Are all languages equal?

Lucy Weston-Taylor

Thesis submitted for the degree of Master of Arts (Philosophy)
University of Otago
Dunedin, New Zealand
June 2013
Abstract

There is a general orthodoxy in linguistics that all languages are equal. Only in the past 10-15 years have linguists started critiquing this claim. In this thesis I critically examine structural interpretations of the thesis that all languages are equal. I argue that the claim should be interpreted as a claim about the equi-complexity of language structure, language as abstracted away from usage. There are two structural interpretations of the claim for which there is tentative evidence, but more linguistic research needs to be done to test these properly. The weak equilibrium hypothesis and the double threshold hypothesis both claim that language varieties, when measured according to a commensurable metric, fluctuate around a norm. I also claim that languages with writing systems are more complex than languages without such systems. I conclude that if either of the structural interpretations are true, all languages are equal. However taking written language into account, then not all languages are equal.
Acknowledgements

Thanks to my supervisor James Maclaurin, for his critiques, suggestions, task-mastering, and moral support. It is no overstatement to say that this thesis would be significantly poorer without his contributions.

Thanks to Heather Dyke, for some early comments, and Greg Dawes for invaluable final crafting and shaping advice. Martin Taylor’s heroic last minute proofing was greatly appreciated.

Thanks also to all staff and students of the Otago University Philosophy Department. It wouldn’t have been the same without you all.

And finally thanks to Glenda Weston, Martin Taylor, Emily Weston-Taylor and Frederik Leonhardt for putting up with me through this project. You now know more about the equi-complexity of languages than you ever thought possible. Count yourselves lucky.
1. Chapter One: Introduction

1.1 Introduction

1.1.1 Wider implications

1.1.1.1 Language rights

1.1.1.2 Intra-language dependencies

1.1.1.3 Human nature

1.1.1.4 Interdisciplinary interest

1.1.1.5 Proto-language

1.1.2 An emerging field

1.2 Structural interpretations:

1.2.1 Total grammatical complexity

1.2.2 The minimum argument

1.2.3 The maximum argument

1.2.4 Conclusion:

1.3 Functional interpretations in the literature

1.3.1 What does function mean?

1.3.2 A teleological account

1.3.3 A propensity account

1.3.4 Which accounts of function do we find in the literature?

1.3.5 Conclusion

2. Chapter Two: Assumptions

2.1 Assumption 1: We know what a ‘language’ is

2.1.1 Mutual intelligibility

2.1.1.1 Intelligibility is a continuum

2.1.1.2 Symmetricality is assumed

2.1.1.3 Intransitivity can lead to counterintuitive results

2.1.1.4. Summary of the problems

2.1.2 Isogloss boundaries

2.1.2.1 How to use isoglosses to delineate boundaries

2.1.2.2 Issues with isoglosses picking out linguistic boundaries

2.1.2.3 Conclusion

2.2 Assumption 2: There is a commensurable way to measure language structure

2.2.1 Measuring complexity

2.2.1.1 Absolute complexity

2.2.1.2 Relative complexity

2.2.2 Conclusion

3. Chapter Three: Setting out the structural interpretations
3.1 Force analogies and threshold analogies ................................................................. 45
3.2 The hypotheses ........................................................................................................... 49
  3.2.1 The static value hypothesis: ................................................................................. 50
  3.2.2 The varied value hypotheses: .............................................................................. 50
    3.2.2.1 The weak maximum hypothesis ................................................................. 51
    3.2.2.2 The weak minimum hypothesis .................................................................. 56
    3.2.2.3 The weak equilibrium hypothesis .............................................................. 57
    3.2.2.4 Double threshold hypothesis ..................................................................... 58
3.3 Evaluating the hypotheses ......................................................................................... 59
  3.3.1 The strict equilibrium hypothesis: ................................................................. 60
  3.3.2 The weak maximum hypothesis ...................................................................... 61
  3.3.3 The weak minimum hypothesis ...................................................................... 64
    3.3.3.1 Morphologically isolating ........................................................................ 65
    3.3.3.2 Syntactically monocategorical ................................................................. 66
    3.3.3.3 Semantically associational ...................................................................... 66
  3.3.4 The weak equilibrium hypothesis .................................................................... 68
  3.3.5 The double threshold hypothesis ................................................................... 70
3.4 Conclusion .................................................................................................................. 71
4. Chapter Four: Writing systems ................................................................................... 72
  4.1 Writing systems are a part of a language ............................................................ 72
    4.1.1 Two analogies ................................................................................................. 76
      4.1.1.1 The relationship between plays and films. ............................................... 76
      4.1.1.2 The relationship between live music and recorded music ....................... 77
    4.1.2 Interaction between spoken and written language ........................................ 80
    4.1.3 Effect of written language on speech ............................................................ 81
    4.1.4 Effect of spoken language on writing ............................................................ 83
  4.2 The advantages of a writing system ..................................................................... 84
    4.2.1 Propositional storage ..................................................................................... 84
    4.2.2 Metalinguistic storage .................................................................................... 85
    4.2.3 Storage of cultural information .................................................................... 86
    4.2.4 Writing is subject to different constraints than speech ................................ 87
    4.2.5 Editing ............................................................................................................ 87
    4.2.6 Some languages are better than others ......................................................... 88
5. Chapter Five: Conclusions ......................................................................................... 89
  5.1 Chapter 2 Conclusion ............................................................................................. 89
    5.1.1 Assumption 1 .................................................................................................. 89
    5.1.2 Assumption 2 .................................................................................................. 90
  5.2 Two live structural options .................................................................................... 90
5.2.1 Both hypotheses are plausible................................................................. 92
5.2.2 The hypotheses are empirically different............................................... 93
5.2.3 Current data is inconclusive................................................................. 94
5.2.4 Conclusion .............................................................................................. 95

5.3 Chapter 4 Conclusion ............................................................................... 95

5.4 Summary.................................................................................................. 96
  5.4.1 Metric choice ....................................................................................... 96
  5.4.2 Longitudinal studies ............................................................................ 96

References: .................................................................................................. 98
“Linguists have long argued that no single language is more primitive nor complex than any other” (Barrutia 1967: 23–24).

“It does not make any kind of sense to suppose that any one dialect is in any way linguistically superior to any other” (Chambers & Trudgill 1998: 3).

“Almost all linguists take it for granted that, in some important sense, all languages are equal” (Christiansen & Kirby 2003: 4).

“The statement that ‘All Languages are Equally Complex’ (ALEC) is often portrayed as one of the fundamental tenets of modern linguistics” (Deutscher 2009: 243).

“It is a finding of modern linguistics that all languages are roughly equal in terms of overall complexity” (Dixon 1997: 118).

“A central finding of linguistics has been that all languages, both ancient and modern, spoken by both ‘primitive’ and ‘advanced’ societies, are equally complex in their structure” (Forston 2010: 4).

“The received view among linguists is that all languages are equally complex in their grammars” (Miestamo 2006: 12).

“The practically universal [in linguistics] adoption of the principle of uniformitarianism, namely the idea that all languages are in some important sense equal” (Newmeyer 2003: 63–64).

“The long-received truism that all languages are ultimately about equal in complexity” (Nichols 2009: 110).

“All human groups speak equally well. There are no languages that are superior to others” (Pagel 2013).

A “well-known assumption is that all human languages are overall equally complex. This statement is present in most introductory classes in linguistics or encyclopedias of language” (Pellegrino 2011: 539).

“For much of the twentieth century, linguistics was strongly attached to a principle of invariance of language complexity as one of its bedrock assumptions – and not just the kind of assumption that lurks tacitly in the background, so that researchers are barely conscious of it, but one that linguists were very given to insisting on explicitly” (Sampson 2009: 1).

“As far as current methods of linguistic analysis can determine, all known languages have reached the steady state” (Wang 1976: 62).
1. Chapter One

1.1 Introduction

The thesis that all languages are equal is a central tenet of both historical and modern linguistics. Sampson describes the thesis as a historically “unquestioned truism” (Sampson 2009: 1), and Deutscher claims “such beliefs seem to be dominant in current linguistics” (Deutscher 2000: 165). New students are taught it in introductory classes; linguistic textbooks include such a statement as a matter of course. The thesis that all languages are equal is an assumption that has underpinned linguistic theory since the late 1950s.

The earliest formulation in print is found in *A Course in Modern Linguistics*, (Hockett 1958), and the idea was not controversial when first printed. As such, it is likely the thesis that all languages are equal was common in linguistics before 1958. Hockett mentions a version of the claim that all languages are equal only once (180–181), and does not defend the view. Looking at the responses to his book at that time, there are no comments challenging the claim. There don’t appear to be contrary views at that time in the literature, instead it seems that the claim was widely accepted. “The total complexity scores of related and unrelated languages come out remarkably even, as Hockett and others have pointed out” (Barrutia 1967: 24). Wang states that “all languages are said to have about the same degree of complexity” before further referencing Hockett (Wang 1976: 62).

1.1.1 Wider implications

Often the thesis that all languages are equal is mentioned once in a paper and not referred to again. This lack of explicit argumentation involving the thesis might seem to suggest that the thesis makes no predictions and isn’t used in other areas, in short that nothing rests on the claim. However this is not the case. There are a number of related fields, both within and outside linguistics, where the equality of languages is relevant. In the next section I will briefly set out some of these fields, and explain the relevance of the thesis that all languages are equal.

1.1.1.1 Language rights

Of the 7,000 languages spoken today, a large number are minority languages of which many are in danger of extinction. Language deaths continue at a high rate. A number of people including linguists, speakers of endangered languages and social justice advocates wish to reduce this rate of language death. There are many pushes to revitalise ailing languages. One of the so-called success stories is the revitalisation of the Maori language. However such
revitalisations don’t take place without support. In the Maori example, an increase in the number of Te Reo (Maori) speakers was only achieved with huge community support, implementation of language-specific infrastructure (e.g. Kohanga Reo (Maori language preschools), Maori language TV) and of course government funding.

Governmental funding is crucial to protecting and promoting ailing languages, and such measures require justification. The equi-complexity thesis can be used to justify such funding. While it may not be used as direct evidence for language preservation, it is easy to see how a proof or refutation of the thesis ‘all languages are equal’ might influence this debate.

1.1.1.2 Intra-language dependencies
If all languages are equal, then language components must be closely related to each other. That is, for whatever measure of equality, there must be some dependency between language domains such that they balance each other out. For example, all other things being equal, a gain in complexity in verb conjugation ought to be followed by a loss in complexity in another realm (e.g. plural formation). However, it would be surprising if this were the case, as it is unclear what mechanisms would achieve this purpose. In this respect a positive statement of the thesis would necessitate further research identify these intra-language relationships and the mechanisms by which they interact.

1.1.1.3 Human nature
Sampson suggests that even if nothing explicitly rested on the thesis, it would still be worth investigating, as it deals with a fundamental assertion about language. “There cannot be many current topics of academic debate which have greater human importance than this one” (Sampson 2009: 1). Whilst I think this is slightly overstating the issue, I agree with Sampson’s general point. Language is a fundamentally human trait, other species can be taught rudimentary parts of language, but none can wield language as we do. 99.99% of humans have spoken or signed language; it is enormously rare to find someone without language. Such people typically have some sort of severe disability and/or a severely abnormal upbringing. It takes a lot for a human to not have language. We don’t know, however, just how fundamental language is to humans, this is still a matter of much debate. Because having language is a huge part of being human, any claim about all languages, such as the thesis at issue, is worth investigating, if only for implications about the humans who speak such languages.
1.1.1.4 Interdisciplinary interest
There are a number of linguistics-related fields such as evolution of language, evolution of cognition, philosophy of language, where interdisciplinary input is high. One might think that the thesis that all languages are equal is only talked about (and infrequently at that) in the field of linguistics. However the claim is asserted with such vigour that it is not surprising to see some non-linguists who work alongside linguistics asserting the claim as well. Recently the evolutionary biologist Mark Pagel gave a Royal Society of New Zealand lecture on language. During his talk, Pagel claimed “there are no languages that are superior to others” (Pagel 2013).

1.1.1.5 Proto-language
Major transitions are irreversible changes in the transmission of (typically) genetic information; they mark significant differences in units of selection, e.g. replicating molecules to populations of molecules, and asexual clones to sexual populations. A feature common to major transitions is that “entities that were capable of independent replication before the transition can replicate only as part of a larger whole after it” (Maynard-Smith and Szathmáry 1995: 6). According to Maynard-Smith and Szathmáry, the last major transition is that of primate societies to human societies (14), and the development that made this possible was language.

They explicitly claim that language and not proto-language allowed for this transition, proto-language wasn’t enough for human societies to form. Given this claim, it would seem that Maynard-Smith and Szathmáry would have a well-developed definition of both proto-language and language, and would provide some criteria for distinguishing between the two. However Maynard-Smith and Szathmáry only define proto-language in relation to language. They claim that “the emergence of protolanguage in Homo erectus [is] a cultural inheritance system with limited potential in which only certain types of statement can be made” (14), whereas “human language… [has] a universal grammar and unlimited semantic representation” (14). Maynard-Smith and Szathmáry do eventually give some criteria for a proto-language. They claim that proto-language lacks word-order rules, there are few or no grammatical items and phrase structure is minimal (305). It isn’t clear why these criteria were chosen, and they don’t argue for their criteria.

Like Maynard-Smith and Szathmáry, a large number of evolution of language scholars often discuss proto-languages without having clear definitions of this phenomenon. The transition of proto-language to language is an important step that needs to be investigated. If all languages are equal according to some criterion, perhaps that criterion could be used to
distinguish proto-language from language. At the very least, it could be used to help pinpoint some of the characteristic properties of languages.

A common folk assumption is that some languages are better than others. In contrast, the standard assumption in linguistics, and among scholars working in and around linguistics, is that languages are equal. However none of these groups has properly investigated the thesis.

1.1.2 An emerging field

Looking at the statement that all languages are equal, it is by no means clear how we should measure this ‘equality’. What are the conditions for a language to be ‘equal’? Though the claim that all languages are equal has been in the literature for more than 50 years, discussion of the claim has been oddly lacking. There hasn’t even been any clarification of what the claim actually means. It isn’t clear from the literature how the ‘equality’ of languages ought to be measured, let alone any investigation of whether it is true or false.

In the past 10-15 years a select band of linguists have started tackling the problem. Most linguists who have approached the topic of language equality have only done so after having conducted research where the equi-complexity of language was assumed. Dahl 2009, for example, investigated two closely related languages, and on the basis of his research wanted to claim that one was more complex than the other.

That is, while there are a number of papers and books on the subject, there haven’t been many intra-disciplinary debates. After a conference in 2007 at the Max Planck Institute of Evolutionary Anthropology in Leipzig, one of the first edited books directly addressing the thesis (Sampson et al. 2009) was published. Although the number working in the field is growing, it is still a niche discussion.

Most of the articles and books written in this field have been from linguists putting forward their own positive arguments. As it is such a new field, there is no framework for discussion within the literature. There hasn’t been a critical review of the field, setting out the framework and critically examining the arguments that have been put forward. Before we can really engage with the claim, the conceptual space needs to be mapped out. That is what I hope to achieve in this thesis.

In this thesis I will investigate the linguistic claim that all languages are equal. In chapter one I will present a literature review of the field, explaining the main views. In chapter two I will set out and explain some of the assumptions I have to make in order to engage with the question. In chapter three I will put forward and explain the various potential
structural interpretations of the thesis, after which I will evaluate these hypotheses. In chapter four I will address the issue of written language, and chapter five contains my concluding remarks.

1.2 Structural interpretations:
Scholars agree that the thesis ‘all languages are equal’ needs interpretation before we can assess the claim. Suppose I say to you ‘all holidays are equal’. The statement alone, without any context, is open to multiple interpretations, and as such isn’t a meaningful utterance. On the one hand, there are different things I could mean by ‘holidays’; for instance vacations, statutory holidays, New Zealand statutory holidays or religious holidays. Furthermore, there are different measures by which we can evaluate the above and different ways they can be classified as equal. Is weather important? Is a holiday in summer better than a holiday in winter? Is it important whom you spend the holiday with? Is it just that it’s a day off work? Is a holiday commemorating something important better than a ‘mere’ stat holiday? Is a whole day off better than a half day off? Which measures we choose to evaluate specific holidays will depend on which interpretation of the word we are interested in. Just as ‘all holidays are equal’ has to be interpreted, so too does the thesis that all languages are equal, and indeed scholars (either explicitly or implicitly) put forward their own interpretations of the claim. In this section I will explain what the literature says about the claim that ‘all languages are equal’: in further sections I will examine these theses in more depth.

There are two broad categories in the literature: structural interpretations and functional interpretations. Structural approaches abstract away from usage; they claim that there is some structural measure according to which all languages turn out to be equal. Functional approaches, on the other hand, focus on the performance of a language; what (a) language enables. For example, all languages are equally able to express or communicate information; all languages facilitate cooperative ventures such as cooperative hunting, boat building etc. Most of the discussion so far has involved strict structural interpretations (e.g. Dahl 2009; Miestamo 2006; Deutscher 2009) but there some linguists who have put forward a functional claim in conjunction with a structural claim (Gil 2009, Hockett 1958, Hawkins 2009). I’ll look at the various structural interpretations in the literature first, and then set out the functional interpretations.

There are three structural interpretations of the thesis that all languages are equal that are explicitly expressed in the literature. In chapter three I will introduce some other terms to refer to these positions, but for this section I will use the terms found in the literature. The three interpretations found in the literature are: total grammatical complexity, where language
complexity is equal for all languages, and interdomain complexity balances; *the minimum argument*, where there is a minimum threshold for language structure, and everything over that threshold is equal; and the *maximum argument*, where there is a maximum threshold for language structure. In the following section I will set out the general approaches that scholars are taking, before going into the arguments in more depth in later sections.

### 1.2.1 Total grammatical complexity

Of the structural interpretations of ‘all languages are equal’, total grammatical complexity is discussed far more than others (e.g. Hockett 1958; Kusters 2008; Dahl 2009; Deutscher 2009; Kieraś 2010). Hockett is credited with first formulating this:

> Objective measurement is difficult, but impressionistically it would seem that the total grammatical complexity of any language, counting both morphology and syntax, is about the same as that of any other. This is not surprising, since all languages have about equally complex jobs to do, and what is not done morphologically has to be done syntactically. Fox, with a more complex morphology than English, thus ought to have a somewhat simpler syntax; and this is the case.

(Hockett 1954: 180–181)

Those in favour of this interpretation claim that all languages are equal in their overall complexity. However this isn’t a claim that languages are static; although the theory claims that *overall complexity* remains constant, domain complexity is fluid. A domain is a particular subset of language structure, for example grammar, phonemic diversity, or syntax. Overall complexity is then calculated by aggregating domain complexity. As overall complexity is (supposedly) constant, when complexity in a particular domain increases, this must be compensated for by a decrease in complexity in a different domain, and vice versa.

Sometimes it does seem that a loss of complexity in one domain necessitates an increase in complexity in another. Aitchison discusses the Maori passive, where phonemic change in word complexity has influenced the passive construction:

> Once, Maori words ended in consonants, and the passive was formed by adding [-ia] on the end… Then consonants at the end of words were gradually lost [while] the passive forms have not changed at all... Maori speakers have reanalysed the passive... assuming that the consonant on the end of the verb is part of the passive:
Maori infinitives seem to have simplified over time. Infinitives now all end in a vowel, a characteristic which aids recognition of unknown infinitives. They are also phonologically simpler than their ancestral forms. That is, modern infinitives involve fewer sounds than the ancestral infinitives. Whilst infinitives have (arguably) become simplified, as the above table shows, the modern passive is more complex than the ancestral passive. Passive construction used to treat all infinitives the same, just add ‘ia’. However to construct the modern passive, speakers must now which group an infinitive belongs to, and choose the appropriate passive suffix. The Maori passive is an example of the ‘balancing out’ of complexity in different domains, which the thesis of total grammatical complexity claims is the mechanism that keeps languages equal.

While total grammatical complexity is often discussed, there is no consensus within the literature. In the following I’ll set out the disagreement regarding total grammatical complexity.

Hockett, who first proposed total grammatical complexity as an interpretation of the thesis that all languages are equal, only proposes the thesis tentatively. As we see in the quote above he says, “objective measurement is difficult” and “impressionistically it would seem” (Hockett 1955: 180–181). Moreover, the latter part of his statement makes a separate functional claim (which we will look at later). This is the only time in the book that Hockett mentions total grammatical complexity (by name or concept). It’s not clear how seriously he takes the thesis.
More recently, Aitchison argues for a version of total grammatical complexity, though she doesn’t mention the theory by name.

A language which is simple and regular in one respect is likely to be complex and confusing in others. There seems to be a trading relationship between the different parts of the grammar which we do not fully understand.

(Aitchison 1991: 214)

Aitchison’s version is weaker than Hockett’s; she doesn’t think that all languages always have the same overall complexity: rather that language complexity fluctuates around a norm. “Language is ebbing and flowing like the tide, but neither progressing nor decaying, as far as we can tell” (215).

Dahl and Nichols, in separate chapters agree that total grammatical complexity is the right interpretation of the ‘all languages are equal’ thesis, which they both think is false. Dahl carries out a study to explicitly test total grammatical complexity, and Nichols’ approach is similar. “This chapter is an attempt to test the complexity invariance assumption… the idea is that… it will be possible to identify the points at which the languages differ in complexity and see whether these differences are compensated for elsewhere in the grammar” (Dahl 2009: 50).

