Investigations in item stability:
In pursuit of the optimal meaning list for use in the initial stages of the comparative method

Examination number: 6866540

Master of Science by Research Linguistics
The University of Edinburgh

2012

Abstract

The lack of a widely-accepted, objectively-defined standard list of 'basic' meanings for use in the initial stages of the comparative method is identified as a priority in the resolution of areas of unnecessary subjectivity in historical and comparative linguistics. A methodology is presented, capable of ranking meanings by a score fully representative of the four features identified as necessary for a meaning to be considered optimal for use in the initial stages of language comparison: maximal item stability, maximal resistance to replacement of form by borrowing, maximal conceptual simplicity, and maximal universality. The stability of 67 meanings is quantified using a procedure described, but not adequately implemented, by Dolgopolsky (1986) and Lohr (1999); the results are integrated with Tadmor et al.'s (2010) borrowed, analyzability, and representation scores, to form a composite score by which the meanings are ranked. The resultant ranking, while not representative of the definitive list of meanings optimal for use in the initial stages of the comparative method, owing to the limitation on the number of input meanings, demonstrates the viability of the methodology presented here. Statistical results are presented to support the hypothesis that there is a strongly significant relationship between item stability and variation in stability; however, contrary to expectations and conflations evident in the literature, no evidence is found to support the hypotheses that there are strongly significant relationships between item stability and item borrowability, analyzability of form, or item universality. Finally, the results are used to test the validity of the glottochronological hypothesis of a constant rate of replacement; no support whatsoever is found to support this hypothesis.

Acknowledgements

First and foremost, I would like to extend my sincerest thanks to my supervisors, Prof. Ronnie Cann and Dr. Patrick Honeybone. Their suggestions, criticisms, and encouragement have been absolutely invaluable; without their input, this dissertation simply would not have been what it is.

I am extremely grateful the specialists consulted to verify the data collected in the course of this dissertation: Prof. James Copeland (Tarahumara), Prof. Paul Dundas (Sanskrit), Gaddam Renuka Devi (Telugu), and Maziar Toosarvandani (Northern Paiute). Particular thanks are owed to Prof. Ron Asher (Tamil and Malayalam), who went over and above the verification process, cheerfully answering my questions about the intricacies of comparative Dravidian linguistics. Brian Stubbs was equally happy to help with questions about Uto-Aztecan, and kindly made available the most recent edition of his comparative dictionary. Thanks are due to Dr. Simon Greenhill, who permitted the use of data from the Austronesian Basic Vocabulary Database in this study. For helping to refresh my understanding of the mathematics of probability, I would like to thank Prof. Simon Kirby. I am obliged also to Prof. Ben Fortson, whose stimulating comments encouraged me to refine and develop certain aspects of my approach to meaning stability. All errors, of course, remain my own.

Financial support for the completion of this dissertation was generously provided through a full studentship from the Arts and Humanities Research Council.

While the probability that they will ever see this is close to zero, thanks are certainly due to the brothers Hartnoll and the brothers Sandison and Eoin, for shaping my experience of the data collection. Thank you, Alan and Sonia Fynn, for the kind use of your cherry tree in the final stages of writing up. Finally, Tom: you always ask the right questions, whether regarding meaning stability, or stability of meaning.

Contents

1 Introduction

A genetic relationship between languages is a relationship which posits a common ancestor for the languages concerned. By analogy with the concept of a family tree, languages in the same language family are those which are thought to have diverged from this common ancestor. Features are inherited from the parent language throughout the daughter languages' linguistic systems, giving rise to similarities with a genetic origin between those languages which retain the same features. Similarities occurring between languages can also be caused by non-genetic factors, which can be broadly divided into three types: that of chance alone (notably forms which are similar semantically and phonologically purely by chance, e.g. English /bæd/ and Persian /bæd/, both meaning 'bad'); the presence of universal tendencies (such as onomatopoeic forms, e.g. English *splash*, *clap*, and *click*); and the natural proclivity for speakers of languages to borrow features at all levels of the linguistic system, but especially at the lexical level, from speakers of other languages, if the socio-cultural, geographic, and linguistic contexts facilitate it (e.g. Anglo-Norman *parsone* > Present Day English *person*; English *sandwich* > French *sandwich*).

To date, the comparative method remains the only methodology widely considered to be capable of providing support for hypotheses positing a genetic relationship between languages (see, for example, Campbell and Poser 2008; Fox 1995; McMahon and McMahon 2005). While other methodologies have been proposed to determine the validity of such proposals, notably Multilateral Comparison (Greenberg 1963, 1987, 2000; Ruhlen 1991, 1994) and some applications of lexicostatistical or glottochronological methods to provide initial support for genetic relationship (for example, Swadesh 1972:279; Tovar et al. 1961, cited in McMahon and McMahon 2005:36), these methodologies have all been criticised for not being able to show beyond reasonable doubt that the observed similarities have arisen due to inheritance, rather than non-genetic factors such as borrowing or chance resemblance (see, for example, Campbell 1988, Campbell and Poser 2008:264-278, Matisoff 1990, and McMahon and McMahon 1995, for detailed rebuttals of Multilateral Comparison; Campbell 1998:185-6, Trask 2007 [1996]:458-9 for criticisms of initial classification using lexicostatistical methods).

