdfs/thinking_linguistically_about_keywords.pdf

Meyer, Charles F. and Gerald Nelson. 2006. Data Collection. In Bas Aarts and April McMahon (eds.), *The handbook of English linguistics*. Malden, Massachusetts: Blackwell. 93-113.

Meyler, Michael. 2007. *A dictionary of Sri Lankan English*. With Dinali Fernando and Vivimarie VonderPooten. Colombo: Michael Meyler.

Moag, Rodney F. 1982. The life-cycle of non-native Englishes: A case study. In Braj B. Kachru (ed.), *The other tongue: English across cultures*. Urbana, Illinois: University of Illinois Press. 270-88.

Mufwene, Salikoko M. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.

Murray, James. 1900. *The evolution of English lexicography*. Oxford: Clarendon.

Nelson, Gerald. 2006. World Englishes and Corpora Studies. In Braj B. Kachru, Yamuna Kachru and Cecil L Nelson (eds.) *The Handbook of World Englishes*, 733-750. Malden, MA: Blackwell.

Newman, John. 1996. *Give: A cognitive linguistic study*. Berlin: Mouton de Gruyter.

OED Online. Oxford University Press. Accessed 1 December, 2014.

Owusu-Ansah, Lawrence K. 1994. Modality in Ghanaian and American personal letters. *World Englishes* 13 (3), 341-349.

Palmer, F. R. 1981. *Semantics*. 2nd edn. Cambridge: Cambridge University Press.

Platt, John, Heidi Weber and Ho Mian Lian. 1984. *The New Englishes*. London: Routledge and Kegan Paul.

Poutsma, H. 1926. *A grammar of Late Modern English*. Groningen: P. Noordhoff.

Quine, Willard van Orman. 1960. *Word and object*. Boston: MIT Press.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1972. *A grammar of contemporary English*. London: Longman.

Quirk, Randolph. 1990. Language varieties and standard language. *English Today* 6, 3-10.

Rao, G. Subba. 1954. *Indian words in English: A study in Indo-British cultural and linguistic relations*. London.

Ramson, W. S. 1966. *Australian English: An historical study of vocabulary 1788-1898*. Canberra: Australian National University Press.

Rosch, Eleanor. 1973. Natural categories. *Cognitive Psychology* 4 (3). 328-50.

Rosch, Eleanor. 1975a. Cognitive reference points. *Cognitive Psychology* 7, 532-47.

Rosch, Eleanor. 1975b. Cognitive representations of semantic categories. *Journal of Experimental Psychology* 104 (3). 192-233.

Ruhl, Charles. 1989. *On monosemy: A study in linguistic semantics*. Albany: SUNY Press.

Saussure, Ferdinand. 1995 [1916]. *Course in general linguistics*. Trans. R. Harris. London: Duckworth.

Schmied, Josef. 2004. Cultural discourse in the Corpus of East African English and beyond: Possibilities and problems of lexical and collocational research in a one million-word corpus. *World Englishes* 23, 251-60.

Schneider, Edgar W. 1994. How to trace structural nativization: Particle verbs in World Englishes. *World Englishes* 23, 227-49.

Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the world.* Carmbridge: Cambridge University Press.

Schneider, Edgar W. 2008. Clustering global Englishes automatically: Neutral input, meaningful results. In Hans-Georg Wolf, Lothar Peter and Frank

Polzenhagen (eds.), *Focus on English: Linguistic structure, language variation and discursive use.* Leipzig: Leipziger Univeritatsverlag. 51-63.

Schneider, Edgar. 2011. English into Asia: From Singaporean ubiquity to Chinese learners' features. In Michael Adams and Anne Curzan (eds.), *Contours of English and English language studies*. Ann Arbor: University of Michigan Press. 135-156.

Sinclair, J. M. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Smakman, Dick and Stephanie Wagenaar. 2013. Discourse particles in Colloquial Singapore English. *World Englishes* 32, 208-34.