As I will explain in chapter three, such arguments only work against the thesis if total grammatical complexity is the correct interpretation. There are of course other scholars who think that total grammatical complexity is not the right interpretation, and offer differing interpretations in its stead.

1.2.2 The minimum argument

Deutscher (2009) introduces the term ‘minimum argument’, but the idea is most clearly expressed in Gil (2009). At its most basic, the minimum argument claims that there is a minimum standard of complexity that languages must reach in order to be considered ‘equal’. Languages can differ in complexity; all that is important is that they meet the standard, and all languages do in fact meet this standard.

One reading of the minimum argument is that it is not actually an interpretation of the thesis that all languages are equal, instead it argues against it. Languages can and do differ in complexity, thus according to the thesis itself, languages are not equal. However I argue that this is the wrong way to interpret this minimum argument; actually there is only one
important criterion for language equality, that languages meet this minimum requirement. Languages can rise above this requirement in different ways; however this is irrelevant to the equality question.

Deutscher and Gil both identify the minimum argument with a related functional claim. Gil claims that his paper will help identify

the necessary linguistic abilities: how much grammar does it really take to build a boat and sail it to a distant island? Or more generally: how much grammar does it take to do all of the things that, amongst all living species, only humans are capable of doing, such as, for example, worshipping God, waging war, and working in offices; inventing, manufacturing, and using sophisticated hi-tech tools; and engaging in multifarious commercial, scientific, and artistic abilities?

(Gil 2009: 20)

For Gil, grammar is complex in order to give humans the ability to do complicated tasks. Structural complexity is what gives a language its functional ability to express or communicate a wide variety of propositions. This imposes some minimal complexity on language, such that there is a significant difference between a language variety, and a language-like system that is less complex than the minimum.

I won’t get into discussion of this functional claim here; I will simply note that the minimum argument is related to a functional claim. I want to sketch out the structural approaches first, and we’ll return to this later when investigating functional approaches.

1.2.3 The maximum argument

The final competing structural interpretation presented in the literature is the maximum argument. There are two versions of this argument, a strict version, where all languages are at this maximum, and a weak version, where some languages are. I will discuss this further in chapter three.

The maximum argument claims that there is a maximum threshold for language complexity; there is an upper limit above which no language can be more complex. McWhorter runs a version of this.

After countless millennia of usage and drift, we might expect that by a certain point all grammars had, by the sheer dictates of chance, developed various random
complexities in parts of their grammars… We might propose that the volume of such excrescence in each grammar eventually reached the limit of human propensity to process it… Under this scenario all natural languages would be equally complex by virtue of having all come to rest at a certain ‘surplus complexity quotient’.

(McWhorter 2001: 131)

McWhorter claims there is a ‘limit of human propensity to process’ language, which imposes an upper threshold to language complexity. Like a lid on a jar, this limits the amount of information that can be stored. Thus no language can be more complex, as that would require more resources than the human brain (at least according to McWhorter) has. Once a language has reached this limit, the only way that new complexity can develop is if complexity somewhere else is lost, freeing up processing space.

However, while he thinks that most languages have reached this upper limit, McWhorter doesn’t think that all languages are there yet. He makes this claim about a certain class of languages, stating that “creoles display less complexity than the rest of the world’s natural grammars” (133).

Note that here, as in the case of the minimum argument, there is a related functional claim, which McWhorter explicitly mentions. The maximum threshold is a cost/benefit trade off. The higher the language complexity, the greater neurological processing ability needed to process it. I’ll discuss the functional claim in the following section.

1.2.4 Conclusion:

In this section I have introduced the three structural interpretations of the thesis that ‘all languages are equal’ which are found in the literature. So far I have merely explained these claims as they are expressed in by their proponents. Later, in chapter three I will set out the hypotheses in more detail, and assess the strength of the various claims. Kieraś (2010) formulates the thesis that all languages are equal as being about grammatical complexity, i.e. what I am calling structural complexity. I agree with Kieraś that the thesis as discussed is about structural complexity. However, some linguists use a functional equality claim to support their structural equality claim. In the following section I will set out some functional interpretations of the equi-complexity thesis.
1.3 Functional interpretations in the literature

While structural interpretations are predominant in the literature, there are some references to functional interpretations. It is important to note that these interpretations, while related to structural approaches, are in fact interpretations of a different thesis. In this section I will discuss the functional claims found in the literature, explain how they differ from the structural interpretations, and explain why such functional interpretations are formulations of a different thesis. I’ll look briefly examine the four functional interpretations found in the literature, before coming back to them in more detail at the end of the chapter.

As discussed previously, Hockett formulates a structural hypothesis; however in that same passage he also puts forward a functional hypothesis. “All languages have about equally complex jobs to do, and what is not done morphologically has to be done syntactically” (Hockett 1958: 180–181). According to this reading of Hockett, complexity in language is necessary for language to carry out its function; it is this function which drives structural complexity.

Both Deutscher and Gil claim similar links between functional and structural measures. Gil’s “How much grammar does it take to sail a boat?”, as the title explicitly suggests, investigates the minimum grammar requirements needed to support human task ability.

It is important to note that claiming that all languages are equal because they have the same (teleologically pinpointed) function does not entail a claim about structural complexity. Though the two claims are commensurable (e.g. Deutscher and Gil), extra hypotheses are needed to claim that the function explains the minimum complexity.

McWhorter claims that there is a maximum threshold for structural complexity because of a functional threshold. He states that there is a “limit of human propensity to process [language]” (McWhorter 2001: 131).

Hudson (1983) claims that linguists talk about potential equality, the idea that languages are not restricted, that they have the potential to be equal. “Linguists have tended to say that all varieties are potentially equal in that any differences between them can easily be wiped out by adaptation on the part of the 'weaker' variety, but in practice differences between actual and potential equality are glossed over” (personal communication). This is a functional claim about the flexibility of languages. A language that is hindered in adapting itself/being adapted (depending on your theory of language change) to novel situations is functionally constrained in a way that a language without this hindrance isn’t.

There are a number of different functional approaches to the thesis that linguists have taken. The functional accounts that each linguist puts forward differ substantively, and yet
they all have something in common. These linguists, though they may mention functional accounts, actually all put forward structural accounts.

Though Hockett makes reference to a functional interpretation, he explains equi-complexity in terms of structural complexity. Indeed the name he uses for his formulation of the thesis is total grammatical complexity. Hockett’s mention of function to support a structural equality claim follows the same pattern used by most linguists who make references to function.

Gil, for example, talks about a functional analysis: “A widespread assumption is that linguistic complexity is necessary in order to support complexity in other [non-linguistic] domains. This accords with a functional approach towards the evolution of language, whereby greater linguistic complexity enables humans to accomplish more tasks, and in doing so confers an evolutionary advantage” (Gil 2009: 19). However when it comes to investigating the thesis, Gil focuses on a structural interpretation, he discusses “the level of grammatical complexity” (20) of various languages.

These mentions of function followed by a structural evaluation might seem to imply that the two are necessarily linked. It might also imply that a functional interpretation of the thesis should be investigated by looking at structural complexity. As I will show in this chapter, functional interpretations are separate to structural interpretations, and can be studied on their own. However as it happens, all the references to function found in the literature consist of a functional explanation for the claim that languages are structurally equal. These accounts claim that some function of language causes all languages to be structurally equal.

In the following section I will separate out functional accounts from structural accounts, and show that the two are not necessarily related. I want to show that functional interpretations are substantively different from structural interpretations, and that the two approaches make different claims. Due to space constraints, I will only sketch out some of the issues involved in looking at functional interpretations of the thesis. Then I will discuss the functional claims found in the literature, discuss which accounts of function they use, and discuss how the authors link their functional and structural claims.

1.3.1 What does function mean?

There are two main accounts of ‘function’ that are applicable to biological items, or traits: a teleological approach, “that analyzes the function of a trait only in terms of those effects of the trait which have in the past contributed to the selection of organisms with that trait” (Allen 2009). This approach characterises the function of an item as that for which an item was
selected. This is contrasted with a propensity approach to function which analyses function as what an item can do. I’ll discuss a teleological account first, and then a propensity account.

Some theorists adopt an etiological or backward-looking approach that analyzes the function of a trait only in terms of those effects of the trait which have in the past contributed to the selection of organisms with that trait. Others adopt a dispositional or forward-looking approach that analyzes function in terms of those effects it is disposed to produce that tend to contribute to the present or future maintenance of the trait in a population of organisms.

1.3.2 A teleological account:

Neander (1991b) gives one formulation of the teleological approach she calls an etiological account. A trait’s function is that which in the past gave it fitness advantages, specifically those fitness advantages which contributed to that trait being selected. “Roughly speaking, on the etiological theory I favour, the proper function of a trait is to do whatever it was selected for” (Neander 1991b: 455). A trait’s function is that for which it was selected. As such, this account only covers traits that were selected for in the past. If a particular trait wasn’t selected for, according to a teleological account, it will not have a function. In the next section I discuss a propensity account of function, which does not have this constraint.

So when we are figuring out what the function of a trait is, we look at that trait’s evolutionary history. Neander gives us an example: “The function of your opposable thumb is to assist in grasping objects, because it is this effect which opposable thumbs contributed to the inclusive fitness of your ancestors” (461).

In order to apply this account of function to language then, we need to know what it was that language did to increase the fitness of our ancestors; why (proto-) language was selected.

Note that I am making the assumption that (proto-) language was (mainly) selected for, i.e. not a spandrel. Some scholars claim that “certain specific aspects of the faculty of language are spandrels—by-products of preexisting constraints rather than end products of a history of natural selection” (Hauser et al. 2002: 1574). If language as a whole wasn’t selected for, and instead certain parts of language are spandrels of various other traits, then there isn’t one coherent ‘function’ that language has, according to a teleological account. In such a case, we’d have to use a different account of function, which I will discuss in the next section.

There are numerous theories of language’s selective history. Számadó & Szathmáry give a summary of the current adaptationist hypotheses of language evolution: those who claim that language was selected for because it increased our ancestors’ fitness.
Theories of language evolution

**Gossip:** menstrual ritual can be a costly signal of commitment; hence participating in such rituals can create female groups of shared interest in which sharing information about the social life of others (i.e. gossiping) can be beneficial [Power 1998].

**Grooming hypothesis:** language evolved as a substitution for physical grooming [Dunbar 1998]. The need for this substitution derived from the increasing size of the early hominid groups, which mean that physical grooming became more time consuming, whereas it was possible to ‘groom’ more than one individual simultaneously via vocal communication.

**Group bonding and/or ritual:** language evolved in the context of intergroup rituals, which first occurred as a kind of ‘strike action’ against non-provisioning males. Once such rituals were established, a 'safe' environment was created for further language evolution [Knight 1998].

**Hunting theories:** ‘our intellect, interests, emotions, and basic social life – all are evolutionary products of the success of the hunting adaptation.’ [Washburn & Lancaster 1968]. Later, Hewes in his paper about the gestural origins of language [Hewes 1973] takes up the idea and argues that the probable first use of language was to coordinate the hunting effort of the group.

**Language as a mental tool:** language evolved primarily for the function of thinking and was only later co-opted for the purpose of communication [Burling 1993].

**Mating contract and/or pair bonding:** the increasing size of the early hominid groups and the need for male provisioning also necessitated 'social contract' between males and females [Deacon 1997].

**Motherese:** language evolved in the context of mother–child communication. Mothers had to put down their babies to collect food efficiently, and their only option to calm down babies was to use some form of vocal communication [Falk 2004].

**Sexual selection:** language is a costly ornament that enables females to assess the fitness of a male. According to this theory, language is more elaborate than a pure survival function would require [Miller 2001].

**Song hypothesis:** language evolved rapidly and only recently by a process of cultural evolution. The theory assumes two important sets of preadaptations; one is the ability to sing; the other is better representation abilities (i.e. thinking and mental syntax) [Vaneechoutte 1998].

**Status for information:** language evolved in the context of a so-called ‘asymmetric cooperation’, where information (that was beneficial to the group) was traded for status [Dessalles 1998].

**Tool making:** assumes a double homology: ‘a homologous neural substrate for early ontogeny of the hierarchical organisations shared by two domains – language and manual object combination – and a homologous neural substrate and behavioural organisation shared by human and non-human primates in phylogeny.’ [Greenfield 1991]

(Számadó & Szathmáry 2006: 557, Box 3)

As you can see, even though these all agree that (proto-) language increased our ancestors’ average fitness, and was thus selected for, they disagree as to what those fitness advantages are. Each of the above theories specifies a different thing “which items of X’s type did to contribute to the inclusive fitness of O's ancestors” (Neander 1991a: 174), i.e. a different function of language according to Neander’s analysis.

It may appear that calculating the etiological function of language involves rejecting all but one of the above scenarios. However this isn’t the case. Yes, an etiological analysis requires a selective history; however this doesn’t mean that there was one and only one
selective pressure on the trait in question. Often a trait has been affected by more than one selective pressure. It is very plausible that this was the case for language. If so, then we aren’t restricted to one of the above scenarios, rather we are looking for a combination of two or more.

For example, suppose that language first arises in accordance with the grooming hypothesis above, and is very shortly after co-opted for gossip. We then have selection pressure for language that approximates grooming analogues, and also pressure for language that enables gossip. In this case we have two sets of costs and benefits for the same trait, and both sets will impact the trait’s fitness.

1.3.3 A propensity account

Having examined the teleological approach to the notion of function as applied to biological traits, I turn now to examine the propensity account. As Neander states: “The main difference between the two theories [the etiological account and the propensity account] is that while the etiological theory says that functions are determined by past selection, the propensity theory says they are determined by aptness for future selection” (Neander 1991b: 455).

A propensity account looks at the capabilities of the system now; it is not concerned with the history of the trait or item. Cummins gives a propensity account that analyses function in terms of capacities of the system. To find the function of a trait or item, you enumerate the capabilities of that trait or item. A full description of its capabilities is the list of the trait’s functions (Cummins 1975: 765).

A propensity analysis of functions of a trait/item will more often than not result in more functions than an etiological analysis of the same trait. This is because most of the time, propensity analyses pick out a class of functions that include the teleologically-specified function. The etiological approach gives a narrow range of functions; only those which contributed to ancestors’ fitness, and so contributed to that particular trait being selected count as functions. If a trait in our ancestors increased their fitness in one particular way, the same trait is likely to have the same effect nowadays. As such, a propensity account will also count it as a function. However a propensity analysis will include more functions than an etiological analysis, as it is not constrained by historical facts.

As Neander says, the etiological function of opposable thumbs is grasping. Our ancestors’ fitness was increased because of superior grasping ability afforded by the opposable thumbs, which explains modern humans’ thumbs.
If we look at opposable thumbs nowadays, applying a propensity functional analysis, we see that grasping is still a capacity of the system: a description of opposable thumbs that didn’t include grasping ability isn’t a good description of the system. As such, grasping is a propensity function of opposable thumbs. There are also other propensity functions of opposable thumbs: e.g. they fit in rounded holes (e.g. bowling balls, rings), they bend at the knuckle, there is a fingernail at the end that grows continuously.

A propensity analysis won’t always pick out the etiologically specified function—sometimes there is a vestigial trait, which no longer does that for which it was selected. The coccyx, or tailbone, in humans is one such example. The etiological function of the coccyx is to support a tail. Our ancient ancestors had tails, and the ancient coccyx was where the tail attached to the body. However humans don’t have tails, the coccyx is a vestigial organ left over from our evolutionary past. As such, a propensity analysis for the human coccyx today is not going to claim that one of its functions is to support a tail, as it clearly doesn’t do that.

To carry out a propensity functional analysis of a trait, we look at the trait now and describe its capacities. The claim that all languages are translatable is an example of a capacity that is a Cummins function: anything you can say in language A can also be said in language B. It is clear that this capacity doesn’t fall under a Neander analysis, as translatability is highly unlikely to have reliably increased fitness in our distant evolutionary past.

1.3.4 Which accounts of function do we find in the literature?

As we have seen, there are competing functional interpretations of the thesis that all languages are equal in the literature. Now that I have set out the competing analyses of function, we can see which analysis each functional interpretation uses.

Hockett and Gil use a teleological functional analysis in their hypotheses (following Neander’s approach). Though he doesn’t state what the function of language is, Hockett clearly sees the function as goal oriented. Language is for something, and in order to reach this goal, language needs to have certain kinds of features.

Gil does narrow down what he thinks language’s function is: “greater linguistic complexity enables humans to accomplish more tasks” (Gil 2009: 19). Like Hockett, Gil thinks that language complexity is related to function; however, unlike Hockett, Gil doesn’t think that all language complexity is due to functionality. Gil expressly states that “the amount of grammar that is needed in order to support the vast majority of daily human
activities is substantially less than often supposed, in fact less than that exhibited by any contemporary human language, and far less than that exhibited by most such languages” (20).

McWhorter and Hudson give us interpretations using a propensity functional analysis. In order to use Neander’s etiological account of function, we have to assume that language was the focus of natural selection, and that the function we are after was what was selected for. This is implausible for both McWhorter’s and Hudson’s accounts.

McWhorter claims that there is a maximum threshold for structural complexity because of a functional threshold. He claims there is a “limit of human propensity to process [language]” (McWhorter 2001: 131). In language’s evolutionary history, it is highly unlikely that it was advantageous (i.e. gave a fitness advantage) to have a limit to language complexity. Far more plausible is the idea that when language ability was undergoing selection pressure, language complexity at that time was nowhere near McWhorter’s proposed limit. In that case, there isn’t selection for a limit on language complexity, instead there is selection pressure against superfluous brain capacity, and thus the limit appears.

As such, McWhorter’s account doesn’t invoke an etiological notion of function, but it does fall under a propensity account. Cummins’ account says that to find a trait’s function you simply have to describe the capabilities of the trait. If McWhorter is right, and there is a maximum complexity threshold, then describing such a threshold will be an important part of a full description of language.

Hudson’s notion of potential equality is the claim that all languages have the ability to adapt or be adapted to any novel situation. Like McWhorter’s maximum threshold, it is unlikely that language underwent selection for this particular utility. However any full description of the capabilities of language (i.e. what Cummins requires for function) will include this.

1.3.5 Conclusion

As I have shown, functional interpretations of the thesis that all languages are equal are significantly different to structural interpretations. However there are only a few functional accounts in the literature, and all of these are used to explain a structural equality. There is no purely functional account of the thesis that all languages are equal. Therefore, in this thesis, I will follow the majority of the literature and focus on structural interpretations of the thesis that all languages are equal.
2. Chapter Two: Assumptions

I want to set out and evaluate the possible structural interpretations of the claims that all languages are equal. However before I can start evaluating those, I need to discuss two assumptions that structural interpretations rest on. As a comparative claim, we need to know which units to compare in order to investigate this further. Secondly, we need a measure of language structure which we can use to compare languages. In this chapter I will explain these assumptions, and argue for particular formulations of these. I will suggest a language definition, and argue for one approach to measuring language structure. Someone who wishes to advance a different structural interpretation can disagree with these specific formulations; however they will have to put forward alternatives.

2.1 Assumption 1: We know what a ‘language’ is

The thesis that all languages are equal is making a comparative claim. To investigate this claim we have to know what is being compared. To decide whether ‘all Xs are equal’ we first have to know what an X is. At the start of the previous chapter, I used the statement that ‘all holidays are equal’ as a crude analogy for ‘all languages are equal’. As I pointed out, ‘holiday’ is open to multiple interpretations (for instance statutory holidays and vacations). Standards to rank these are specific to each interpretation – what makes a good statutory holiday is different to what makes a good vacation. As standards differ with the specific interpretations, we need to define which interpretation we are using so that we can compare like with like, using the relevant standard.

There are a variety of meanings of the word ‘language’ and most are vague. Amongst other things, language refers to a human-specific communication system (humans have language, other animals don’t), specific versions of these communication systems (English vs Maori), a narrower conception of this (NZ English vs American English), one person’s idiolect (Lucy’s English vs Mike’s English) and one person’s idiolect at one point in time. Common to the core of these definitions is the idea that whatever else a language is, it is a means of communication. We could go further; a language is a uniquely human means of communication. However while common to all the meanings of ‘language’ listed above, such a definition is too broad. Humans communicate through various means, and we don’t want to call all of them language (e.g. dance, facial expressions, art).

For this thesis, I’m restricting the scope of the term ‘language’ to natural languages, excluding things like programming languages, logical systems, and created languages (e.g. Klingon and Esperanto). The debate I’m addressing is concerned with natural human
languages, not logically possible languages. As such, if a system is not a natural language, for my purposes it is not a language.

I’ve picked out the class of languages; however I want to be able to further identify individual languages within that class. I want a principle by which I can individuate languages, i.e. I want to know what things to compare to test the thesis that all languages are equal.

2.1.1 Mutual intelligibility

Within the class of natural languages, there are some clearly defined language boundaries. Everyone agrees, for example, that English and Maori are different languages. Extrapolating from these clear cases, linguists use the principle of mutual intelligibility to find language boundaries.

On purely linguistic grounds, two speech systems are considered to be dialects of the same language if they are (predominantly) mutually intelligible.

(Crystal 2000: 8)

One way of looking at this [how to define a language] has often been to say that ‘a language is a collection of mutually intelligible dialects’.