Nonetheless, while this remains the only means by which one can provide support for hypotheses of genetic relationship between languages to the satisfaction of the general linguistic community, the comparative method itself is, at present, far from being a scientific methodology. As Nettle summarises: "Different people with different degrees of knowledge or different assessments of evidence may come up with substantially different results, without either having departed from the 'method' " (1999:406). It would be inappropriate to mechanise all aspects of the comparative method, owing to the significant role played in its application by the language specialist's intimate knowledge of the languages and cultures concerned; however, if historical and comparative linguistics is to progress as a replicable and objective discipline, we must, as a matter of priority, identify, explore, and ultimately attempt to resolve those areas which have lead to the state of affairs described by Nettle. This dissertation will therefore address the most immediate concerns in this area of historical and comparative linguistics: the identification of areas in the application of the comparative method which permit unnecessary and undesirable levels of subjectivity.

In the process of identifying these areas, it will be argued that the rigorous and objective identification of which meanings maximally facilitate the implementation of the comparative method – which meanings are most 'basic' – should be our priority. Defining meanings which are maximally 'basic' in the context of the facilitation of the comparative method as those which are maximally stable, maximally resistant to borrowing, maximally conceptually simple, and maximally universal, we will demonstrate that there does not, at present, exist an adequate standardised list of meanings, selected solely for their optimality for use in the initial stages of the comparative method. Specifically, we find that, while Tadmor et al.'s (2010) determination of item borrowability, simplicity, and universality is adequate to identify which meanings are optimal in terms of these features, no study has yet quantified item stability in a manner satisfactory for our purposes.

Therefore, using a procedure described, but not appropriately implemented, by both Dolgopolsky (1986) and Lohr (1999), the focus of this dissertation will be on the transparent quantification of the stability of 67 meanings using cross-linguistic data. This quantification will then be integrated with the quantifications of item borrowability, simplicity, and universality presented in Tadmor et al. (2010). The

product of the four scores is a composite score maximally representative of the features of item stability, borrowability, conceptual simplicity, and universality; the input items will be ranked by this composite score, thus determining the input items' relative optimality for use in the initial stages of language comparison. While not resulting in a definitive list, owing to the necessary restriction on the number of meanings for which stability will be investigated, this dissertation will demonstrate the viability of this methodology for identifying which items are most basic in the context of the comparative method cross-linguistically.

The careful separation of the feature of item stability from the other three features of a meaning will enable a detailed and statistical exploration of further issues pertinent to item basicness: the relationship between the stability of an item and the variation in stability cross-linguistically, and the relationship between item stability and item borrowability, conceptual simplicity, and universality. With the results gained in the course of this dissertation, we will also have the opportunity to determine the validity of the central hypothesis of glottochronological theory – that there is a constant rate of replacement of form in the most basic meanings. Statistical evidence will be presented to support the hypothesis that the more stable an item is cross-linguistically, the less this stability will vary from language family to language family; it will be shown, however, that there is no evidence to support a hypothesis of a strongly significant relationship between item stability and the three other features of a basic item, contrary to the expectations and conflations often present in the literature. Finally, it will be shown that there is no statistical support whatsoever for the glottochronological hypothesis of a constant rate of change in basic meanings.

The quantification of item stability has, to date, never been used in the way presented here to explore issues regarding variation in stability, the relationship of the four features of a basic meaning, or the glottochronological constant. The interpretations and implications of these results will thus be considered thoroughly. The significance of the results presented for historical and comparative linguistics, as well as the discipline of linguistics more generally, supplement and further justify the main focus of this paper: the demonstration of a methodology capable of determining the optimal meanings for use in the initial stages of the comparative method.

2 Meaning lists in language comparison

2.1 The identification of areas of subjectivity in the comparative method

The comparative method is a methodology which has been developed to identify which similarities between languages exist because they are features inherited from a common ancestral language, and which have non-genetic origins.[1] Two main stages of the comparative method are distinguished in the literature: the initial determination of whether features have been inherited from a common ancestor, and, if so, demonstrating which ones, with the goal of determining which languages are genetically related; and the reconstruction of the ancestral proto-language of those languages identified as genetically related, based on those features established as having been inherited (Antilla 1989:318; Fox 1995:60; McMahon and McMahon 2005:8-10). These stages can be repeated recursively in order to sub-group languages within a language family, often resulting in a diagrammatic representation of the relationships determined with a family tree. This discussion will focus primarily on the initial stages: the lexical, phonological, and morphological comparison used to determine which features are similar between languages because they have been inherited from the same parent language – which are cognate – and therefore which features provide evidence to support a hypothesis of genetic relationship. As will be shown in this section, there are numerous issues concerning the application of the comparative method in the initial stages of language comparison which remain unnecessarily subjective and controversial; as errors in the initial stages of any method will be compounded as one progresses to later stages of implementation, it is of critical importance to resolve any such issues as a matter of priority.