Smith, Nicholas and Geoffrey Leech. 2014. Verb structures in twentieth-century British English. In Bas Aarts, Joanne Close, Geoffrey Leech and Sean Wallis (eds.), *The verb phrase in English: Investigating recent language change with corpora*. Cambridge: Cambridge University Press. 68-98.

Speelman, Dirk. 2014. Logistic regression: A confirmatory technique for comparisons in corpus linguistics. In Dylan Glynn and Justyna A. Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*. Amsterdam: John Benjamins Publishing Company. 487-534.

Taylor, John. 2003. *Linguistic Categorization*. 3rd edn. Oxford: Oxford University Press.

Taylor, John R. 2012. *The mental corpus: How language is represented in the mind.* Oxford: Oxford University Press.

Tuggy, David. 1993. Ambiguity, polysemy, and vagueness. *Cognitive Linguistics* 4 (3), 273-90.

van der Meer, G. 2006. It's about *Time*: On coherence and simplicity in dictionary entries. *English Studies* 87.5, 602-16.

Wackerly, Dennis D., William Mendenhall III and Richard L. Scheaffer. 2008. *Mathematical statistics with applications*. 7th edn. Belmont, California: Brooks/Cole.

Wallis, S. A. *forthcoming*. That vexed problem of choice: Reflections on experimental design and statistics with corpora. London: UCL Survey of English Usage. http://www.ucl.ac.uk/english-usage/staff/sean/resources/vexedchoice.pdf. Accessed 1 August, 2012.

Wallis, S.A. 2014a. What might a corpus of parsed spoken data tell us about language? In: Ludmila Veselovská and Markéta Janebová (eds.) *Complex Visibles Out There*. Proceedings of the Olomouc Linguistics Colloquium 2014: Language use and linguistic structure. Olomouc: Palacky University.

Wallis, S. A. 2014b. Is language really a set of alternations? In *corplingstats*. https://corplingstats.wordpress.com/2014/02/20/is-language-alternations/. Accessed 1 December, 2014.

Weinreich, Uriel, William Labov and Marvin Herzog. 1968. Empirical foundations for a theory of language change. In W. Lehmann and Y. Malkiel (eds.), Directions for historical linguistics. Austin: U. of Texas Press.

Wierzbicka, Anna. 1982. Why can you have a drink but you can't *have an eat? *Language* 58 (4), 753-99.

Wolf, George. 1991. Translator's introduction: The emergence of the concept of semantics. In George Wolf (ed.), *The beginnings of semantics*. London: Duckworth. 3-17.

Woodward, Kathryn. Concepts of identity and difference. In Kathryn Woodward (ed.), *Identity and Difference*. London, Thousand Oaks, New Delhi: Sage.

Xu, Qi. 2013. 'A corpus-based study of the alternating ditransitive verb TELL in native and Chinese learner English corpora'. *ICAME Journal* 37, 185-206.

Yao, Xinyue and Peter Collins. 2012. The present perfect in World Englishes. *World Englishes* 31 (3), 386-403.

Yule, H. and A. C. Burnell. 1886. *Hobson-Jobson: A glossary of colloquial Anglo-Indian words and phrases, and of kindred terms, etymological, historical, geographical and discursive*. London: John Murray.

Zauner, Adolf. 1903. Die romanischen Namen der Korperteile: Eine onomasiologische Studie. *Romanische Forschungen* 14: 339-530.

Ziegeler, Debra and Sarah Lee. 2006. Analysing a semantic corpus study across English dialects: Searching for paradigmatic parallels. In Andrew Wilson, Dawn Archer and Paul Rayson (eds.), *Corpus linguistics around the world*. Amsterdam: Rodopi. 121-39.

Zwicky, Arnold M. and Jerrold M. Saddock. 1975. Ambiguity tests and how to fail them. In John P. Kimball (ed.), *Syntax and semantics*, vol. 4. New York: Academic Press. 1-36.