(Chambers & Trudgill 1998: 3)

According to the principle of mutual intelligibility, if person A (who only speaks language B) can’t understand person X (who only speaks language Y), and person X can’t understand person A, then languages B and Y are two different languages. If person A and X both understand each other, then language B is the same as language Y. That is to say, two languages are mutually intelligible if speakers of both languages can understand each other. If neither speaker can understand the other, the languages are mutually unintelligible. Carving up ‘languages’ in this way is roughly equivalent to the folk conception of language. Examples of such languages include English, Spanish, Bavarian, High German (so-called German dialects are not mutually intelligible with High German), and Scandinavian (with the sub-variants Swedish, Danish and Norwegian). Using this interpretation, these are the Xs that the thesis that ‘all languages are equal’ is about.

Invoking mutual intelligibility is a good principle to use to identify the clear-cut cases (e.g. Mandarin and French), but it becomes harder to use when dealing with related
languages. The issue here is one of overlap, or similarity. If two systems overlap 100% (i.e. are mutually intelligible), they are one and the same language variety, if they don’t share any similarities (i.e. mutually uninteresting) they are two languages. However there are a lot of cases where linguistic systems overlap to some degree and the principle of mutual intelligibility is not able to account for these.

There are problems with using mutual intelligibility as a criterion of language boundaries. In this section I will set out and explain three problems, all variants on a theme, then I will put forward a solution that addresses these. The problems are that mutual intelligibility does not give definitive answers in all cases to the question of whether X and Y are different languages.

### 2.1.1.1 Intelligibility is a continuum

There are some clear-cut cases, pairs of languages that are uncontroversially mutually (un)intelligible. While New Zealand English and Australian English, for example, are mutually intelligible, Maori and English, are mutually unintelligible. In such cases we can use the principle of mutual intelligibility to easily delineate the language boundaries. However there are also pairs of languages that don’t easily fit into either extreme – e.g. Samoan and Maori, and NZ English and Glaswegian English.

<table>
<thead>
<tr>
<th>Mutually intelligible</th>
<th>?</th>
<th>Mutually unintelligible</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Zealand English &amp; Australian English</td>
<td>Samoan &amp; Maori</td>
<td>NZ English &amp; Maori</td>
</tr>
<tr>
<td>NZ English &amp; Glaswegian English</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The existence of the pairs in the middle of the spectrum is problematic for mutual intelligibility as a criterion. Mutual intelligibility treats intelligibility as a discrete unit; languages either are, or are not, intelligible. However as examples such as Samoan and Maori show, there are a number of languages that don’t fall neatly into either end of the spectrum. Intelligibility is a continuum, and pairs of languages will fall on different points along the
spectrum. Mutual intelligibility does not provide any guidance on what to say about such semi-intelligible language pairs. Because it simply can’t deal with these cases, it fails as a criterion of language boundaries.

### 2.1.1.2 Symmetricality is assumed

Mutual intelligibility considers pairs of languages, and assumes that the intelligibility of one language to another is symmetric. That is to say, the criterion for languages being mutually intelligible are that both are understandable to the other, and vice versa for unintelligibility. However sometimes intelligibility is asymmetric. A speaker of language A can understand a speaker of language B, however the speaker of language B cannot understand the language A speaker. For example most speakers of German language varieties (e.g. Baierisch,) understand Hochdeutsch (Standard German); however Baierisch is unintelligible to speakers of Hochdeutsch.

There are a number of reasons why a language pair would be asymmetrically intelligible. One possibility is that the two languages are related, and one has undergone a greater change from the ancestor language than the other, in response to various pressures. In such an instance, the slower-changing language could be intelligible to the former, with semi-archaic similarities, however intelligibility in the other direction will be more difficult.

Quite apart from purely linguistic phenomena, there could be other social, political, historical reasons, for the asymmetry. One language could be the language of the media and ofﬁcialdom, the other the language at home. In such a case the various social realms will dictate the necessity for individuals to understand each language.

Regardless of the reason for language pair asymmetric intelligibility, the fact that there are such pairs is problematic for mutual intelligibility. The assumption of symmetric intelligibility means that mutual intelligibility as a criterion does not have an answer for some language boundary questions.

### 2.1.1.3 Intransitivity can lead to counterintuitive results

Mutual intelligibility is not transitive. While this is not a problem in itself, the characteristics I have already mentioned (intelligibility as a continuum, assumption of symmetricality) combined with intransitivity lead to some counterintuitive results. There is a phenomenon known as dialect continua where mutual intelligibility as a criterion of language boundaries fails. Dialect continua occur when a number of geographically-related language varieties are mutually intelligible with their neighbours, but unintelligible to their neighbours’ neighbours.
If we arrange villages along our route in geographical order, while speakers from village A understand people from village B very well and those from village F quite well, they may understand village M speech only with considerable difficulty, and that of village Z not at all. Villagers from M, on the other hand, will probably understand village F speech quite well, and villagers from A and Z only with difficulty. In other words, dialects on the outer edges of the geographical area may not be mutually intelligible, but they will be linked by a chain of mutual intelligibility.

(Chambers and Trudgill 1998: 5)

This diagram shows a number of dialect continua (Chambers and Trudgill 1998: 6).

The chain of mutually intelligible neighbours with mutually unintelligible edges is typically considered to be a problem for mutual intelligibility, as it is unclear where the
language boundaries are. If A and Z are mutually unintelligible, then there must be a language boundary separating them, as they are separate languages. However closer examination shows that each language variety is mutually intelligible with its closest neighbours. There doesn’t appear to be a language boundary at any particular point.

2.1.1.4. Summary of the problems:

Using mutual intelligibility as a criterion of language boundaries has some problems. Intelligibility is a continuum, there are pairs of languages that are neither mutually intelligible nor mutually unintelligible. Mutual intelligibility assumes symmetrically, however there are some asymmetric intelligibility relations. Intransitivity together with the continuum of intelligibility leads to some odd cases. Mutual intelligibility as it stands does not have a response to any of these issues. There are two approaches to take here, both of which address the three problems raised.

One solution is to claim that language is vague. As such, we should expect issues such as those detailed above, and the existence of intermediate language boundaries and dialect continua is not a problem with the theory. There are many definitions of what vagueness is; nonetheless Sorensen claims that “there is wide agreement that a term is vague to the extent that it has borderline cases” (Sorensen 2012). The existence of borderline cases seems to be the problem with intelligibility and intransitivity as discussed previously. However vagueness is a solution that brings other problems with it. I’ll put this aside for the moment to consider the competing solution.

2.1.2 Isogloss boundaries

The basis of this solution is that you can work out ways of comparing the language of different groups without having to know whether those groups are speaking different languages. All you need are populations defined by linguistic phenomena.

The problem of delineating languagehood is irrelevant to investigating the thesis that all languages are equal. This may seem counterintuitive, but we don’t need to say what languagehood is in order to talk about the equi-complexity thesis. The thesis is that all languages are equally complex. All Xs are equally complex. To investigate this we only need to know how to pick out the Xs. One way of doing that is to find out the criterion of languagehood. However that is not the only way.

As I mentioned earlier, we can delineate the class of Xs from non-Xs. The class of ‘natural human languages’ includes all and only all Xs. However we want to know which
parts of this group we can compare to each other in order to test the thesis that all languages are equal.

The solution that I am proposing is that all Xs are bounded by isoglosses. An isogloss is a boundary surrounding a grouping of a particular linguistic phenomenon. In common parlance, what this picks out are dialects, or language varieties. In this section I’ll explain what an isogloss is and how we would use these. Later I will argue that isoglosses are suitable conditions for delineating Xs.

This approach says that an X is any linguistic group that is bounded by at least one isogloss. The claim that all languages are equal is applicable to groupings such as English, NZ English and Parisian French.

ISOGLOSSES, [are] the lines marking the boundaries between two regions which differ with respect to some linguistic feature (for instance, a lexical item, or the pronunciation of a particular word).

(Chambers & Trudgill 1998: 89)

An isogloss picks out the geographical spread of one particular linguistic feature. For example, if you take a map of the South Island of New Zealand and draw a circle around the southernmost quarter of the island, you’ve drawn the isogloss for the Southland Burr – “r” (as in nurse, hearse). An isogloss can map any linguistic difference: pronunciation, word choice, grammatical rules, etc. Often there are a number of linguistic features whose distribution roughly corresponds to the same geographic area. For example the isogloss for the Southland burr is similar to that for the distribution of the word ‘crib’. “The coincidence of a set of isoglosses is called a BUNDLE” (Chambers & Trudgill 1998: 94). These bundles often correspond to what non-specialists think of as language boundaries. The border of France and Spain, for example, will be the approximate location for a large number of isogloss boundaries, a large bundle. The bundle (set of isoglosses) characterises the differences between French and Spanish.
This diagram shows a collection of isogloss boundaries. For the majority of Germany, the different isoglosses line up along a geographical boundary that cuts just north of Berlin. However, these come apart when they reach western Germany, and the Rhenish Fan results. Using isoglosses as a linguistic boundary marker gives a number of options for languagehood. We could pick the dat/das boundary, which roughly goes through Coblenz, for example. The crucial issue is that any boundary based on linguistic difference is acceptable. Isoglosses solve the boundary problem for languages by treating any linguistic difference as a potential boundary.
2.1.2.1 How to use isoglosses to delineate boundaries

I’ll use generations as an analogy to show how we can use isoglosses to pick out units to compare. We talk about different generations as if they are discrete units. ‘Baby-boomers are more entrepreneurial than Gen Y’, ‘Gen X don’t realise how good they have it’. However the boundaries around a particular generation are not clear-cut. There isn’t one specific date marking the end of the Baby-boomers from the start of Gen X. And yet we can still carry out meaningful discourse about generational differences. We can even carry out research to support claims about different generations. How do we do this if we can’t pinpoint the generational boundaries? The solution is to use birth dates, instead of generation-membership, to pick out groups to investigate.

The generations (e.g. Baby boomers) are analogous to language varieties that fit the languagehood criteria (e.g. English, Maori). These categories don’t have clear boundaries, and there are often disputes around what should or shouldn’t be included in a particular instance. However even though generational boundaries are not well-defined, we can still engage in
generational comparisons. Paradoxically, demographers achieve this by bypassing ‘generational’ investigation. They acknowledge the fuzzy boundaries, and instead specify groups of birth dates.

As a society, we have labelled the generations of the 20th Century. However, those labels and the years those labels represent are often inconsistent… The two generational groups prevalent in today’s workforce are called the Baby Boomers (Boomers) and Generation X (Gen X-ers)… Although there is agreement as to the labels for the Boomers and Gen X-ers, there is little agreement on the years encompassing them. The Boomers’ birth years are variously reported to begin anywhere from 1940 to 1946 and to end in 1960 or 1964. There is even less agreement for the Gen X-ers’ birth years, reported to begin somewhere in the early 1960s and end in 1975, 1980, 1981, or 1982.

(Smola & Sutton 2002: 364)

Smola and Sutton carry out a study and draw conclusions about Baby boomers and Gen X-ers by comparing a 1974 survey with a 1999 survey. They use birth dates to draw conclusions about generations. “Analysis of the 20 items of this study… suggest that Gen X-ers’ work values are significantly different from those of the Baby Boomers” (378).

Demographers sidestep the question of defining generational boundaries by focusing on birth dates. Using isogloss boundaries to pick out language varieties, and putting the question of languagehood to the side is closely analogous. Isogloss boundaries and birth date ranges are hard-edged. Given the relevant information, there is a definitive answer as to whether something fits within the category or not. Languages and generations, on the other hand are much broader, fuzzier categories. It is possible to know all the relevant information, and still not know whether something fits within the category. Avoiding definitional issues with the fuzzy categories, we can still investigate these by way of using the hard-edged categories.

This is not to say that there are no varieties that fit the languagehood criteria (things we commonly call ‘languages’). My claim is that defining languagehood is a separate issue, and not one that I need to solve in order to look at the thesis that all languages are equal. All things we call languages (such as English, Maori and German) are language varieties. If we investigate enough language varieties, the conclusions we draw from those will stand for the smaller group that fulfill the languagehood criteria. Just as demographers can draw conclusions about generations from birth date groups, we can draw conclusions about
Languages from investigating language varieties.

I want to use this diagram to show how we can use isoglosses to pick out units to compare. We don’t need to know how to delineate different generations (i.e. we don’t need to know how show the sharp boundary between Gen X and Y). Instead we can use an isogloss analogue to pick out age groups. So we group people based on dates of birth (e.g. 1950–1980, 1950–1955, and 1980–1982). By varying the time span we can pick out larger or smaller groups of people.

We can use isogloss boundaries to pick out large linguistic systems (e.g. the spread of the construction “I haven’t”) or smaller linguistic systems (e.g. the spread of “eh?” as a tag question). There are many potential isogloss boundaries; any linguistic property that some but not all linguistic systems have is a potential isogloss boundary. The principle is agnostic with regards to the importance of linguistic difference: a boundary between two dialects (e.g. NZ English and Australian English) is just as much a boundary as something more substantial, such as that between English and Maori. We can sidestep the problem of languagehood, and yet still be able to test the thesis that all languages are equal.

Linguists typically use the term language variety to refer to any linguistic system they wish to pick out.

We shall use ‘variety’ as a neutral term to apply to any particular kind of language which we wish, for some purpose, to consider as a single entity. The term will be used in an ad hoc manner in order to be as specific as we wish for a particular purpose. We can, for example, refer to the variety ‘Yorkshire English’, but we can equally well refer to ‘Leeds English’ as a variety, or ‘middle-class Leeds English’ – and so on.

(Chambers & Trudgill 1998: 5)

In this thesis I will define a language variety as any linguistic group bounded by an isogloss. I’ll use this term from now on to make it clear that I’m using isogloss boundaries to pick out the comparable units. The claim that all languages are equal is to be understood as the claim that all language varieties are equal. Importantly, note that this thesis is actually broader than the former; it includes claims about things we call languages (English, Maori, German) as well as claims about other units (Southland NZ English, Yorkshire English, Baierisch). Investigating the claim that all language varieties are equal includes investigating the claim that all languages are equal.
2.1.2.2 Issues with isoglosses picking out linguistic boundaries:
In this section I’ll set out some possible objections to using isoglosses to delineate linguistic boundaries, and explain why they fail.

2.1.2.2.1 It is arbitrary
One response might be to reject an isogloss-based approach to language variety boundaries, because it appears to be arbitrary. An isogloss can be based on any linguistic difference. Isoglosses flatten out the significance of differences; they provide no justification for claiming that one linguistic difference is more important than another.

It is undeniable that some isoglosses are of greater significance than others, in the sense that some mark distinctions ‘felt’ to be culturally important while others do not, some persist while others are transitory, and the like. It is equally obvious that some bundles are more significant than others, in the same sense. Yet in the entire history of dialectology, no one has succeeded in devising a satisfactory procedure or a set of principles to determine which isoglosses or which bundles should outrank some others.

(Chambers & Trudgill 1998: 96–97)

This doesn’t present a problem for my thesis. I want to be able to compare languages to investigate the claim ‘all language varieties are equal’. Isoglosses that mark so-called ‘significant’ boundaries might be more interesting to investigate than less ‘significant’ ones but we can compare them all. Having multiple language varieties as bounded by isoglosses doesn’t impact negatively on the project. Using isoglosses to determine language variety boundaries deliberately sidesteps issues of degrees of linguistic significance. This allows us to portion off that question and focus on what it means to be equal.

2.1.2.2.2 It is subjective
For any isogloss, there is another that could have been created. The existence of any one isogloss is dependent on a linguist deciding that that linguistic difference is important. As such, the existence of isoglosses is not representative of linguistic differences per se, rather, it is representative of linguistic differences that linguists deem important. Carving up languages according to isogloss boundaries equates to finding language boundaries according to linguists’ interests.
Though this looks like a theoretical concern, it is really a practical concern which isn’t important here, as there is no restriction on isoglosses. If there isn’t an isogloss already in place, and someone thinks there is a linguistic difference, they can simply create a new isogloss. Moreover, this is only pressing in cases where two groups have almost identical language use. So for our purposes this isn’t a pressing concern. The cases where two groups have almost identical language use are unlikely to be cases where someone would want to say that one language is ‘better’ than the other.

### 2.1.2.2.3 It is geographically bound

There are two concerns here; this approach ignores non-geographically-bound linguistic differences; and, this approach doesn’t account for emigration.

Isoglosses map linguistic differences to geographical areas. This approach may work for a number of linguistic differences, but it ignores linguistic differences that are correlated to things other than geography. In cities and towns there are often linguistic differences between the language varieties spoken by different social classes. An isogloss won’t be able to pick out these differences, as the linguistic difference doesn’t correlate to geographical locations.

The solution is to broaden the definition of an isogloss. We’re interested in any collection of natural language where you can draw a boundary of linguistic difference. Any time we have two speaker populations with a linguistic difference between them, we have a linguistic boundary. “What is important is that that variability correlates with other factors, such that certain variants are more closely associated with one village than another, or with labourers more than managers, or with people speaking to close friends rather than to strangers, or with some other factor” (Chambers & Trudgill 1998: 70). When it comes to investigating the thesis that all language varieties are equal, we can compare anything that has a linguistic boundary around it.

Isogloss boundaries are idealised concepts. A man raised and living in Scotland, who moves when he is middle-aged will not be considered a representative of the New Zealand accent, even if he happens to be in New Zealand. Any isogloss will have to identify and remove such instances from the analysis. However this issue isn’t as dire as it might initially seem. Isoglosses are constructed from linguistic data. Suppose a linguist collects such data from 100 people. If one or two of these are recent immigrants, this won’t contaminate the whole data. Once you look at the data as a whole, it will be obvious that these one or two are outliers, and the general pattern of the area will be visible nevertheless. If there is a large enough immigrant community, however, the analysis will count this variant as a significant...
part of the language variety in that area. Rather than being a flaw, this is exactly what we want an isogloss to measure. An isogloss of the English language around South Auckland will measure a number of traits traditionally associated with Pacific languages. Though these traits may have arrived in Auckland from a different background than English, these traits characterise some of the language typically found in South Auckland. Thus an isogloss that includes such traits is exactly what we would want an isogloss to cover.

2.1.2.3 Conclusion
I have argued that language variety boundaries can be identified using a version of isoglosses. Mutual intelligibility fails as a language boundary criterion. Instead, according to population isoglosses we have a multiplicity of possible language boundaries which don’t all accord with the non-specialist conception of language boundaries.

However the criterion of mutual intelligibility is still useful. It picks out the cases that are interesting to compare. Mutual intelligibility picks out pairs of language varieties with significant linguistic differences from each other. This doesn’t mean that these cases somehow ‘more’ of a language than other cases, it simply picks out the more difficult comparative cases. Mutual intelligibility acts as a useful heuristic when choosing what to compare.

To investigate the thesis that all language varieties are equal I need a definition of the comparative unit. Mutual intelligibility seemed like a good candidate; however there are problems with this approach. The problems force a choice to be made between vagueness and abandoning the criterion. I argue that mutual intelligibility as a criterion should be abandoned, though it can be a useful heuristic. A modified isogloss criterion is all that is needed to find a language boundary. As isoglosses bound any system with a linguistic difference, they pick out more systems that satisfy the languagehood criterion. Every language is a language variety, but not every language variety fulfils the languagehood criteria (whatever these criteria are). To avoid confusion I will refer to the systems picked out by isoglosses from now on as language varieties.

2.2 Assumption 2: There is a commensurable way to measure language structure
All structural hypotheses assume that there are commensurable measures of language structure. That is, central to the hypotheses is the idea that language structure is measurable and comparable across different language varieties. In this section I will explain why this
assumption has to be made, argue that complexity is the right way of measuring language structure, and then discuss two possible types of complexity to measure.

When comparing a specific language trait, e.g. phonology, you need to measure the trait you’re investigating. When Miestamo claims that “a language that has 34 phonemes, e.g. Kwazá (Kwaza; van der Voort, 2004: 45-46), has a more complex phoneme inventory than one that only has 18, e.g. , Tauya (Trans-New Guinea, Madang; MacDonald 1990: 21-31)” (Miestamo 2008: 24), he is assuming that there is a cross-language way of measuring phonemes. For Miestamo’s comparison to hold, we have to assume that the thing he is counting (phonemes) picks out a class of things that is not defined in relation to one particular language variety, i.e. we have a definition of what a ‘phoneme’ is; that we know that different language varieties can have phonemes, and that we can compare phonemes between language varieties.

One might think it’s simple to compare phonemes across language varieties, however actually this isn’t as straightforward as it seems. A phoneme is a “distinctive sound: that is, [a] sound which, being used instead of another, in the same language, can change the meaning of a word” (International Phonetic Association 1999: 27). It is a sound that is a meaningful contrast in a specific language variety - phonemes describe a relationship between sounds. The letter ‘b’ in English, for example, is meaningless without reference to the Latin alphabet or spelling conventions. Just as we use and understand ‘b’ in relation to the alphabet and spelling, so too are phonemes to be understood in relation to other sounds in the language variety (which are analogous to the alphabet), and how sounds are put together (which is analogous to spelling rules).

When a symbol is said to be suitable for the representation of sounds in two languages, it does not necessarily mean that the sounds in the two languages are identical. Thus [p] is shown as being suitable for the transcription of pea in English, and also for pis in French... but the corresponding sounds are not the same in the two languages.

However the differences between the consonant at the start of the French pis and the English pea are so small that we can talk about the phoneme [p] and group them together in a meaningful sense. Using the French [p] instead of the English [p] when pronouncing pea does not change the meaning of the word. Phoneme specification does not claim that there is a
phoneme for every distinct sound; rather there is a phoneme for every meaningful distinction in sound.

The use of the phrase ‘distinctive sound’ above implies that there are other sounds which do not change the identity of a word, sounds which are not ‘distinctive’ in this technical sense. Central to the notion of a phoneme is the recognition that many finely distinct sounds can be phonetically identified which do not have the word distinguishing role of, say, English /k/ and /t/ (as in /ki/ key vs. /ti/ tea). For instance, the English /k/ phoneme is made with a tongue closure further forward in the mouth before a front vowel, (such as the /i/ of key) than before a back vowel (such as the /ɔ/ of caw). But crucially it is not possible; in English, to exchange these two varieties of /k/ to make two new words, so the two varieties of /k/ are not ‘distinctive’ in English.

Phonologists (those who study phonemes) and phoneticians (those who study phonetics) can acknowledge that a given phoneme (e.g. [p]) does not specify an exact sound, and yet talk meaningfully about that phoneme. For example the phoneme [p] is defined as a voiceless bilabial plosive (169). This definition doesn’t pick out an exact sound – for example, differences in tone and length aren’t specified; however the definition does pick out a specific class of sounds. We can talk meaningfully about different phonemes and we can count phonemes and compare the number of phonemes in a language variety. It isn’t straightforward to make cross-language phonemic comparisons, but we can nevertheless make such comparisons.

As shown, the seemingly easy cross-language comparison of phonemes turns out to be harder than first thought. This suggests that justifying a cross-language comparison of language structure is also going to be harder than anticipated.

In order to claim that language variety A has a better language structure than language variety B, we need to be able to measure this ‘language structure’. There needs to be a measuring stick to compare language varieties against. Here’s a simple analogy to explore what a metric for measuring language structure should take into account. I have created a metric to measure people’s hairstyles. It covers all the important parts of a hairstyle, and aggregates them into one easy to understand number. You calculate your score by aggregating where you lie on a number of different scales. Moreover the metric is not hairstyle specific: you can measure different people’s hairstyles and compare them with respect to the metric.
To calculate your hairstyle number, find your score for colour, length, texture and thickness, and add them together. For example, my score is fourteen.

However the scores that this metric gives aren’t very meaningful, aggregating the individual scales obscures information. Knowing that my score is fourteen doesn’t give you much information about my hairstyle. Such a score does rule out some combinations (eg. that I have white, bald, straight and thin hair, or that my hair is black, long, very curly and thick), and in this sense the metric isn’t exactly uninformative; scores calculated from the metric do rule out some styles in a minimal way. However while the metric can rule out some extremes, it doesn’t positively identify most styles. Moreover because the metric doesn’t tell you anything about my hairstyle, comparing people’s styles according to this metric is also not useful. If I tell you that my friend has a score of twelve, there isn’t anything you can infer about the differences and similarities of our hair. Is her hair longer than mine? What length and colour are our respective hairstyles?

There seems to be something odd going on here. Knowing how the metric works, and knowing a person’s score is not very informative. However knowing individuals’ scores on the parts of the metric is informative. When I tell you that my colour score is 14 you have learned information about my hairstyle. Somehow individually informative data aggregated with other informative data turns into meaningless data. The problem is that the individual scales aren’t related to each other. As such, aggregating them actually obscures the information.
In order to investigate structural interpretations of the thesis that all language varieties are equal, I have to assume that there is a metric for measuring language structure. As I have just explained, what the metric measures must be commensurable across languages, and it must track a real phenomenon. In the rest of the thesis I’m going to refer to a metric that satisfies these criteria as a commensurable metric.

2.2.1 Measuring complexity

Most linguists (e.g. McWhorter, Sampson, Miestamo, Dahl, Nichols, Chipere and Deutscher) refer to linguistic complexity when talking about the equality (or otherwise) of language varieties. Something more complex has a ‘higher value’ than something less complex. Sampson for example, interprets the thesis that all language varieties are equal explicitly in terms of complexity. “For much of the twentieth century, linguistics was strongly attached to a principle of invariance of language complexity as one of its bedrock assumptions” (Sampson 2009: 1). Deutscher echoes this: “the statement that ‘All Languages are Equally Complex’ (ALEC) is often portrayed as one of the fundamental tenets of modern linguistics” (Deutscher 2009: 243).

The way it is presented, it often seems as if the claim that all language varieties are equal has to be interpreted in terms of complexity. Miestamo, for example, when summarising the debate puts it in terms of complexity: “The consensus that all languages are equally complex is being challenged by a growing number of authors” (Miestamo 2008: 23). There are other ways to interpret the claim, however it does seem that complexity is the best fit.

Instead of invoking the notion of complexity, we could count the parts of a language variety. One problem with counting parts is that different language varieties can use different numbers of parts to achieve the same goal, and we don’t always want to claim that the more parts are harder. For example, there is a word in German, “Schadenfreude” that translated into English means “taking pleasure in the misfortune of others”. English takes seven words to relay a concept encapsulated in one German word. If we were merely counting parts of a language variety, on our language structure scale the English phrase has a higher score than the German word. However this doesn’t seem right. Both utterances are expressing the same concept, and it seems that one should not have a higher score for merely having more ‘parts’.

As Miestamo points out, there are some problems with this approach. “Obviously, the idea behind counting parts of systems does not always mean looking at lists of elements that make up an inventory, and the parts are not always as straightforwardly countable – this idea
is to be taken in a more general sense” (24). However it is not clear exactly how to take this in a more general sense.

### 2.2.1.1 Absolute complexity

There are two different types of complexity that we could use to measure languages: absolute complexity and relative complexity. In the following sections I will introduce and explain both types, and argue which should be used to measure language structure.

Miestamo summarises the differences between the two types of complexity.

> The absolute approach defines complexity in objective terms as the number of parts in a system, of connections between different parts, etc. The relative approach to complexity defines complexity in relation to language users: what is costly or difficult to language users (speakers, hearers, language learners) is seen as complex.

(Miestamo 2009: 81)

Absolute complexity is measured by looking at language as a system, abstracted from usage. For example, you could choose to measure verbs, nouns, grammatical rules, phonological rules, and/or a combination of these. “The basic idea behind the absolute approach is that the more parts a system has, the more complex it is” (Miestamo 2008: 24). Absolute complexity can “be seen as a measure of the content that language learners have to master in order to be proficient in a language, in other words, the content of their competence” (Dahl 2009: 50).

Using absolute complexity to measure language is a decision about what approach to take. It doesn’t give you a metric, however. Whether you choose to count the brute number of grammatical rules, Greenbergian implicational hierarchies, syllable structure, or clausal embedding complexity, to name some examples, comes down to decisions about how to measure absolute complexity. Central to all these metrics is the idea that language complexity is to be measured.

Which parts to measure is an issue of metric choice, which I’m not going to get into. I simply want to demonstrate that it is plausible that there is a way of measuring language structure so that you can compare different languages. In the next section I’ll explain the alternative approach to measuring language complexity. I’ll argue that it is not suitable for our purposes as a way of measuring language structure.
2.2.1.2 Relative complexity

Relative complexity, on the other hand, refers to the cost or difficulty of the language variety for the user. However, as Miestamo points out,

> the question “complex to whom?” is central to the relative approach to complexity. Whether a phenomenon is to be seen as simple or complex, depends on whether one takes the point of view of the speaker, hearer, L1 acquirer [first language learner, usually a child] or L2 learner [second language learner].

(Miestamo 2008: 25)

In the following paragraphs, I will set out six language roles that could be the ‘language user’ for computing relative complexity, and I’ll argue that none of these are suitable candidates by themselves. Then I’ll discuss whether aggregating some or all of these could develop a useful metric.

The first candidate I’ll investigate is a child learner. Children play an important part in language transmission, and what children find easy or difficult is going to have a significant impact on the language variety. However, even though children’s capabilities are a crucial bottleneck for language, a child is not a suitable candidate for evaluating the complexity of a language variety. In order to explain why, I have to explain what linguistic data we could get from children. There are two types of data that would be relevant here. The first is the average age at which children learn and/or master a particular linguistic feature. The second is the percentage of children who master a particular feature (if not all children achieve this). For linguistic features that generally some but not all children learn, the percentage of children that do learn the feature will be of interest.

For any actual child learner data that we collect, there are going to be at least two explanations. The first explanation is that the child language data is explained by some linguistic facts. If a particular construction typically appears in children’s speech only when they are eight or nine years old, we want to say that particular construction is difficult for children, therefore it appears reasonably late. From this interpretation of the data, we could then claim that that particular construction is linguistically complex (without qualification).

However there is a competing explanation for the child learner data that doesn’t refer to any linguistic facts, instead it appeals to developmental processes. An explanation along these lines says that rather than telling us about linguistic difficulty, child learner data tells us about children’s cognitive capabilities. The fact that a certain grammatical utterance typically isn’t heard until a child is eight then, should not be taken as evidence for linguistic difficulty,
but rather as evidence of the fact that the concept is conceptually difficult. Young children don’t typically have any interest in thinking of such a thing, let alone wanting to express it. Any language patterns that are found in children’s language have a competing explanation that refers to children’s development. As such, child learners don’t seem like a useful candidate for ‘language user’, when we’re examining the thesis that all language varieties are equal.

An adult, on the other hand is capable of thinking of those concepts that might be hard to express in language. So we won’t have the same problem that we came across with children.

As we’ve already seen, Miestamo draws a distinction between language speakers and hearers. What counts as language complexity differs for the two groups, and can even be in conflict. As such, we shouldn’t look at an ‘adult user’, because this conflates the two roles. Redundancy, for example, increases speaker complexity while making comprehension easier for the hearer. In the German sentence ‘Ich fahre Fahrrad’, (I ride a bike) both ‘Ich’ (I) and ‘fahre’ (to go, first person singular) are needed to signal the subject of the sentence – the first person singular. In Spanish it isn’t necessary to have a grammatical subject of a sentence, so the German sentence can be translated as ‘Voy en bicicleta’ (lit. to go, first person singular, by bike). The German sentence contains redundancy; the subject is signalled twice, both by ‘Ich’ and by ‘fahre’. The German sentence is more complex for the speaker than the Spanish. However arguably this decreases complexity for the listener, or at least increases ease of comprehension. If the listener mishears either ‘ich’ or ‘fahre’, the other word still contains information about the subject. In Spanish, on the other hand, if “voy” is mumbled, or not heard for any reason, the rest of the sentence doesn’t contain any clues about the subject for the hearer.

Homophones, on the other hand, are an example of something that increases complexity for the hearer but not the speaker. Homophones are different words that sound identical for example, the utterances “the cat’s bed” and “the cats’ bed”. Without further knowledge, the hearer is left with two meanings congruent with the utterance, and no way of differentiating between them.

However this isn’t to say that speaker and hearer complexity is always in conflict. Indeed the very point of most interactions between speaker and hearer is to share information, so there is a strong incentive for cooperation. Speakers want to decrease hearer complexity, so that hearers can understand, and hearers work hard to understand speakers’ utterances, so that they (the hearers) don’t have to constantly interrupt the speakers. In this vein, tongue twisters and hard-to-pronounce words increase both speaker and hearer complexity; the speaker is faced with pronunciation difficulty, and the hearer has to make sense of something that can
often sounds like gibberish. Shared referents, e.g. ‘he’ and ‘that’, can decrease both speaker and hearer complexity. The speaker can simplify their utterances, and the hearer can take advantage of the simplified utterance.

Speaker complexity and hearer complexity are separate measures, and while they are often in agreement, they can conflict; they have a degree of independence from each other. So, given that we can’t use an adult with the target language variety as their first language as our ‘language user’, should we take an adult speaker or an adult hearer as the ‘language user’? I don’t think we should use either. Choosing one privileges that language role over the other, and I don’t think this is justified. It doesn’t make sense to consider speakers in isolation, and hearers in isolation; language is a cooperative venture between speakers and hearers, both are necessary and integral roles.

Neither speakers, hearers, adults with the target language variety as their first language, or child learners seem suitable for our ‘language user’. What then about adults learning the target language variety after having already acquired a first language variety (a second language learner)? Miestamo is sceptical; “if we want to reach a definition of complexity that could be applied as generally as possible, the primary relevance of L2 learners is by no means obvious” (Miestamo 2006: 15). He points out that language varieties must have speakers, learners, adult users and child learners – all of these roles are central to language varieties, whereas L2 learners are not. As such, Miestamo thinks that taking L2 learner complexity as the complexity of a language variety is an error. While I’m tempted by this position, I don’t need to take a stance on this, as there is another way to reject measuring complexity by looking at L2 learners.

How easy it is for a L2 learner to pick up a target language variety depends on how close the target language variety is to their first language variety. An Italian speaker will find it much easier to learn French as a second language than a Mandarin speaker; if your first language variety is Samoan, it will be easier to learn Maori than Latin. So how do we measure L2 complexity of a language? Different L2 learners will have difficulty with different things dependent on their first language variety, so if we take data from all L2 learners, we’ll end up with a mishmash of ‘complex’ things.

I’ve argued that no specific user-based relative complexity measure is a suitable method for measuring language structure for our purposes. If we take a speaker or hearer as the basis for measuring language complexity, this won’t give a general answer to the question ‘are language varieties equal?’ Instead we’ll come out with a conclusion somewhat like ‘language varieties are (not) equally complex for speakers’ – a statement about a specific class of user. It might turn out that statements about classes of users are as general as we can
get, that there is no user-neutral claim to be made about the equi-complexity of languages. However it would be pre-emptive to abandon the search for a user-neutral claim based on issues with a relative complexity approach. Instead, we should look at a different approach to measuring language complexity, and see if we can make a user-neutral claim from that standpoint. Only if it turns out that no such general conclusion can be reached, should we can return and look at claims about classes of users.

Following a similar discussion, Miestamo claims that we should abandon relative complexity altogether. “There will always be some conflict between definitions of complexity based on different types of users, and no general user-type-neutral definition is possible” (15). However this is a pre-emptive move. So far we haven’t considered any general user-type neutral definitions. The problem with the relative complexity measures I’ve considered is that language difficulty is itself relative to different types of user. As such no one user can capture general language complexity. However there is another way we could use relative complexity. We could aggregate a number of specific-user-based metrics, to get one overall metric. This avoids the problems that individual relative complexity metrics ran into, namely that we are elevating one language role over others unjustifiably. However it is unclear whether we can aggregate these user-based difficulty measurements.

As I have shown, there are informative and uninformative aggregations. In the following passage I’ll examine two example metrics that are examples of each type of aggregation. These will help specify what makes an aggregation informative. After considering these, I will then return to aggregations of various user-type relative-complexity metrics. And we can see that such an aggregation is likely.

Mike’s sister (who looks nothing like him) is arriving at the train station, and he sets his friends a challenge. Mike wants to see whether his friends can pick his sister out from a crowd. He tells Laurel his sister’s height (1.6 m) and weight (81 kg), and he tells Juan his sister’s Body Mass Index (25).1 Clearly, Laurel is going to be much more likely to spot the sister than Juan, while Juan is unlikely to be able to pick her out. The problem is that there are a number of height/weight combinations that give a BMI of 25. Aggregating height and weight loses information, so the resulting figure is not specific enough for Juan to find the target. The BMI information picks out a narrow range of people, but this range is much larger than the range specific by Laurel’s information. Here’s a list of some other height/weight combinations with a BMI of 25.

---

1 “Body Mass Index (BMI) is a simple index of weight-for-height that is defined as the weight in kilograms divided by the square of the height in metres (kg/m²)” (World Health Organisation website).
<table>
<thead>
<tr>
<th>Height (m)</th>
<th>Weight (kg)</th>
<th>BMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>1.8</td>
<td>81</td>
<td>25</td>
</tr>
<tr>
<td>1.6</td>
<td>64</td>
<td>25</td>
</tr>
<tr>
<td>1.4</td>
<td>49</td>
<td>25</td>
</tr>
<tr>
<td>1.2</td>
<td>36</td>
<td>25</td>
</tr>
</tbody>
</table>

In this case the aggregation isn’t very informative; the information lost is crucial. However in other situations we do seem to find the BMI informative. Medical professionals use the BMI as a health predictor. Aggregations are not informative in a vacuum. Whether an aggregation is informative or not depends on its purpose. As a future health predictor, the BMI is a reasonably informative aggregation, but it is not a good aggregation for identifying strangers.

There are some cases where aggregating metrics makes sense. For example, in the colder months, the MetService (New Zealand Meteorological Service) gives a temperature prediction, a wind speed prediction, and a wind chill prediction. ‘Wind chill’ (what it ‘feels like’ outside) is calculated from temperature and wind speed input.

\[
W = 13.12 + 0.6215 \times T - 11.37 \times K^{0.16} + 0.3965 \times T \times K^{0.16}
\]

Where \( W \) is the wind chill in degrees Celsius, \( T \) is the air temperature in degrees Celsius, and \( K \) is the average wind speed in km/h at a standard height of 10 metres above ground. Note that wind chill is only defined for \( K \) with a minimum speed of 5 km/h, and is only designed for air temperature of 10\(^\circ\)C or less.

(McDavitt 2010)

This is an example where two separate measurements (temperature and wind speed) can be aggregated to give an informative result; in this case the aggregation is actually more informative than the separate two metrics. If you know that it is 10 degrees Celsius outside, and that there’s a 20km/hr wind, hearing that the wind chill is 7.4 degrees Celsius is
informative. However this aggregation is not so informative to everyone. Pilots need to know wind speed and temperature in order to calculate their take off speed and angles. A pilot would not be happy to be given the wind chill aggregation instead of the separate metrics.

What is important is not what things are aggregated together, but whether they are fit for their purpose. Without a purpose, we can’t evaluate whether an aggregation is informative or not. It seems that we can’t dismiss any aggregation; all we can say is that aggregations are informative relative to a context.

Let’s revisit my simple hairstyle metric (discussed in Section 2.2, page 34).

<table>
<thead>
<tr>
<th>Colour</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>White or Grey</td>
<td></td>
<td>Blonde</td>
<td>Red</td>
<td>Brown</td>
<td>Black</td>
</tr>
<tr>
<td>Length</td>
<td></td>
<td>Bald or shaved</td>
<td>Short</td>
<td>Ear length</td>
<td>Shoulder length</td>
</tr>
<tr>
<td>Texture</td>
<td></td>
<td>Straight</td>
<td>Slight wave</td>
<td>Wavy</td>
<td>Curly</td>
</tr>
<tr>
<td>Thickness</td>
<td>Thin hair</td>
<td>Slightly thin</td>
<td>Medium</td>
<td>Pretty thick</td>
<td>Thick hair</td>
</tr>
</tbody>
</table>

To calculate your hairstyle number, find your score for colour, length, texture and thickness, and simply add them together. I think this aggregation is fundamentally different to the two I’ve just discussed, as there doesn’t seem to be any context in which this metric would be informative. Perhaps there is a distinction to be made here, that there are aggregations for which there is a context that would make them informative, and that there are aggregations for which no such context exists. But that would be a very strong conclusion to draw. I think what we should conclude from this discussion is that an aggregation is context specific. That is, aggregations are informative in certain contexts, for certain purposes. We can’t evaluate an aggregation without considering its purpose.

It is unclear what purpose a relative complexity aggregation would have – that is, over and above being an aggregation. Without a purpose, we can’t evaluate an aggregation, so for now I’ll leave this question aside. There are no aggregation attempts in the literature, and
without seeing any attempts; I am going to remain agnostic on the issue. If aggregating various individual-user-based relative complexity metrics is informative, then I think such a metric could be used to measure structural interpretations of the thesis that all language varieties are equal. However without any evidence that a ‘super’ relative complexity metric will be informative, I don’t see any reason to think that this would be the case.

2.2.2 Conclusion

In this section I argued that the thesis that all language varieties are equal should be understood as making a claim about the equi-complexity of language varieties. If that is true, then we need a way of measuring language complexity. That is, there must be a way to measure language complexity according to which we can compare different language varieties. There are two possible types of language complexity that we could use: absolute complexity and relative complexity.

As I have argued, relative complexity is unsuitable for investigating the thesis that all language varieties are equal. Specifically, relative complexity measures language difficulty as itself relative to different types of user. By definition, relative complexity gives no measure of general complexity. This does not in itself rule relative complexity out; however, given that no one language user (e.g. hearer or listener) is primary to language, relative complexity would unfairly promote one user over another, and would give a skewed view of language complexity.

Absolute complexity, on the other hand, is an approach that looks at language as abstracted away from usage. It avoids the problems that relative complexity faces. By looking at the content of knowing a language variety, we can investigate the complexity of various language varieties and compare them against each other.

Just how to measure absolute complexity is a separate issue. I haven’t argued for a specific metric; I’ve instead claimed that we should take a certain approach when coming up with a metric. I haven’t shown that such a metric exists; there is still a query as to whether it is in fact possible. What I have shown is that such a commensurable metric is necessary in order to posit structural interpretations of the thesis that all language varieties are equal, and that there is a plausible approach to such a metric.

I have argued that it is plausible that there is a metric for measuring language structure by which you can compare different languages, and I will be assuming this in the following chapters. Metric choice is a separate issue, which I’m going to put aside. Clearly there will be
debates about various potential metrics, which linguistic features should be measured, and how they should be weighted.

For those creating metrics to carry out language investigations, there is going to be another issue. There could be more than one metric that fits the desired qualities, and if so the competing metrics will most probably give different results. Which metric we choose will then determine which results we have. While I’m not discussing the problem of metric choice in this thesis, I claim that there is at least one metric that is suitable for making comparisons of language structure across language varieties. If I can make that claim, I can discuss specific structural interpretations that all language varieties are equal.
3. Chapter Three: Setting out the structural interpretations

In chapter one, I set out the interpretations found in the literature of the thesis that all language varieties are equal. While the attempts made so far are valuable, the accounts lack philosophical rigour. I will aim to set out the hypotheses clearly, I will introduce some structural hypotheses that haven’t been formulated in the literature, and will explain the differences between these approaches. I will set out five hypotheses, competing interpretations of the thesis that all language varieties are equal. These five hypotheses fall into three groupings. Initially I will present and explain the hypotheses. In later sections I will analyse them in more detail.

All structural interpretations have two parts to the claim: a structural claim and an explanatory claim. One is a data prediction; the other is a claim about the dynamics that create this data. The structural claim is what the hypothesis claims the data should look like, if all language varieties were measured according to a commensurable metric and the data was plotted. The explanatory claim accounts for the structural claim; it describes the mechanism(s) at work that create the predicted data pattern. For example, if the predicted pattern is that language varieties all share the same value, then the explanatory claim needs to explain why this should be the case. This means there are two ways of arguing for against such interpretations, which I will later discuss.

3.1 Force analogies and threshold analogies

In the following sections I’m going to be setting out different structural hypotheses. As I mentioned, these will also have to provide an explanatory claim. When it comes to discussing these explanatory claims, I’m going to use force analogies and threshold analogies which I’ll explain first.

Force analogies are often used quite prominently in biology. “Selection for X, against Y, and so on, are component forces that combine vectorially to determine the dynamics of the population” (Sober and Lewontin 1982: 160). Sober later claims “[evolutionary] theory may be thought of in analogy with Newtonian mechanics. Various forces are described, but the theory has at its conceptual center a view of what will happen to the systems it describes when no forces at all impinge” (Sober 1984: 14). Forces acting on a ball in flight together account for the ball’s arc. We can also think of evolutionary processes as forces acting on the units of selection. I want to use similar force analogies. There are various forces that act on language varieties which we can use to account for certain features of language.
Linguists also refer to forces, and their forces seem to be directly analogous to biological forces. Aitchison claims that “that language… may be marking time, or treading water, as it were, with the advance or decline held in check by opposing forces” (Aitchison 1991: 7). When talking about language change, Bloomfield refers to prestige as “the most powerful force of all in fluctuation” (Bloomfield 1935: 403). Dahl even explicitly draws the analogy with biology. “It is sometimes suggested that language change is due at least partially to forces that are similar to mutations of genes” (Dahl 2004: 60).

Using force analogies in linguistics seems unproblematic. They are fulfilling the same function here as they are in biology. Therefore just as they are accepted in biology, these metaphors should be accepted in the linguistic field. However ‘forces’ are not the only relevant analogy here. Linguists also talk invoke threshold analogies. McWhorter doesn’t use the word ‘threshold’, he prefers ‘limit’ (McWhorter 2001: 131); however, it seems clear that his limit is the same thing as what I’m calling a threshold.

If we can use force analogies, we should also be able to use threshold analogies. In the following section I will argue that force and threshold analogies are fundamentally different and that maintaining this difference is important. I will show that the analogies are useful for the topic under discussion. Both analogies play an important part in the structural hypotheses I will consider.

I’m claiming that a threshold is a constraint on a force. That is, a threshold is a point at which a force ceases to have an effect. This effectively means that the object upon which the force is acting comes to a halt. I’ll go through an example to explain this concept and to show how the difference between threshold and force is useful.

Consider an example. Jamie is packing a suitcase for a domestic flight. There is a threshold for the suitcase’s weight; the airline won’t accept luggage over 20 kg. But this threshold will only come into Jamie’s consideration if the suitcase is getting close to 20 kg. While packing, if the suitcase weighs 18 kg Jamie might start thinking about the threshold, but if the suitcase is 10 kg, he’s not going to think about the threshold.

That’s not to say that there are no other forces at work to regulate the suitcase’s weight. One force will act to increase the weight: Jamie wants to bring enough supplies for his trip. Another will act to decrease the suitcase’s weight. All things being equal, Jamie would rather a lighter suitcase. He can lift a 20 kg bag, but he’d rather lift an 18 kg bag, and would prefer a 16 kg bag over that. So when considering whether to bring a second pair of shoes, Jamie considers whether the shoes are necessary for his trip, i.e. whether the first force dictates that he should take the shoes. Once he decides that they aren’t necessary, the second
force dictates that the shoes should be left behind. If no other forces are in play, weight should be minimised.

Here’s a diagram of the forces involved in determining the weight of the suitcase. The red line is the threshold; this is the maximum weight the suitcase is allowed to be. The arrows represent the two forces, the downward arrow is Jamie’s desire to have a light suitcase, and the upward arrow is the desire to have enough supplies for his trip.

As the diagram in (1) shows, the threshold didn’t actually have any effect on the weight of the bag. The two forces combined explain the weight of the suitcase.

Suppose that Jamie will be dropped off and picked up. He won’t be carrying the bag, so he no longer has a preference for a light bag. All other things being equal, without the downward acting force (Jamie’s preference for carrying a lighter bag), the weight of the suitcase will rise unchecked until it hits the threshold. Now the threshold has an effect on the suitcase’s weight.
Now I can explain why it is useful to differentiate between forces and thresholds. Let’s look back at the initial scenario, represented in chart (1). However let’s suppose there is no distinction between ‘threshold’ and ‘force’. Chart (3) is an illustration of this concept.
Though it may seem that chart (1) and (3) represent the same information, chart (3) is less informative than chart (1). By losing the distinction between threshold and force, we’ve lost information.

What (1) makes clear, but (3) does not is that the maximum weight has no effect in this instance. What (1) also contains is counterfactual information. The diagram tells us what would happen if the force ‘the less weight the better’ were removed, or counteracted. The information in chart (2) is contained within (1). Not only does chart (3) not tell us that the maximum weight has no effect in this instance, it also doesn’t contain counterfactual information. It isn’t clear what would happen if the ‘the less weight the better’ force were removed from the system. Counterfactual information is important. If you’re wondering about what would happen if the downward force were removed, you’re not going to get the right result if the threshold isn’t represented in some way. A representation that lacks this information is not as good as a representation that contains this information. I think it is useful to maintain a distinction between ‘force’ and ‘threshold’. For the rest of the thesis, when I mention ‘threshold’ I’m using the definition I mentioned earlier; a threshold is a constraint on a force.

3.2 The hypotheses

Now that I have discussed the background assumptions we need to investigate the thesis, I can start setting out interpretations of the thesis that all language varieties are equal. In the following section I will set out the five hypotheses I have found in the literature and then I will evaluate them.
3.2.1 The static value hypothesis:

This hypothesis states that when you measure all language varieties with a commensurable metric, all language varieties will have the same value. Whatever score English has according to this rating; it will be the same score as Maori, Latin and Indonesian. According to this hypothesis, all language varieties are equal insofar as they share this value. The reason that all language varieties have the same value is because language structure is in an equilibrium.

As I have explained, a hypothesis must posit both a structural and an explanatory claim. According to the static value hypothesis, the explanatory claim is that there are at least two forces which combine to keep complexity scores static according to the commensurability metric. These balance each other out so well (and in every case) that language varieties measured according to the commensurable metric turn out to all share the same value. This hypothesis predicts a pattern similar to chemical equilibrium reactions, hence the name.

Note that this hypothesis doesn’t stipulate what these forces are. There are a number of competing possibilities, and proponents of this hypothesis will have to put forward a particular description of the dynamic.

3.2.2 The varied value hypotheses:

Now I’m going to discuss some other hypotheses that don’t make this claim; instead they propose a variety of language values according to a commensurability metric. In this section
I’ll introduce and set out four hypotheses; the weak maximum hypothesis, the weak minimum hypothesis, the weak equilibrium hypothesis and the double threshold hypothesis. These hypotheses claim that language varieties measured according to the commensurable metric will have differing values. The weak maximum and weak minimum make no claim about the distribution of language value according to the metric, whereas the weak equilibrium and double threshold do. The weak equilibrium and double threshold both posit a relatively constrained set of values.

The weak maximum, weak minimum and double threshold hypotheses all posit a variety of language values and one or two thresholds that constrain these values. We have to be careful that these thresholds are not trivially confirmed. If it is the case that language values do vary, given that there are a finite number of language varieties, it is trivially true that there will be a language variety with the lowest value. If there is a lowest value, it is trivially true to say that there are no values below it. Similarly, if language values vary, there will trivially be a highest value. One of the consequences of having a minimum threshold is that there will be no language varieties with values below that threshold. However the setting of a lowest or highest value does not prove the existence of a threshold. As such, the burden is on threshold hypotheses to justify their claim that a threshold exists.

3.2.2.1 The weak maximum hypothesis
This hypothesis posits a different structural claim from the first three. The proposed pattern for the weak maximum hypothesis is that different language varieties will have a range of different scores when measured according to the commensurability metric.
I have included two diagrams to show that the pattern this hypothesis predicts is not as rigid a pattern as some of the others. Both of these diagrams are compatible with the weak maximum hypothesis. It could be the case that only one language variety has hit the maximum threshold (i.e. diagram (a)) or that many language varieties have hit this threshold (i.e. diagram (b)). Common to both diagrams, and central to the hypothesis, is the claim that there is a maximum threshold for language varieties, and that at least one language variety has reached this threshold.

While there will be a variety of scores, according to McWhorter there will be at least one language variety with the highest score. This is not a necessary condition of the theory, but it is of McWhorter’s formulation. He infers the existence of the threshold from a phenomenon he claims exists, namely, that most (if not all) language varieties have reached this threshold. As such, he needs at least one language variety to have achieved maximum complexity.

The weak maximum hypothesis claims that there is a maximum threshold for language structure. Like a lid on a jar, there comes a point where you cannot fit anything more in. There is simply not enough space, and this imposes an upper limit. If you are storing rice in the jar, in order to fit more in, you must first remove some rice, to make space for the incoming grain.

Initially it seems unclear what the grains of rice are analogous to in language. We can approach this in two ways. We could look at what the commensurable metric is counting or we could see what could account for a threshold effect in or on language.
As stated, the metric is going to be measuring absolute complexity in some formulation. There seems to be no inherent upper limit for linguistic complexity, rather a threshold would be imposed for non-linguistic reasons. McWhorter, as I mentioned in chapter one, runs a version of a maximum hypothesis.

After countless millennia of usage and drift, we might expect that by a certain point all grammars had, by the sheer dictates of chance, developed various random complexities in parts of their grammars… We might propose that the volume of such excrescence in each grammar eventually reached the limit of human propensity to process it… Under this scenario all natural language varieties would be equally complex by virtue of having all come to rest at a certain ‘surplus complexity quotient’.

(McWhorter 2001: 131)

McWhorter thinks that the threshold comes about because of a relation between linguistic complexity and human processing power. Namely that the greater the complexity, the more processing power required. Moreover he claims that there is a limit to human linguistic processing ability and this limits linguistic complexity accordingly. The grains of rice are analogous to humans’ linguistic processing ability.

It seems fair to assume that linguistic knowledge is somewhat like working memory; linguistic knowledge, like memory decays over time; if a word or structure is not used or heard, knowledge of that word or structure will decline, and will no longer be part of that speaker’s idiolect. Moreover, language depends on individuals’ idiolects; if a particular word or phrase is not part of any speaker’s idiolect, it will no longer be part of the language variety.

A system like McWhorter’s is internally consistent. If language and linguistic processing are as McWhorter describes then we should expect to see a maximum threshold on language. However human linguistic processing is not as McWhorter describes it. In order to see this, I’ll make explicit some of the assumptions he needs to make. In order for a mental processing limit to impact language varieties as McWhorter claims, we have to assume that:

- Higher linguistic complexity causes increased human language processing.
- There is a limit as to how much linguistic processing an individual can do.
- There is an average amount of linguistic processing a human can undertake, and that this is fairly standard throughout the population.
- For a given language variety, A, most speakers of this language variety have reached their linguistic processing limits.
There is an increase in complexity in the language. If these conditions obtain, we have a system stretched to the limit, and the increase in language complexity puts more pressure on the system. Something has to give, and in a system like this, it’s going to be overall language complexity.

A language variety supervenes on a group of idiolects. If a particular word or grammatical structure is in no idiolect, that word or structure is not in the language variety. Suppose for the sake of simplicity that there is a language variety that has 200 speakers: there are 200 idiolects the language variety supervenes on. In order for the speakers to communicate with each other, they must share basic vocabulary and grammar. That is, the majority of each person’s idiolect will be identical to others’ idiolects. Without such a large amount of shared linguistic knowledge, communication would be very difficult. The requirement that idiolects of the same language variety are significantly similar means that if some idiolects undergo significant language change, other idiolects must also undergo such change, otherwise they will no longer be of the same language variety.

Within an individual’s idiolect, there will be words and structures they use frequently, those they hear frequently, those they use rarely and those they hear rarely. If most speakers frequently use a word, that word will also be heard frequently. If most speakers rarely use a word, it is likely that that word will be rarely heard. This relationship is not nearly as strong as the former, however. There could be a few dedicated individuals who promote the word, in which case it would be more frequently heard. While acknowledging that the correlation, if there is one, will not be very strong, it nevertheless seems reasonable to assume that there is a weak correlation. It’s the linguistic phenomena that are both rarely used and rarely heard that will play the crucial role here.

If, as McWhorter claims, humans have a maximum linguistic processing capability that most have reached, in order for an individual’s idiolect to expand, it must contract elsewhere. For example, when you learn a new word, you lose a different word from your vocabulary. It seems likely that the words or grammatical structures that an individual loses will be words or structures that an individual rarely uses and rarely hears. We forget arcane structures, words we used to use as children, and specialised terms that are no longer relevant to our lives.

So suppose that there is a new innovation in the language variety, one that constitutes an increase in complexity, and this innovation—word, structure, accent change etc.—has prestige and spreads like wildfire. In order for an individual to adopt this innovation, they must cut a word or structure out of their idiolect. The linguistic knowledge that will be lost is
going to be knowledge that the individual hasn’t thought about for a while, phenomena they rarely use or hear.

If, as I think likely, there is a correlation between rare usage and rare hearing, then there will be a correlation between the words or phrases that idiolects lose. If what is rarely accessed linguistic knowledge (either from use or from hearing) for one person is likely to be rare for the other speakers, it’s likely that the linguistic knowledge that is dropped from their idiolects will be similar.² If this is happening to all the individuals in a speech community, then not only will individuals not produce the word/structure, but they will not hear it. As it is not reinforced, it will drop out of the idiolects, and in doing so, the language variety. As language varieties supervene on idiolects, if a word or phrase is part of no idiolects it is no longer part of the language variety. Even though the mechanism that McWhorter is proposing doesn’t work directly on language, a threshold effect can filter through from individuals’ idiolects to language.

It is important to also note that McWhorter requires that most individuals in a speech community have reached this linguistic processing capacity. In order for such a threshold to have an effect on language, it must be being consistently reached. There will be some variation on the processing capacity across individuals, as there is variation in height, running ability etc. Accordingly, not every individual has to have their linguistic processing capacity reached, but McWhorter does need to claim that most have. If most individuals in a speech community have extra linguistic capacity, there won’t be a significant effect on the language variety. In order to explain the proposed pattern, the strict maximum hypothesis needs to posit that language varieties are complex enough and the threshold sufficiently low enough that the two interact.

Both the (a) and (b) weak maximum hypothesis diagrams are consistent with the hypothesis; however the patterns differ in how much support they give to the threshold. Diagram (a) is consistent with the weak maximum hypothesis, but the pattern is also consistent with other hypotheses. Diagram (b) on the other hand, provides more evidence of a threshold, as we can see the threshold’s effects on multiple language varieties.

Moreover, McWhorter, who explicitly presents the weak maximum hypothesis, proposes the second diagram. He thinks that most language varieties are at the threshold, except creoles which are too young; “creoles display less complexity than the rest of the world’s natural grammars” (McWhorter 2001: 133).

In the diagrams of the weak maximum hypothesis, there are no proposed forces pushing language varieties up or down, but there is a threshold (constraint on a force). This

² I am merely claiming there is a correlation, I make no assumptions as to the strength of the correlation.
doesn’t mean that there are no any such forces at work; indeed a number of forces are compatible with this theory. However the theory doesn’t need to posit any forces.

Unlike the strict maximum hypothesis, the weak maximum hypothesis doesn’t posit another process at work (though there are a number that are compatible). So how does the process explain the pattern? The weak maximum hypothesis posits a maximum threshold for language, an upper limit, but no lower limit. As such we can expect a variety of values according to the hypothesised commensurable metric.

3.2.2.2 The weak minimum hypothesis
This hypothesis is similar to the strict minimum hypothesis (2). It posits that when you measure all language varieties with the hypothesised commensurability metric, all language varieties are above a particular value, which is a lower bound. It is where a language variety stands with respect to the threshold that is important – everything above the threshold is ‘equal’.

![Weak minimum hypothesis](image)

This lower bound marks an important divide between the usability of communication systems. Systems above this line are usable language varieties, whereas systems below the line lack some specific structure such that the usability of such systems is noticeably diminished. This should map on to the boundary between proto-language and language. That is, you would not expect proto-languages to have features above this lower boundary. While there is no list in the literature of what exactly this lower boundary could be, there are some examples of the types of things that would distinguish between communication systems above
and below this boundary. For example, that language varieties above the boundary have a lexicon, are able to refer to the past, present and future; are capable of recursion (embedding phrases within sentences), have indexicals (e.g. ‘I’, ‘that’, ‘she’).

Gil is a proponent of the weak minimum hypothesis. He thinks that there is a minimum threshold for language, and that is what’s important when deciding the ‘equality’ of languages; anything above that is irrelevant to the topic. Though Gil thinks the minimum threshold is due to functional reasons, he explicitly refers to language structure; specifically, minimum grammar levels. “How much grammar does it take to do all of the things that… only humans are capable of doing?” (Gil 2009: 20) He doesn’t explain where exactly he thinks the minimum boundary is; rather in his chapter he describes some characteristics of language that he thinks are definitely above the threshold.

3.2.2.3 The weak equilibrium hypothesis

This hypothesis states that when you measure all language varieties with the hypothesised commensurable metric, all language varieties are clustered around the same value, because language structure is an equilibrium. This view is similar to the static value hypothesis (see 3.2.1, p. 50). It posits similar processes/forces acting to push language varieties towards the ‘normal’ from both above and below. This view, however, proposes a slower return to equilibrium than the static value hypothesis. That is, over time language varieties that are above the ‘normal’ should move down, and language varieties that are below the ‘normal’ should move up.

![Weak equilibrium hypothesis](image-url)
Aitchison puts forward a version of this hypothesis. “We must conclude therefore that language is ebbing and flowing like the tide, but neither progressing nor decaying, as far as we can tell” (Aitchison 1991: 215). She acknowledges that language varieties do not all have equal complexity, but claims nevertheless that there seems to be certain optimum complexity levels for language. “Language seems to have a remarkable tendency to restore its patterns and maintain its equilibrium” (137).

### 3.2.2.4 Double threshold hypothesis

When you measure all language varieties with the hypothesised commensurable metric, all language varieties are clustered between two bounds. This proposes both a maximum threshold and a minimum threshold.

At first glance, the double threshold and weak equilibrium seem to be merely different versions of the same hypothesis. However this is not the case. The double threshold hypothesis differentiates between small, ‘everyday’ forces (e.g. pronunciation changes, new vocabulary) and large ones (e.g. loss of cases). Clearly, language varieties are constantly changing, and both hypotheses acknowledge that. However most of these changes are small, and don’t make a significant difference to the language variety.

The weak equilibrium hypothesis doesn’t differentiate between small and significant language change, whereas the double threshold hypothesis can account for this difference. Here are some explanations of the types of language change I am referring to.
One example of small, seemingly not significant language change is slang: whatever slang 10 year old children were using two years ago will be different to the slang today. The current day slang is not a significant language change from that of yesteryear: adults stereotypically mock the current slang, however they do understand it. Slang is a great in-group identifier.

When a child uses a new slang term, someone who doesn’t know the term will still (usually) understand the utterance. The meaning gets across. I argue that slang is a minor language change, because new slang terms don’t impede understanding. That is, even if you don’t understand the particular term being used, you can understand the utterance.

There are language changes that do seem to be significant, for example the loss of cases from English, the creation of third person plural pronouns in various English sub-groups (e.g. ‘yous’ as heard in utterances such as ‘yous guys’). These changes seem to be different from the former group, not merely of scale, but of type.

The double threshold hypothesis accommodates both of the discussed types of language change processes. All language varieties must be within the threshold boundaries. However we will only see a threshold actively impacting a language variety when the variety is being pushed towards a threshold by a force.

There is lots of language change that does not constitute a significant change, in that it does not have a large impact on a language variety’s score according to an appropriate metric. This minor language change (such as the invention of a new slang term) is not necessarily going to be related to a threshold. The double threshold hypothesis can differentiate between ‘everyday’ forces which cause minor change, and thresholds which are involved in significant change, for example the creation of a new tense. The weak equilibrium hypothesis, on the other hand, makes no differentiation between these processes; it posits the same processes to explain both phenomena.

It is an empirical question which of these hypotheses describes the phenomenon of language change correctly. As such, we can find out which of these two hypotheses (the weak equilibrium hypothesis and the double threshold hypothesis) actually describes language change.

### 3.3 Evaluating the hypotheses

I have just set out the five structural interpretations of the thesis that all language varieties are equal (the static value hypothesis, weak maximum hypothesis, weak minimum hypothesis, weak equilibrium hypothesis and the double threshold hypothesis). In the following section I will return to each hypothesis and evaluate them.
3.3.1 The strict equilibrium hypothesis:

If we measure and find at least two language varieties with differing values according to the hypothesised metric, this is evidence against the strict equilibrium hypothesis. As I will explain, there is such evidence, so this hypothesis is false.

Dahl analyses Elfdalian and Swedish, language varieties which share a common ancestor, but are mutually incomprehensible. “Both Elfdalian and Swedish belong to the North Germanic languages. Elfdalian... is sufficiently far from Swedish not to be mutually comprehensible with it. Elfdalian and Swedish differ in many respects... [but] are still of course quite closely related” (Dahl 2009: 52). The aim of his research was overtly to test the strict equilibrium hypothesis, which he refers to as “the complexity invariance assumption” (50). “The idea is that by comparing such a pair of languages, it will be possible to identify the points at which the language varieties differ in complexity and see whether these differences are compensated for elsewhere in the grammar” (50). Dahl compares the two language varieties by surveying phonology, morphology and syntax, and claims that Elfdalian has a more complex phonology and morphology than Swedish, while he concludes that “a difference in syntactic complexity between Elfdalian and Swedish is not easily demonstrable” (63).

According to the strict equilibrium hypothesis, any increase in complexity in one domain must be compensated for by a decrease in complexity in another domain. Elfdalian and Swedish share a recent common ancestor. Any difference between the two language varieties is due to language change from the common ancestor language variety to now. So if one of the static value hypotheses is true, for every complexity Elfdalian has that Swedish doesn’t, we should see a complexity that Swedish has that Elfdalian doesn’t. Dahl concludes that this compensatory complexity is not found, that Elfdalian is simply more complex than Swedish. From this, we can conclude that the static value hypotheses are false.

Dahl’s conclusion (that Elfdalian is more complex than Swedish) is only evidence against the strict equilibrium hypothesis if the method Dahl used to compare the two fits our requirements for the hypothesised commensurability metric. I think that Dahl’s survey does fit these requirements. Choosing closely-related language varieties restricts the variables – instead of having to take inventory of every part of the language variety, you can focus on the differences. Then it’s a straight one to one comparison. Dahl’s project shows that there isn’t a compensatory mechanism to maintain language complexity value, i.e. that the static value hypotheses are false.
Nichols also concludes that this hypothesis is false. Explicitly testing Hockett’s total grammatical complexity hypothesis, (a formulation of the strict equilibrium hypothesis) Nichols carries out “a large enough cross-linguistic survey of enough parts or grammar to indicate whether the assumption of equal complexity appears viable” (Nichols 2009: 111). She compares a number of elements of grammatical complexity across 68 language varieties in a pretty comprehensive survey. The aim of the survey is to cover enough language varieties to test the waters, as it were. Nichols doesn’t think that her survey is large enough to give a definitive answer as to whether language varieties are equal as Hockett claims, however she expects it to point towards an answer. After looking at specific ‘grammatical complexity’ features, Nichols concludes that total grammatical complexity is false. She states that “if there were a preferred complexity level or trade-off between complexity levels of different grammar components, one could expect it to have at least made itself felt in this survey; but it has not done so” (121).

Assuming that Dahl and Nichols’ approaches are correct, their evidence shows that the strict equilibrium hypothesis is false.

3.3.2 The weak maximum hypothesis

McWhorter gives a version of the weak maximum hypothesis. He thinks that most language varieties have reached the threshold, but that there are some (in particular, creoles) that are less complex than the rest. McWhorter explains that the existence of the threshold is due to limitations on human processing ability.

We might propose that the volume of such excrescence [complexity] in each grammar eventually reached the limit of human propensity to process it… Under this scenario all natural language varieties would be equally complex by virtue of having all come to rest at a certain ’surplus complexity quotient’.

(McWhorter 2001: 131)

Kusters takes the claim at face value, and tries to defend McWhorter’s threshold:

Let’s try to defend it [the maximum threshold] by supposing that it is due to limitations on mental capacities: the mind/brain can process and contain only a limited amount of complexity. An increase in complexity in one component, then, necessarily
leads to a decrease elsewhere. This argument presupposes that the average speaker already uses the maximal amount of the ‘complexity space’.

(Kusters 2008: 11)

What McWhorter calls ‘the limit of human propensity to process [language]’ and Kusters identifies as ‘complexity space’ is not general intelligence, or general brain processing power. This has to refer to a specific complexity space in the brain for language. That is, there is some specialised part of the brain that deals with all and only language. Or at least there is some specialised part of the brain that deals with all and only language complexity of the type that the hypothesised commensurable metric measures. The reason for this strict regulation is that if ‘complexity space’ is related to general processing, and so its capacity can change based on how much processing or memory is being used, the maximum threshold concept falls apart.

Let’s look at an analogy. Suppose you have a bookshelf with a particular section for recipe books. Once you fill that section up, you can’t store any more recipe books, even though you may have space on other parts of the bookshelf. This imposes a maximum threshold on recipe books; there is only so much space allocated. Once you reach this maximum, the only way of storing new recipe books is to first get rid of some current recipe books. How many new recipe books you can store depends on how many recipe books you already have. This is how McWhorter conceives of human ability to process language: there is some fenced off space such that it imposes a maximum threshold. The bookshelf is analogous to the brain’s processing power, recipe books to language.

I mentioned that without this designated space the concept of a maximum threshold falls apart. Suppose then that you are no longer limited by designated sections in your bookshelf. You can fill almost the entire thing with recipe books if you so wish, but only almost the whole bookshelf. After all, you have to store some good literature, trash novels, and textbooks there. In this scenario, how many recipe books you can store depends on how many other books you are storing. When you have lots of other books, you can only have a few recipe books, and when you have few other books, you can store lots of recipe books. So if you want to store lots of recipe books, you can choose to have fewer other books. However as you always have to store some other books (literature, trashy novels, textbooks) you will never fill up the entire bookshelf with recipe books. Thus, in this scenario, there is no maximum threshold for recipe books; how many you are able to store simply depends on how many other books you have. This scenario is analogous to the situation where language does not have a designated section of brain processing power. It is impossible for us to use all of
our processing power for language even if we wanted to: we use parts of our brain for non-language things, for example spatial awareness, memory, and recognition. So even if language can simply colonise any non-active parts of the brain, it won’t hit a constant threshold: how much language can take over simply depends on what other things the brain is doing at that point in time. So as McWhorter wants to posit a maximum threshold that doesn’t depend on non-language related phenomena, he has to claim that there is a specific language-related part of the brain. Remember too that McWhorter thinks that most language varieties have hit this maximum threshold.

As Kusters points out, learning a second language variety, bilingualism and multilingualism pose a serious problem for the maximum hypothesis.

When... arguing that mono-linguals use the complete ‘complexity space’, the acquisition of foreign languages becomes difficult to explain. It would entail either that the acquisition of a second language has the effect that the native language becomes less complex, or that the acquisition of a second language occupies a quite different ‘complexity space’.

(Kusters 2008: 10)

Here we have a dilemma. In order for the maximum threshold to affect language, at least some people with some language varieties must have achieved complete ‘complexity space’. Only once this complexity space has been filled up will we start seeing trade-offs as a necessary part of language change. However it is unclear how someone using complete ‘complexity space’ for one language variety can learn another. But this happens all the time. Millions of people are bilingual, lots are multilingual, there are bilingual communities, a significant number engage in learning a second language variety.

As Kusters points out, there are two moves that McWhorter can make here, although neither looks particularly appealing. The first would be to claim that learning a second language variety involved a decrease in your first language capability. Clearly this is absurd. Sure, it may be the case that you lost some complexity in your first language variety, but not nearly enough to compensate for the gain in complexity in your second language variety.

Perhaps the language complexity space is used to store knowledge of universal grammar. As universal grammar, by definition, is used by every language variety, you only need to store one copy of this. When you learn a second language variety, you learn different vocabulary, however you only need to learn superficial grammar rules, your knowledge of
universal grammar is almost all the complexity you need. Of course this move is only possible if universal grammar exists, which is still hotly debated.

This may seem a promising option; however it is not one that McWhorter can use here. He claims that not all language varieties have reached the maximum threshold, that some are simply less complex than others. If complexity is how much universal grammar one has, then following McWhorter, some language varieties (e.g. creoles) fall short of universal grammar. As such, the situation above doesn’t describe these language varieties. This approach fails if the theory of universal grammar is false.

A second option available to McWhorter is to claim that different language varieties occupy different complexity spaces. As such, learning a new language variety has no affect on the first, as it is subject to a completely different maximum. This would seem bizarre. There is no reason to think that each different language variety occupies a separate space in an individual’s brain. Indeed there are bilingual communities where different language varieties are used interchangeably. It seems that such fluid language mixing would not be easy if an individual’s language varieties are completely separate. When you’re learning a second language variety, you are typically taught in your first language; the second language variety’s vocabulary, grammar, syntax and other rules are stated in the first language variety.

**3.3.3 The weak minimum hypothesis**

In the previous section I pointed out we can’t simply go from there being evidence that there is one language variety with a ‘lowest’ value according to the commensurable metric to the existence of a minimum threshold. We don’t want the threshold to be trivially confirmed. So a proponent of the hypothesis has to do some work to convince us that the threshold exists.

In section 3.3.2 I set out Gil’s weak minimum hypothesis. He thinks that the minimum threshold is a function of specific language structures; only with these structures can a language variety enable humans to carry out complicated tasks. “How much grammar does it take to do all of the things that, amongst all living species, only humans are capable of doing, such as, for example, worshipping God, waging war, and working in offices; inventing, manufacturing, and using sophisticated hi-tech tools; and engaging in multifarious commercial, scientific, and artistic abilities?” (Gil 2009: 20). The idea is that if a language variety is under this threshold, it will not enable humans to do these tasks.

So, what is the proposed minimum and how does it correlate with task enabling? Gil claims that this is something like an Isolating-Monocategorical-Associational Language (IMA). An IMA is “an idealized language prototype with the following three properties:”
1. morphologically isolating: no word-internal morphological structure;
2. syntactically monocategorical; no distinct syntactic categories;
3. semantically associational: no distinct construction-specific rules of semantic interpretation

(Gil 2009: 20)

In the following sections I will set out and explain each of these three claims.

3.3.3.1 Morphologically isolating

A language that is morphologically isolating is a language where there are no suffixes, prefixes, or other such affixes. In such a language, every morpheme (atomic unit of meaning) is a separate word. English is not morphologically isolating; it has lots of words that are made up of multiple morphemes. Think of the word *antidisestablishmentarianism*, “opposition to the disestablishment of the Church of England”. Looking at the table overleaf, this single word is made of six morphemes, each applied in turn to give a new meaning. The definitions in the table are taken from the online Oxford English Dictionary, June 2013.³

<table>
<thead>
<tr>
<th>Establish</th>
<th>From 16th c. often used with reference to ecclesiastical ceremonies or organization, and to the recognized national church or its religion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dis-establish</td>
<td>To deprive (a church) of especial State connection and support; to remove from the position of being the national or state church</td>
</tr>
<tr>
<td>Disestablish-ment</td>
<td>The withdrawal of especial State patronage and control from a church</td>
</tr>
<tr>
<td>Disestablishment-arian</td>
<td>An adherent of disestablishment</td>
</tr>
<tr>
<td>Anti-disestablishmentarian</td>
<td>An opponent of disestablishment</td>
</tr>
<tr>
<td>Antidisestablishmentarian-ism</td>
<td>Opposition to the disestablishment of the Church of England</td>
</tr>
</tbody>
</table>

In a morphologically isolating language, such prefixes and suffixes cannot be added, that is to say there are no means by which you can change the meaning of one word by way of altering it. In order to add information, you simply add words.

### 3.3.3.2 Syntactically monocategorical

A language that is syntactically monocategorical makes no grammatical distinctions between different types of words. In English, for example, there are many grammatical groupings (such as nouns, verbs, adverbs, etc.). These groups determine, for example, word order, which words can modify others, and whether a string of words is a sentence. In a syntactically monocategorical language, there are no such categories and thus no such rules based on these categories.

### 3.3.3.3 Semantically associational

A semantically associational phrase is one where the interpretation of that phrase does not depend on any grammatical rules. That is, interpretation depends purely on the morphemes. In English, ‘dog bites man’ and ‘man bites dog’ have different meanings, even though they contain the same words. The difference in meaning comes from English word order rules with respect to subject and object. A semantically associational language is one where word order, or any other grammatical rule, does not impact the meaning of the sentence. A sentence can then have many different meanings, and context will decide between them.
Gil claims that “one example of a Relative IMA [a language approaching IMA categorisation] is Riau Indonesian” (Gil 2009: 22). He argues that Riau Indonesian is only slightly more complex than an IMA language. It is not strictly speaking morphologically isolating, though most words are monomorphemic. Riau Indonesian has some syntactic categories, but not many, and it is fairly semantically associational.

He claims that as Riau Indonesian is sufficient for speakers to accomplish normal human tasks, that an IMA language variety is more complex than the minimum standard. However, even if we accept Gil’s claim that Riau Indonesian is a Relative IMA, this doesn’t actually explain where the minimum threshold is.

Gil’s argument fails when it comes to the equi-complexity thesis. Here’s a formulation of his claim in argument form:

1. Languages differ in complexity
2. Simpler languages achieve everything that more complex languages do
3. There is a minimum threshold
4. All languages are over this threshold

Therefore
5. All languages are equal

Gil’s argument that all languages are equal relies on a premise that states that languages differ in complexity. He claims that there is a minimum complexity threshold, and that insofar as languages have surpassed this threshold, they are equal. There is no inherent problem with this approach. Gil is talking about two different complexity comparisons, and there could be a reason why one is more significant than the other. However in order claim that one range of complexity variance is more significant than another Gil needs to argue for the difference in significance. However he offers no justification for his claim that some complexity variance is significant, while other variance isn’t.

Moreover, Gil’s evidence for a weak minimum threshold is not conclusive. As I mentioned earlier, when it comes to thresholds, we need to make sure that they are not trivially confirmed to be true. Having a lowest or highest rating language is not enough evidence to show that there is a threshold, and Gil offers no evidence beyond this.

Due to these two issues, the weak minimum hypothesis fails. Gil’s evidence for a minimum threshold is unconvincing, and more to the point, even if this threshold exists, it is
far from clear that this would be an acceptable conclusion.

### 3.3.4 The weak equilibrium hypothesis

Aitchison is a proponent of the weak equilibrium hypothesis. Aitchison explicitly compares contemporary language varieties with historical language varieties, and different contemporary language varieties with each other. Assuming that we can make such comparisons, Aitchison assumes that language varieties are in one of three states. She looks at three possibilities: that language is decaying, evolving to a fitter state, or “that language remains in a substantially similar state from the point of view of progress or decay. It may be marking time, or treading water, as it were, with the advance or decline held in check by opposing forces” (Aitchison 1991: 7).

Aitchison thinks that language varieties fluctuate around a particular level. “Language seems to have a remarkable tendency to restore its patterns and maintain its equilibrium” (137). “We must conclude therefore that language is ebbing and flowing like the tide, but neither progressing nor decaying, as far as we can tell” (215).

What evidence does she have for this? Aitchison doesn’t claim to have exhaustively surveyed the data, but she does think that there are a lot of data that seem to show compensatory behaviour over periods of time. Here’s a more detailed account of Aitchison’s Maori example I presented earlier (see 1.2.1, page 7).

Once, Maori words ended in consonants, and the passive was formed by adding [-ia] on the end. For example:

<table>
<thead>
<tr>
<th>Active</th>
<th>Passive</th>
</tr>
</thead>
<tbody>
<tr>
<td>awhit 'to embrace'</td>
<td>awhit-ia 'to be embraced'</td>
</tr>
<tr>
<td>hopuk ‘to catch’</td>
<td>hopuk-ia ‘to be caught’</td>
</tr>
<tr>
<td>maur ‘to carry’</td>
<td>maur-ia ‘to be carried’</td>
</tr>
<tr>
<td>weroh ‘to stab’</td>
<td>weroh-ia ‘to be stabbed’</td>
</tr>
</tbody>
</table>

Then consonants at the end of words were gradually lost, so awhi, hopu, mau and wero became the standard active form of the verbs listed above [while] the passive forms have not changed at all... Maori speakers have reanalysed the passive... assuming that the consonant on the end of the verb is part of the passive:”
<table>
<thead>
<tr>
<th>Original construction</th>
<th>New analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>awhit + ia</td>
<td>awhi + tia</td>
</tr>
<tr>
<td>hopuk + ia</td>
<td>hopu + kia</td>
</tr>
<tr>
<td>maur + ia</td>
<td>mau + ria</td>
</tr>
<tr>
<td>weroh + ia</td>
<td>wero + hia</td>
</tr>
</tbody>
</table>

As we can see, a change in one part of the language variety (losing the last consonant) correlates with a different change (complexification of the passive rule). Arguably the loss of the final consonant in the infinitive is a phonological simplification, the resulting word has fewer component parts and is shorter. Aitchison also presents examples from other languages.

One researcher… compared the progress of Turkish and Yugoslav children as they acquired their respective languages. Turkish children find it exceptionally easy to learn the inflections of their language, which are remarkably straightforward, and they master the entire system by the age of two. But the youngsters struggle with relative clauses (the equivalent of English clauses beginning with who, which, that) until around the age of five. Yugoslav children, on the other hand, have great problems with the inflectional system of Serbo-Croatian, which is ‘a classic Indo-European synthetic muddle’, and they are not competent at manipulating it until around the age of five. Yet they have no problems with Serbo-Croatian relative clauses, which they can normally cope with by the age of two.

The Turkish children have inflections down by age two, but don’t master relative clauses until the age of five. The Yugoslav children can use relative clauses at age two, however not the inflectional system until age five. Aitchison suggests that the reason for the different rates of domain acquisition is that there is a difference in complexity in the domains in both language varieties. The Turkish inflectional system is simpler than the Yugoslavian inflectional system, and because of this children find it easier to learn the Turkish rather than the Yugoslavian inflectional system. Aitchison then claims that, given the respective inflectional systems, we would predict that another domain would have opposing complexities.

So, as Aitchison points out, there seems to be a trading relationship here between the inflectional system and relative clauses.
A language which is simple and regular in one respect is likely to be complex and confusing in others. There seems to be a trading relationship between the different parts of the grammar which we do not fully understand.

(214)

The data currently available can’t rule out the weak equilibrium hypothesis. We have evidence that different language varieties differ in complexity. However we don’t have data on how wide this complexity variation is; that is, whether the difference between various language varieties is large in relation to the possible complexity space. Also, there isn’t enough longitudinal data to test the hypothesis’ predictions of language change over time.

There seems to be tentative evidence for something like the weak equilibrium hypothesis; however there is not enough data to rule one way or the other. We simply need more research. In the final chapter I will discuss this further.

3.3.5 The double threshold hypothesis

As I mentioned when introducing them, the weak equilibrium and double threshold hypotheses are very similar. The difference between them is that the weak equilibrium hypothesis only posits forces, while the double threshold hypothesis posits both thresholds and forces that affect language structure.

Deutscher is an expert on Akkadian, an ancient language variety. He examined surviving written texts from 2500 BC to 500 BC, and concluded that Akkadian became more complex over time. Deutscher’s evidence contradicts the static value hypothesis, strict maximum hypothesis and strict equilibrium hypothesis. It seems to argue against the weak equilibrium hypothesis, depending on the time scale being considered.

As with the weak equilibrium hypothesis, we don’t yet have the data to test the hypothesis thoroughly. Both the double threshold and the weak equilibrium hypotheses make similar predictions, in that they involve patterns of language change over time. But they differ in the particular movement of language change over time they predict. As such, we need longitudinal data to tell whether either of these hypotheses (if at all) are correct. In chapter five I will elaborate the differences between these two hypotheses, and explain the research needed to test them.
3.4 Conclusion
In this chapter I have set out and discussed the five structural interpretations of the thesis that all language varieties are equally complex. I discussed the static value hypothesis, the weak maximum hypothesis, the weak minimum hypothesis, the weak equilibrium hypothesis and the double threshold hypothesis.

As I have shown, the static value hypothesis is false. Though researchers disagree about which metric should be used, there is consensus in the data that this hypothesis has been disproved. The weak maximum and weak minimum hypotheses are also false.

The weak maximum hypothesis relies on a claim that has been proved to be false: that there is a limited amount of language processing power in the brain. While it is true that humans don’t have infinite language processing capacities, the three restrictions McWhorter requires are simply not found. McWhorter’s posited language processing limit must be uniform across individuals, all individuals must have a similar bound in order for that bound to affect language. Language processing must be modular, it has to be language-specific in two ways; language processing must be restricted to this module, and it can’t be co-opted by non-linguistic processing. Moreover the majority of language users must have reached this processing limit. None of these three claims seem true. There is evidence that language ability varies widely between individuals (Chipere 2009). The neuroplasticity of the brain is well-documented, particularly with respect to language (see Cowie 2010: Section 3.2); it is simply false that neurological processing is limited in the way McWhorter claims. Furthermore the millions of multilingual and bilingual speakers show that even if there is such a limit, the majority of individuals have not reached it.

The weak minimum hypothesis relies on incomplete argumentation. Gil’s claim that languages are equal relies on a premise that languages differ in complexity. While this is not prima facie contradictory, the onus is on Gil to argue why one complexity variance is substantive while another is not. Gil does not offer any such argument, nor does he argue for the existence of a minimum threshold; he assumes that a description of the posited threshold is enough. However, it is trivially true that for any varied set of values, there will be one (or more) with the lowest value. We can’t infer the existence of a threshold from the existence of a lowest (or highest) value, but Gil doesn’t give us any further arguments for the minimum threshold.

This leaves us with two potential live options, the weak equilibrium hypothesis and the double threshold hypothesis. In a later chapter I will examine these further, explain the predictive difference in more detail, and set out the research that would distinguish between the two hypotheses.
4. Chapter Four: Writing systems

In this thesis I am investigating the claim that all language varieties are equal. In the past few chapters I have been following the majority of linguists working on this topic, and considering mainly spoken and signed language varieties. Linguists typically have treated writing systems as separate entities from spoken/signed languages. In this chapter I will challenge that assumption. Firstly I will argue that written systems are a part of the language variety they encode, and then I will argue that having a written system gives a huge advantage to a language variety. As such, not all language varieties are equal; those with a written system are better than those without.

4.1 Writing systems are a part of a language

The received view in linguistics is that written language and spoken or signed language are separate phenomena. Crystal claims that “the written language is… not a natural medium of language, as speech is” (Crystal 2006: 23). Rodman explains the difference “writing... is a human invention, like the bicycle, and has to be studied. Talking is a biological trait, like walking, and comes naturally” (Rodman 2006: 8). (See also Pinker 1994: 186; Bloomfield 1935: 282; and Sampson 2013.)

I’m interested in seeing whether all language varieties are equal. So far I have been considering spoken/signed language varieties, while leaving writing systems to the side. Most scholars working in the area of language equality (e.g. Dahl, Nichols and Miestamo) have focused on spoken language, and I have followed their lead. The reason for this split is that some (see above) think that spoken and written language are such different systems, that they can’t legitimately be considered as identical. If that is the case, then we can’t include facts about writing systems in analyses of language equality. It would be like saying some birds are better than others because there are good drawings of them. In order to claim that writing systems give language varieties an advantage, I first have to argue that writing systems are a part of the language variety they encode.

The question is whether or not writing is a part of the spoken language variety it encodes. To put it another way, are writing and speech two separate but related systems (fig.1), or are they two parts of a whole (fig. 2)? What is the relationship between the two modes of language?4

---

4 The above figures present a hugely simplified picture. Clearly there will be a gradient scale for ‘closeness’ of any two related systems. Some will have no interaction whatsoever, while others will be extremely closely intertwined. However there will also be some cases in the middle, systems that have some interaction. I am framing this as a dichotomy, as the question I want to ask is dichotomous: if we ignore the written system of a
Fig. 1 depicts a case where writing systems and speech are essentially two separate communication channels. There can be interaction between the two, however the influence one has on the other is negligible and/or occasional; there is no strong causal relationship either way. A study into one channel would not be remiss if it ignored the other – there would be no major influences left out.

Fig. 2

given language, are we ignoring a significant part of that language? I argue yes, therefore we should include written systems in discussions of the equality of languages.
Fig. 2 depicts the opposing case. In this representation, both systems have a significant influence on the other. Research into one that didn’t survey the other would have a significant deficit.

In systems like that depicted in Fig 2., ignoring half of the equation would be like measuring the presence of Sodium and Chlorine ions (Na\(^+\); Cl\(^-\)) in a solution over time without measuring the quantity of salt (NaCl) poured into the solution. The amount of NaCl explains the amount of Na\(^+\) and Cl\(^-\) ions in the solution. Without that crucial knowledge, part of the puzzle would be missing.

I want to know whether writing is such a separate system from spoken language, that analysing a spoken language variety and excluding its writing system is justifiable? I will argue that such an approach fails in analysing a language; I think Fig. 2 is the best representation of the relationship. In the following section I argue that writing systems are a subset of a language variety.

It is very surprising to find a person who has no spoken or signed language; it is much less surprising, however, to find an illiterate person. 84% of the world’s adult population is literate (UNESCO 2010). This is a significant majority to be sure, however not as high as the percentage of adults that speak a language variety, which is closer to 99.99%.

Writing systems are also much less prevalent than spoken/signed languages, and there isn’t a corresponding writing system for every language variety. Even for those language varieties that do have corresponding writing systems, not every speaker gets the opportunity to learn the writing system. And for those who do have the opportunity, not all manage to learn the system. Pyles and Algeo summarise most of the arguments for treating spoken/signed language and writing separately in the paragraph below.

Human beings have been writing (as far as we can tell from the surviving evidence) for at least 5000 years; but they have been talking for much longer, doubtless ever since there have been human beings. When writing did develop, it was derived from and represented speech, albeit imperfectly, as we shall see in chapter 3. Even today there are spoken languages that have no written form. Furthermore, we all learn to talk well before we learn to write; any human child who is not severely handicapped physically or mentally will learn to talk: a normal human being cannot be prevented from doing so. It is as though we were ‘programmed’ to acquire language in the form of speech. On the other hand, it takes a special effort to learn to write; in the past many intelligent and useful members of society did not acquire the skill, and even today many who speak
languages with writing systems never learn to read or write, while some who learn the rudiments of those skills do so only imperfectly.

(Pyles & Algeo 1982: 9)

The arguments in this paragraph fall into two categories. The first half of the paragraph contains arguments that claim the historical dependency of writing on speech means that nowadays writing is a separate system to speech. As I will show in the following section, this line of reasoning fails. The second group of arguments are more relevant; however I will show that they also fail. I’ll discuss these after the historical dependency arguments.

Human beings have been writing (as far as we can tell from the surviving evidence) for at least 5000 years; but they have been talking for much longer, doubtless ever since there have been human beings.

(9)

The evidence supports the hypothesis that spoken language has been around for a lot longer than writing systems have. But it’s unclear how this fact is supposed to support treating spoken and written language as separate systems. As I will argue shortly, this historical relationship does not determine the relationship of the systems today.

When writing did develop, it was derived from and represented speech, albeit imperfectly, as we shall see in chapter 3.

(9)

The argument here seems to go something like this:

1. When writing was first developed, it was derived from and represented speech
2. If something is derived from, and a representation of another, it is secondary to the original source.
3. If something is secondary to an original source, it should be treated differently to the source.

Therefore

4. Writing is secondary to speech (from 1&2)
5. We should treat spoken and written language differently (from 3&4)
If we know there is a system, and a secondary system based on the primary, it’s not clear that we can infer anything from that information about their relationship. More specifically, in my reconstructed argument above, premises 2 and 3 are not, on the face of it, necessarily true. Both would need to be argued for in order to advance this line of thinking, and neither Pyles and Algeo nor other scholars have put forward arguments for similar statements.

We all learn to talk well before we learn to write. (9)

Like the other arguments in this category, this, while an interesting fact, doesn’t bear any relation to the relationship between writing systems and spoken or signed languages. Almost all humans learn to crawl before they learn to walk. This doesn’t mean that crawling is our primary form of locomotion.

**4.1.1 Two analogies**

There are two characterisations of primary/secondary relationships that seem to be on separate ends of the relatedness continuum. I’ll give two examples which I think fit either end of this continuum. These examples have a history that is analogous in some important respects to the history of language. I’ll use these to show that worries similar to the ones set out by Pyles and Algeo are not a good way to argue that writing and speech should be treated as separate systems. I’ll then identify what is relevant to deciding whether two systems are separate or related, and argue for that written and spoken language are parts of the same system.

**4.1.1.1 The relationship between plays and films.**

Theatre existed long before the technology became available to make movies. When film was first invented, the reels produced were films of plays, that is, the actors would act as in a play, and the performance was filmed. Pretty quickly, though, performances in films started being differentiated from performances in plays. Directors and actors made use of the possibilities that film opened up for them. Nowadays, theatre and film are regarded as separate media. The same stories can be told in both media, indeed there often are film adaptations of plays (e.g. *Proof*, *The Importance of Being Ernest*, *Romeo and Juliet*). There are some professionals (actors, directors etc.) that work in both media, but there are different skills involved. Most
importantly, the media don’t have significant influence on each other, no more so it seems than other cultural niches.

4.1.1.2 The relationship between live music and recorded music.

Similar to theatre, live music existed before recordings were possible. And like films, music recordings were initially ‘live’ recordings of music. That is, the musicians played as they would to an audience, and that performance was recorded. And like films, nowadays artists make use of opportunities the recording system affords them. For example some artists record harmonies with themselves (e.g. Seal’s *Kiss From a Rose*). However unlike the theatre instance, live and recorded music are not completely separate systems. Recordings act as storage units; they can fossilize particular versions of songs. Some artists who play live record some of their music. If someone did a biography of the Beatles’ music and only talked about their live performances, they would be missing a lot.

Both of these relationships are analogous to the relationship between spoken/signed language and writing systems. They all have the historical relationship that Pyles and Algeo think constrains the descendent systems’ relationship. In all the examples there was an initial ‘primary’ system (plays, live music, spoken or signed language) with its own ends and conventions. (This information is also presented in the table below) This primary system existed for a significant period of time, then certain technology developed that enabled the system to be transmitted through a different medium (film, gramophone/records, written language). Initially, this technology was simply used to record the primary system; however quickly a secondary system developed, based on the possibilities opened up by the technology. Pyles and Algeo claim that given certain conditions (rows 2-6 in the table overleaf); a particular relationship between systems is inevitable. However as the table shows, these historical conditions do not determine a relationship.
<table>
<thead>
<tr>
<th></th>
<th>Theatre</th>
<th>Music</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary system</strong></td>
<td>Plays</td>
<td>Live music</td>
</tr>
<tr>
<td><strong>Technological innovation</strong></td>
<td>Film</td>
<td>Gramophone records</td>
</tr>
<tr>
<td><strong>Secondary system</strong></td>
<td>Movies</td>
<td>Recorded music</td>
</tr>
<tr>
<td><strong>Purpose for which the technology was first used</strong></td>
<td>To film plays from an audience’s perspective</td>
<td>To record music from an audience’s perspective</td>
</tr>
<tr>
<td><strong>Later use of the innovation</strong></td>
<td>Specialised movie performances, very rarely filmed plays</td>
<td>Specialised recording performances, sometimes recorded live sessions</td>
</tr>
<tr>
<td><strong>Relationship between primary and secondary systems today</strong></td>
<td>Mostly negligible, only occasional strong causal influence</td>
<td>Quite strong, live music often informs recorded music and vice versa</td>
</tr>
</tbody>
</table>

Pyles and Algeo claim that the historical relationship of writing from speech determines their current relationship. The claim is that for any pair of systems with an analogous historical relationship to language, the current relationship between the systems will be determined by the historical facts. If that is right, then we should expect the theatre descendent systems and the music descendent systems to be strongly analogous. However these systems are not. This shows that the primary-secondary historical relationship is insufficient to determine the closeness of the relationship today.

Pyles and Algeo do mention some arguments that don’t fall make the same mistakes as those discussed above. I’ll briefly discuss these; however I don’t think they pose any serious problems for the claim that writing systems are a part of a language variety.

Any human child who is not severely handicapped physically or mentally will learn to talk: a normal human being cannot be prevented from doing so. It is as though we
were ‘programmed’ to acquire language in the form of speech. On the other hand, it takes a special effort to learn to write. (9)

Spoken (and sign) language is almost universally acquired, whereas written systems are not. While true, this doesn’t illustrate an important difference between spoken and written language. In order to acquire spoken language, a certain amount of scaffolding is necessary: a child has to be surrounded by a language variety, and have interactions with speakers of that language variety. A certain amount of scaffolding is necessary to acquire written language too. The difference in acquisition of the respective systems tells about the difference in scaffolding of those systems. Spoken (and signed) language has very good scaffolding, whereas the scaffolding for written language is less prevalent, and also has a lower success rate. It is unclear what the prevalence of the respective scaffolding tells us about the skills themselves. It could be that aspects of the skills affect how prevalent the respective scaffolding is, or that factors determining scaffolding prevalence are unrelated to facts about the skills themselves. Without further investigation, which I cannot undertake in this thesis, I’ll remain agnostic on the idea.

In the past many intelligent and useful members of society did not acquire the skill (9)

Pyles and Algeo suggest that the ability to acquire a written system is not correlated with intelligence. (Though Marks 2010 suggests that g [a measure of general intelligence] increases with higher literacy skills.) Their comment about ‘useful members of society’ misses the point, however. The skills one needs to become a useful member of a society depend in great part on that society. In the past, for a number of societies, literacy was not considered an important skill. In such societies an illiterate person can be very successful (be a useful, important, or respected member of the group). However in modern Western society you do need to be literate to be successful; literacy has become deeply embedded in our culture. Many manual jobs that didn’t used to require literacy now do. For example drivers need to pass a written test, qualified butchers in New Zealand have to pass some written unit standards, cleaners need to be able to understand written directions. Yes, in some societies (past and present) written language isn’t as prevalent as spoken or signed language, but this doesn’t show that there is an inherent difference in spoken and written language.
Some who learn the rudiments of those skills [literacy] do so only imperfectly.

(Pyles & Algeo 1982: 9)

The implication here is that everyone who learns a first language variety (putting second language learning aside for a moment) speaks it perfectly. This, then, is contrasted with the case of literacy, where it is not the case that if you are literate, you are necessarily perfectly literate. Acquiring a first language variety seems to be an either/or case: you don’t half acquire a first language variety, as contrasted with the slope of literacy skills. However the distinction between these looks less clear when we look into a native speaker’s grasp of the language variety. The claim that native speakers are all perfect speakers of a language variety starts to break down when taken seriously. If all native speakers are perfect speakers, does this mean that all native speakers of some language variety X are equally able to use that language variety in any situation? It’s uncontroversial that some people are better orators than others. While giving a speech involves linguistic skill, arguably it could be the context that creates the difference, not everyone is comfortable with speaking in front of a group. Recently a number of linguists have done research that claims that some native speakers have a better grasp of grammar than others (Chipere 2009). As Chipere claims, some non-native speakers have a better grasp of complex grammatical structures than some of the native speakers.

Clearly there is a significant difference between acquisition of literacy and spoken language, but these characteristics are continuous rather than discrete. For some systems the average skill level of users is high, for others it is lower. There is no inherent difference; one is simply higher on the gradient than the other.

Pyles and Algeo have argued that written systems and spoken/signed language varieties should be treated as separate systems. I have argued against their claims, and shown that they don’t support their conclusion. In the next section I’ll provide evidence that there is significant interaction between written and spoken language, and as such, we should treat them as parts of the same system.

4.1.2 Interaction between spoken and written language

While acknowledging the many differences between spoken and written systems, we can’t deny that they share a very close relationship. Historically, the initial purpose of writing system has almost always been to represent speech (Pyles & Algeo 1982: 62). As writing conventions crystallise, and as literacy rates climb, the effects of one system on another increase.
At least when a writing system is being created, the purpose of that system is to represent speech. We see this clearly in ‘old’ English texts, before spelling reform. There is no fixed spelling, and spellings are usually taken to be written representations of individuals’ speech; thus we can make claims about accents based on spelling evidence. In the following sections I will argue that writing systems have an effect on spoken language, and vice versa. I need to claim this, as I am using the interaction as evidence that the two (spoken or signed and written language) are part of the same system.

4.1.3 Effect of written language on speech

For orthographies where spelling has become fixed, some claim that the relationship between writing and speech has reversed for their speakers. “Nowadays many conceive speech – ideally, at any rate – as the oral representation of writing” (Pyles & Algeo 1982: 62).

A knowledge of spelling has been responsible for changing the pronunciation of certain words whose written forms for one reason or another do not indicate pronunciations that had become traditional. For instance, simply because it occurs in the written language, the t of often has come to be pronounced once again, as it was in earlier days and up until well into the seventeenth century.

Pyles and Algeo (62) mention the changing pronunciation of the English towns of Daventry, Shrewsbury, and Cirencester. The standard pronunciation of these names is now a phonetic match to their spelling, which it did not used to be. Growing up in Wellington, my father impressed upon me that ‘Majoribanks St’ in was pronounced /Marshbanks/ or /Marchbanks/, however /Marshbanks St/ is not a street in Mt Victoria, according to most of my peers.

The phenomenon of spelling pronunciation is not new. Bloomfield noted that “in syntax and vocabulary the message of the written record is unmistakable, and it exerts a tremendous effect upon the standard language” (Bloomfield 1935: 486). Ziegler & Goswami claim that “the acquisition of an alphabetic code is like catching a virus” (2005: 15).

Ziegler and Gowswami give a summary of a number of ways that an alphabetic writing system seems to affect spoken language. Knowledge of an alphabetic orthography affects phoneme analysis, rhyme judgements, and even priming effects.
Ehri and Wilce (1980) show that literate children find it difficult to count the same number of phonemes in /ritʃ/ and /pitʃ/, because the spellings of these words contain a different number of letters (rich vs. pitch)… skilled adults find it more difficult to judge whether two words rhyme when their rimes [sic.] are spelled differently (e.g., rye–tie) than when their rimes are spelled the same (Seidenberg & Tanenhaus, 1979)… Jakimik, Cole, and Rudnicky (1985) studied phonological priming in a lexical decision task. Facilitatory priming effects were obtained only when primes and targets shared both phonology and orthography (i.e., napkin–nap). No priming was obtained when targets shared only phonology (e.g., chocolate–chalk) or only orthography (e.g., fighter–fig; see also Slowiaczek, Soltano, Wieting, & Bishop, 2003).

Deutscher notes that writing systems can counteract some common forces seen in language change.

[A] factor that may contribute to more complex word-structure is the absence of literacy. In fluent speech, there are no real 'spaces' between words, and so when two words frequently appear together they can easily fuse into one. In the written language, however, there are visible gaps between words, and this reinforces our perception of the border between them, and can thus hamper new fusions... So literacy may well be a counter-force that hinders the fusion of words, and thus slows down the emergence of more complex word-structure.

(Deutscher 2005: 273)

In this respect, writing systems can crystallise certain words and phrases. Change, rather than stasis is the norm: “gradual change in the design of a living language is part of its life and is inexorable” (Hockett 1958: 368). As such this crystallisation of written forms is notable. Note, however, that Deutscher doesn’t claim that literate societies cannot undergo this sort of elision change, instead he suspects literacy will tend to slow down such change.

Clearly, written language has a number of varied effects on spoken language. Spelling influences pronunciation, verbal priming, and can retard elision. Literacy affects conception of language, speech perception, and rhyme analysis.
4.1.4 Effect of spoken language on writing

New words tend to arise in spoken, rather than written language. As Deutscher mentions, one way for new words to form is when two (or more) words merge into one. For example ‘going to’ > ‘gonna’; ‘want to’ > “wanna’, both words that are now in the Oxford English Dictionary.\(^5\) Slang tends to originate from situations where spoken language is used, such as children playing in the schoolyard.

Just as spelling can affect pronunciation, so too can pronunciation affect spelling. The English spelling system is not particularly phonetic, and children often have problems when learning to spell words. And so mistakes like ‘patato’ for potato, ‘intelagent’ for intelligent, ‘yore’ for your, are common. Children’s mistakes are corrected, but it would be overambitious to claim that all literate native speakers spell correctly all the time. If enough people make the same mistake enough times, it can lead to a change in spelling.

In a 2009 newspaper article, linguist Laurie Bauer notes “Students now regularly write ‘pronounciation’ because that is what they say, even though the spelling given in dictionaries and the like is still ‘pronunciation’ and the standard pronunciation of ‘pronunciation’ makes it unlike ‘pronounce’” (Bauer 2009).

It seems that spoken language has less of an effect on written language than vice versa. This is probably due to the crystallisation of writing systems: literate people tend to be prescriptivists about writing. It is easier to ‘correct’ a text rather than a speech utterance. Written language becomes somewhat crystallised, change slows down as examples of the target are available long after production. Spoken language varieties tend to change and become more diverse.\(^6\)

As I have argued, writing systems are a part of the language variety they represent. As such, we can use the existence of a writing system (or systems) as a criterion for claiming that

---


\(^6\) That isn’t to say that there is no pressure on speakers to have a certain language variety. Having the same accent as your peers, using the same language as your peers is very important. As Chambers and Trudgill (1997: 76-86) point out, social pressure is a significant source of language change. However therein lies the difference. There are some different ‘literate dialects’ in English: academic English, scientific English, newspaper English; however there aren’t many. The target for ‘correct’ written English is small and stable across a vast number of contexts. When speakers correct a text, they are reinforcing this small number of ‘good’ English literary conventions. There are many many more spoken English varieties than written, and what counts as ‘good’ varies much more as well. In Liverpool, social pressure pushes you to acquire a Liverpudlian accent, in Minnesota the ‘good’ accent is Minnesotan. All these accents and language varieties are constantly changing, so the target is also constantly changing. The same behaviour (pressure to acquire a certain language variety) in written and spoken language leads to different results due to the respective media.
some language varieties are better than others. I do think that language varieties with writing systems have huge advantages over language varieties that don’t, and I’ll consider these advantages in the next section.

4.2 The advantages of a writing system

Writing systems allow linguistic information and meaning to be stored outside of speakers’ brains. Spoken language and signed language are produced and perceived simultaneously. If you are not present for a certain utterance, you have missed your chance to hear that utterance again. The words can be repeated, but the utterance is different. The production and comprehension of written language, however, is not simultaneous, and comprehension can occur at multiple differing times.

The storage function of written texts has a huge impact. Every text is a communication and a record of that communication. Clearly, texts encode meaning. We read *The Odyssey* and extract the story through the symbols. However texts store more than just meaning. There are at least three different types of information that written texts encode. I’ll briefly introduce each type and give examples of textually encoded information.

Speech (and sign) is a pretty good communication channel. People nowadays have knowledge of a number of things passed down through oral histories. Writing, however is a much higher fidelity channel. This high fidelity combined with the various types of information texts store mean that texts store and make available a huge amount of information that speech and sign simply cannot.

4.2.1 Propositional storage

Anyone who has played the childhood game of Chinese Whispers knows how far a message relayed via spoken language message can move from its original meaning. In order for the message to be passed on, it must be replicated, and so there are two points in the process where errors can occur. There are comprehension problems and production problems. A comprehension problem will lead to a failure of production target, which can leads to an error in replication. A production problem is a failure to reach the production target, which can also lead to an error. A written text, however, does not require this comprehension-production chain for meaning to spread. By cutting down on replication errors, this cuts down on a significant amount of communication errors.

Two households, both alike in dignity
In fair Verona where we lay our scene
From ancient grudge break to new mutiny
Where civil blood makes civil hands unclean

(Romeo and Juliet Prologue 1–5)

Millions of people around the world recognise those lines, and know the lines that follow that passage. William Shakespeare’s prose has survived this long because it was written down. We know his version of Romeo and Juliet because we have his folios. No doubt the general outline of the story would survive today without those scripts, as the story had persisted before, but the details would be different.

4.2.2 Metalinguistic storage

Both written and spoken (or signed) versions of language contain meaning as well as metalinguistic information. The sentence ‘I flew in the aeroplane’ contains meaning (that the speaker flew in an aeroplane), and it also contains information about how that information was encoded. This metalinguistic information tells us that the sentence is in English, and also research about the various words contained in the sentence can even date when the sentence was produced. (‘Aeroplane’ was first recorded in 1866⁷, so that’s very probably the oldest that sentence could be.) Indeed this dating can become quite precise. Take the sentence ‘I accessed the internet on the aeroplane’. ‘Internet’ was first recorded in 1974⁸, so we can take that as a back limit; moreover looking at the entry for ‘aeroplane’ we see that it is now chiefly a British word, Americans use ‘airplane’⁹, so it seems likely that the writer of the sentence is British.

Dictionaries explicitly store this kind of linguistic information, however the example sentences above show, it is by no means constrained to texts about languages such as dictionaries, thesauruses and grammar books. One great example of this is verse; when something is written to rhyme we learn about the pronunciation of those words. “Chaucer rimed [sic.] — to quote the words in their present-day spellings — mean with clean, but not with keen, queen, green: he evidently spoke different vowels in these two sets of words” (Bloomfield 1935: 295).

---

To return to Romeo and Juliet, how is it we know how ‘Rosaline’ is to be pronounced? Should the last syllable rhyme with ‘Catherine’ or ‘Caroline’? The reason we know is through Shakespeare’s use of rhyme; ‘Rosaline’ rhymes with ‘mine’.

ROMEO: Thou chid’st me oft for loving Rosaline.
FRIAR LAURENCE: For doting, not for loving, pupil mine.

(Romeo and Juliet 2.3.81–82)

This metalinguistic storage also results in language varieties with a written system having larger vocabularies than language varieties without.

There is one area of language whose complexity is generally acknowledged to depend on culture – this is the size of the vocabulary. The obvious dividing line here is between languages of illiterate societies and those with a written tradition… In illiterate societies… a word that is not actually used by one generation will not be heard by the next generation and will be then lost forever.

(Deutscher 2010: 110)

Illiterate societies with oral traditions can pass down propositional meaning, stories, through the generations. However the initial vocabulary is not necessarily passed down. That is, the vocabulary used to encode the story changes from one retelling to the next, and so the passive vocabulary that Deutscher refers to is not passed on. Having metalinguistic storage dramatically increases the size of vocabulary of a language variety, as arcane terminology is frozen, ready for revival.

4.2.3 Storage of cultural information

Looking at Romeo and Juliet again, one thing we learn is that in Shakespeare’s day, biting your thumb was considered an insult.

SAMPSON: Nay, as they dare. I will bite my thumb at them; which is a disgrace to them, if they bear it.

(Romeo and Juliet 1.1.36–37)
Texts contain cultural information. Not all cultural claims in the text are cultural claims about the society the author lived in. Margaret Atwood certainly didn’t live in a dystopian world where a fertile underclass of women were used as forced surrogates when she wrote *A Handmaid’s Tale*. However with a critical eye, there is cultural information that can be read from a large number of texts. Romeo and Juliet doesn’t tell us that family feuds were common; however it does tell us that it wasn’t uncommon for men in their early twenties to marry teenage girls, Romeo and Juliet’s respective ages are introduced as normal. Homer’s use of colour words doesn’t tell us that the Greeks perceived colour differently to humans nowadays, but it does tell us about how Greeks used colour words.

Texts typically contain all three types of storage. They might be primarily intended to transmit one type of information over others; however it is hard to see how any of the types could be left out entirely. Spoken and signed language can transmit the same information, however written language exceeds those two in detail, fidelity and breadth of transmission.

### 4.2.4 Writing is subject to different constraints than speech

Having two language channels drastically increases the ways in which you can communicate. Different communication channels have different constraints. Speech is constrained by distance; no matter how much I yell I can’t communicate with a person 100km away by spoken language. Speech is also constrained by sound; because of sound amplification at rock concerts, communication is often carried out though a mix of exaggerated facial movements, lip-reading, gesture and mime. Written communication is subject to other constraints, reading (and to a slightly lesser extent writing) requires light, to write you need an implement to create your written symbols, and something to create the symbols on. A written channel in combination with a spoken or signed channel means that some communication constraints don’t apply any more.

Every instance of communication via written language is also a record of that communication. “Before writing there were no historical records of language” (Pyles & Algeo 1982: 63). Speech and sign are instantaneous, and leaves no inherent record. We have ways of recording speech and sign, however these all require foresight. Writing, on the other hand, is both a communication method and a record of that communication.

### 4.2.5 Editing

Writing gives the author the ability to reread a message, to edit, and to double check the wording used. The importance of this cannot be overstated. For short works, individuals can
reframe spoken and signed pieces. I can verbally compose a short poem and edit it without needing a writing system. However working memory is limited, and longer pieces are much harder to improve in this way. Without a writing system, this thesis could not have been created. Newton was a brilliant intellectual, and he would have been brilliant were he illiterate; however it is doubtful whether he would have invented/discovered as much as he did were he not able to write his thoughts down, let alone be able to record the results of his experiments.

4.2.6 Some languages are better than others

In this section I have argued two things. Firstly, I’ve argued that writing systems are a part of the language variety they encode. Rather than being mostly separate systems, writing and speech influence each other to such an extent that to claim something about a language variety with a written system while only talking about the spoken language misses a huge amount. We need to take into account both spoken and written parts (where they exist) of language varieties when considering the equality of language varieties. Secondly, I argued that a writing system gives a language variety significant advantage: a high fidelity communication system where every communication is a record of that communication. Writing stores propositional, metalinguistic and cultural information, it allows for high fidelity transmission of ideas, and enables an author to edit their work quite precisely. Due to these advantages, language varieties with writing systems are, on average, better than language varieties without writing systems.
5. Chapter Five: Conclusions

The claim that all languages are equal is “an idea which ranked for many decades as an unquestioned truism of linguistics” (Sampson 2009: 1). In this thesis I have joined the “flurry of publications from the last couple of years” (Deutscher 2010: 125) that have investigated this claim. In chapter one I conducted a literature review, and set out the interpretations of the thesis that linguists have put forward. In chapter two I set out and explained some of the assumptions that have to be made in order to investigate the thesis. During chapter three I put forward and critically assessed the five structural interpretations of the thesis that I have found. Finally, in chapter four I looked at the impact that including written language might have on the thesis. In this chapter I will review these arguments, explain my findings and discuss where further research might be useful.

5.1 Chapter 2 Conclusion

In chapter two I discussed two assumptions that have to be made before we can get on to evaluating structural interpretations of the thesis that all language varieties are equally complex. Here I will briefly review the assumptions and explain why they need to be made.

5.1.1 Assumption 1

There needs to be a way of picking out the units that the thesis is about. It might seem that this ought to be a way of picking out language boundaries, but as I showed, language boundaries are actually pretty fuzzy. Mutual intelligibility seems like a good way to go: it is typically the criterion linguists use to distinguish between different language varieties. However mutual intelligibility can’t account for a number of linguistic phenomena (asymmetrical understanding, dialect continua).

If we abandon the attempt to pick out languagehood boundaries, we can succeed in delineating the units that the thesis is about. This may seem counterintuitive, however it’s actually just sidestepping a related but different topic. I am remaining agnostic on languagehood criteria. However just as we can research generational attributes by picking out birth-date groups, we can research languages by using isogloss boundaries.

Isogloss boundaries are a way of picking out language varieties. A boundary can be drawn around any geographic or population-based linguistic feature. This approach can pick out narrow varieties (e.g. Southland NZ English) or wide varieties (e.g. French). All linguistic systems that fit the languagehood criteria (assuming there are such criteria) are language
varieties. Using isogloss boundaries allows us to investigate the thesis that all language varieties are equal without getting involved in the debate about languagehood criteria.

5.1.2 Assumption 2
The second assumption is that there is a way of measuring language structure which enables you to compare different language varieties. Without such an assumption, we can’t compare different languages varieties, and so no interpretations of the thesis that all language varieties are equal can get off the ground.

I discussed relative complexity and absolute complexity as approaches to measuring language structure. I argued that absolute complexity (as described by Miestamo) provides the right framework for a complexity metric by which we can compare languages. Exactly which metric to use comes down to technical choices; e.g. whether phonemic complexity should be measured, verb endings, and how they should be weighted. I’m not going to get into a discussion of how to construct such a metric. What I showed is that it is at least plausible that we can construct a metric for comparing language structure. Having these two assumptions, I can go on to consider the structural interpretations of the thesis that all language varieties are equal.

5.2 Two live structural options
I have investigated structural interpretations of the thesis that all language varieties are equal. In the past 15-20 years scholars have started querying the thesis, focusing on structural interpretations of the claim that all language varieties are equal. I have followed their lead and concentrated on structural interpretations as well. As you will recall, in order to consider these hypotheses, I have to assume that there is a metric for language structure against which we can compare different language varieties.

In chapter three I looked at five structural interpretations of the thesis. The static value hypothesis is the claim that all language varieties when compared against the commensurable metric share the same value. The weak maximum hypothesis states that there is a maximum threshold for language and that most (but not all) language varieties have reached this maximum. The weak minimum hypothesis is the claim that there is a minimum threshold for language structure and all language varieties have surpassed this threshold. The double threshold hypothesis is that there are two thresholds that impact language structure, a maximum and a minimum threshold. Lastly, the weak equilibrium hypothesis proposes that there are opposing forces which keep language structure in a band of optimal values.
I concluded that there are two potential true interpretations of the thesis: the double threshold hypothesis and the weak equilibrium hypothesis. Both posit that language varieties measured according to a commensurable metric will have a variety of values. These values lie in an optimal band: however the mechanisms that create the phenomena differ for the two hypotheses. In this section I will set out these hypotheses in more detail, and then set out the data we would need to decide between the two.

All languages are roughly equal in complexity... Some languages do, of course, have complex word structure while others have rather simple word structure; but those with simple word structure tend often to have complex sentence structure, and vice versa. The areas of complexity differ from language to language but the overall level of complexity is about the same... So, all languages are roughly equal in complexity. But are they all exactly equal? Plainly not.

(Dixon 2011: 36)

The double threshold hypothesis claims that there are two structural thresholds for language, one minimum and one maximum. This creates an optimal range for language varieties, as we can see in the diagram.

![Double threshold hypothesis diagram](image)

The weak equilibrium hypothesis gives a subtly different picture. Instead of opposing thresholds, this hypothesis posits opposing forces. It also has an optimal range for language varieties.
The difference between the two hypotheses is in the proposed mechanism by which the range of language values is achieved. Thresholds only have an effect on phenomena that come into close range, whereas forces affect all of the relevant phenomena. Thresholds are like the edges of a pool table; they mark out the possible space that a language variety can be in, and have localised effects. The edges of a pool table have no effect on a ball lying in the middle of the table. The double threshold hypothesis posits constrained forces (thresholds) that have some effect on language varieties, the weak equilibrium hypothesis posits forces that continually affect language varieties.

5.2.1 Both hypotheses are plausible

First of all, as mentioned in the previous chapter, we have to assume that there is a way of measuring language structure such that we can compare different languages. According to these hypotheses we should expect language varieties to fall across a variety of values when measured with this metric. The data that I have reviewed so far support this claim. There are a number of studies which show that language varieties do not have exactly the same value when compared according to a commensurable metric. (Dahl, Nichols, Deutscher, McWhorter etc.) As seen in chapter three, this evidence rules out the other structural interpretations of the all language varieties are equal thesis. The available evidence does not rule out the two hypotheses I’m focussing on in this chapter, although so far there hasn’t been much
supporting evidence either. In the next section I’ll explain how the two hypotheses are empirically different, and consider what evidence would support these two.

**5.2.2 The hypotheses are empirically different**

There is a worry that these are just different ways of describing exactly the same phenomenon. The two hypotheses may appear similar; however the predictions each hypothesis makes are importantly different. They make opposing predictions for language change over time.

Both hypothesise that language scores (according to a suitable metric) fluctuate within a constrained set of values. However that is the extent of the similarity. The mechanisms accounting for the predicted data differ significantly, and these mechanisms mean that the predicted data for each hypothesis is different.

A threshold only interacts with the phenomenon (in this case complexity of language structure) when the phenomenon is at close range. Imagine a ball on a pool table. So long as a pool ball remains somewhere in the middle of the table, the table sides will not have any effect on the ball’s flight. Whether the table sides affect the ball’s motion depends on the sides’ proximity to the ball. The felt, on the other hand, affects the ball’s motion no matter where the ball is on the pool table.

Suppose all we know is that a ball is on a surface. The surface could be infinite, bounded, flat, tilted, we don’t know. However we can track the ball’s movements. And while we can’t directly view the playing surface, we can learn a lot about it by studying the ball’s movements.

Tracking the ball’s movement over time gives us information about the ball’s environment. It can tell us about the mechanisms that shape the phenomena. Importantly the relevant data is a longitudinal study; it tracks the movement over time. A collection of non-time-sequenced points of various balls on the same surface would not be as useful as the longitudinal study.

So far, most linguistic data I have surveyed are cross-sectional studies. These look at a language variety, or language varieties at one point in time. However as we have seen above, longitudinal studies are important to test the double threshold and weak equilibrium hypotheses.

The thresholds of the double threshold hypothesis, like the sides of a pool table only affect the phenomenon at certain points. These longitudinal studies need to track language variety values according to a commensurable metric (as discussed in section 2.2, pages 32–
According to this hypothesis we should expect quite strict bounds shaping language. The data to support these bounds would show a number of language varieties ‘hitting’ an upper and lower threshold.

### 5.2.3 Current data is inconclusive

Most of the data that linguists have collected to investigate the hypothesis that all language varieties are equal has explicitly been to test the strict equilibrium hypothesis. (Dahl, Nichols etc.) These investigations have measured language varieties according to some metric, and looked to see whether two or more language varieties have the same score. This approach, however, won’t be useful for differentiating between the two hypotheses here.

Longitudinal studies are needed to test the hypotheses at issue. The crucial difference between them is in their predictions for language behaviour over time. As I mentioned in the previous chapter, this comes down to the mechanisms posited: thresholds or forces. Forces are constantly acting; the weak equilibrium hypothesis posits two forces that are acting on all language varieties at all times. Thresholds, on the other hand, are constraints on forces, which have localised effects. The effects of these mechanisms at any one point in time will be roughly similar. That is, at any one point in time there won’t be data. However when followed over time, the differences become clear.

Of the studies relating to the thesis that all language varieties are equal there is one longitudinal study relevant here. Deutscher (2000) studied Akkadian texts from 2500 BC to 500 BC. (Tablet writing was reasonably common; there are a large number of surviving Akkadian texts.) However Deutscher’s study does not support either of the two theses in question: he concludes that Akkadian in 500 BC is more complex than Akkadian in 2500 BC.

I urge caution in taking Deutscher’s research as evidence that the double threshold hypothesis and the weak equilibrium hypotheses are false. His data does contradict the two hypotheses in question: a language variety rising steadily for 2,000 years does not look like it has any compensatory mechanisms. However while Deutscher’s study is evidence against the hypotheses it is also only one study. Moreover it is a study of written language, and so it isn’t clear whether this is the right kind of evidence we want.

Aitchison hasn’t carried out any longitudinal studies; however, she has looked at a number of cross-sectional studies to investigate whether language is ‘progressing or decaying’.

“We must conclude therefore that language is ebbing and flowing like the tide, but neither progressing nor decaying, as far as we can tell” (Aitchison 1991: 215). Though not a
longitudinal study, Aitchison’s literature review does seem to point to something like the double threshold or weak equilibrium hypotheses.

There is one study that seems to refute the hypotheses, and one that seems to support them both. We need more data to see what the actual picture is here. There are two questions that need to be answered: are either Aitchison or Deutscher’s broad conclusions warranted? And if Aitchison is right, then does the data support the double threshold hypothesis, or the weak equilibrium hypothesis? We can only decide on this issue after obtaining and analysing more data, and to be specific, more longitudinal data.

5.2.4 Conclusion

The double threshold hypothesis and the weak equilibrium hypothesis are at present both tenable positions. However there is a dearth of evidence, both for and against each hypothesis. Before we can say whether the hypotheses describe the actual phenomenon, we need more longitudinal data to let us see the phenomenon more clearly.

5.3 Chapter 4 Conclusion

In chapter four I discussed another way in which one could interpret the thesis that all language varieties are equal. Linguists typically think of speech/sign language and written language as separate phenomena. Hence so far spoken language has been the focus of investigation into the thesis that all language varieties are equal. However if written systems are a part of the spoken (or signed) language they encode, the thesis that all language varieties are equal is wrong. Some language varieties are better than others, namely those with written systems.

In this chapter I argued firstly that written systems and spoken languages are parts of a whole. There is significant interaction between the two; there are measurable effects of both on the other. As such we ought to treat the two as different parts of a ‘language’.

After having established that, I argued that written systems give a language variety many advantages. These include storing different types of information (propositional, metalinguistic and cultural), making large scale editing feasible, and as a different communication medium, being subject to different constraints. Writing is a high fidelity communication system where every communication is simultaneously a record of that communication. Due to these advantages, language varieties with writing systems are better on average than language varieties without writing systems.
5.4 Summary

In this thesis I investigated the claim that all languages are equal. I found two potential ways in which this could be true. The weak equilibrium hypothesis and the double threshold hypothesis both posit that language structure (according to a suitable metric) varies over time by fluctuating within some close bounds. These hypotheses are contradictory, they can’t both be true. We need more research in order see which, if either, is true. However given the current data these hypotheses seem promising.

It would be interesting to see what impact having a writing system has on language structure. If written systems impact general language to such a degree as to impact language structure as measured by a commensurable metric, then the difference between languages with and without a writing system becomes even more significant. In order to look at this correlation, we’d need the same longitudinal data as mentioned in the previous paragraph.

In this thesis I also found a negative answer to the question of whether languages are equal. Languages with a writing system have a huge advantage over languages without. Therefore languages with writing systems are better than those without.

There are two main directions for further research to go in, metric choice and longitudinal studies.

5.4.1 Metric choice

In this thesis I have argued that it is plausible that there is a metric for measuring language structure by which we can compare language. I argued that absolute complexity (focussing on language abstracted away from usage) is the right approach to take. As I mentioned, there are many potential metrics within this absolute complexity approach. Due to space constraints I did not discuss metric choice. However this is a field ripe for more research. For more discussion of absolute complexity metrics, look at Miestamo 2006, 2008, Kusters 2008, Dahl 2009 and Kieraś 2010.

5.4.2 Longitudinal studies

In chapter three I argued that there are two potentially true interpretations of the thesis that all language are equal, the weak equilibrium hypothesis and the double threshold hypothesis. As I mentioned in this chapter, we do not have access to the data we would need to test these hypotheses. We need longitudinal studies of language change over time to properly test the
weak equilibrium and double threshold hypotheses. Both make testable predictions about language change over time so it is this longitudinal data we need to give a definitive answer.

Currently there is a surge of interest in investigating the thesis that all languages are equal. In the past 10-15 years there has been a small but growing literature including some edited volumes (Miestamo et al. 2008, Sampson et al. 2009) specifically addressing the claim. The upcoming Poznań Linguistic Conference\textsuperscript{10} has scheduled a special debate on linguistic complexity. More research is coming out every day.

References:


http://www.phon.ucl.ac.uk/home/dick/papers/texts/cliewp1.doc.


Sterelny, Kim, and Paul E. Griffiths. Sex and Death: An Introduction to Philosophy of Biology, n.d.