The initial stages of the comparative method focus on lexical comparison, with a goal of phonological comparison and reconstruction. An approach focussing on phonological comparison is justified by Fox:

[1] The two explanations are of course not mutually exclusive; two genetically related languages may have some similarities which have arisen because they are inherited, some similarities which are due to borrowing between the languages, some forms which tend towards similarity cross-linguistically, and some similarities which can be ascribed to chance alone.

> The methods described [i.e. the comparative method] can in principle be applied to all levels of language, phonology, morphology, syntax, lexical semantics, and so on, but it is primarily with phonological reconstruction that we shall be concerned in the first instance, since the procedures can be applied here in a more controlled and consistent manner than elsewhere. (1995:58)

However, in order to achieve phonological comparison and the establishment of phonological correspondences as described below, one must first compare lexical evidence in order to determine which sounds correspond.

Others have suggested that more emphasis should be placed in the first instance on shared morphological (sometimes referred to as 'syntactic') features. Nichols (1996), for example, discusses the need for 'individual-identifying' evidence as initial proof of genetic relationship between languages: one piece of evidence which on its own is highly unlikely to have been innovated independently in the languages concerned. Owing to the statistical threshold that she advocates, $p < 0.000\ 01$, in practice this will often mean a full morphological paradigm or equivalent. Ramifications of this position include the demonstration of cognate sets and sound correspondences not as proof of relationship, but as aids to reconstruction; furthermore, the semantic latitude permitted in the search for cognate vocabulary after individual-identifying evidence has been pinpointed is as wide as necessary to identify correspondences, as long as the forms adhere to the regular correspondences.

While Nichols' dedication to statistical significance is admirable and to be encouraged, Lohr's proposal (1999:8) that several pieces of less paradigmatic evidence combined, such as recurring regular sound correspondences as discussed below, could also reach this level of statistical significance, suggests a means of progressing beyond the initial stages of the comparative method when comparing languages which may not retain such similarities, but which are still demonstrably related. Furthermore, Nichols' approach does not allow for cases in which, owing to phonological shift, evidence which reaches this statistical threshold cannot be uncovered until phonological comparison, based in lexical comparison, has been undertaken. Thus, while individual-identifying evidence may be necessary to demonstrate the genetic relationship between languages, this dissertation takes the view that this can take the form of several pieces of evidence combined, as long as

Proto-Indo-European[3]		**Proto-Germanic**
*/p t k k^w/	>	*/f θ x x^w/
*/b d g g^w/	>	*/p t k ^{w}/
*/b^h d^h g^h g^{wh}/	>	*/b d g g^w/

Table 1: The first Germanic consonant shift

together they reach a level of statistical significance; lexical comparison, enabling phonological comparison, is therefore a valid focus.

Ross and Durie remind us: "When historical linguists talk about the 'comparative method', what they usually have in mind is not just a method but an associated theory" (1996:3). The comparative method relies in large part on the Neogrammarian principle of regular sound change, which dictates that, when a phonological shift occurs in a language (whether conditioned – occurring in specific phonological environments only – or unconditioned), it is regular[2]; in other words, all relevant segments are affected. This will give rise to correspondences in sounds between related languages. A classic example is that of the first Germanic consonant shift, in which a phonological shift, occurring in the development of Proto-Germanic from Proto-Indo-European (PIE), affected the plosives in the manner shown in Table 1.

This phonological shift was regular; Verner's Law (stating that PIE voiceless stops and */s/, when neither word-initial nor preceded by the original PIE accent, became voiced) and the identification that the shift did not occur after some specific consonants (e.g. Latin *noct-* 'night', OE *niht*, rather than the expected **nihθ*) describe the predictable conditioning environments of apparent exceptions. The first Germanic consonant shift resulted in phonological correspondences between the Germanic languages and other PIE languages, shown in Table 2.[4]

[2] Recent work has suggested that phonological change may not always be regular; unfortunately, a discussion of regularity and irregularity in linguistic comparison and reconstruction is beyond the scope of this paper. See Durie and Ross (1996) and references therein for an informed discussion.

[3] This reconstruction of the Proto-Indo-European stop series follows e.g. Fortson (2004:51). Some Indo-Europeanists differ in their reconstruction, favouring the proposed 'glottalic theory'; see, e.g. Gamkrelidze and Ivanov (1995), Hopper (1973), and, latterly, Beekes (2011 [1995]:128-9).

[4] Note some later phonological shifts obscure the original correspondences.

Greek	**Latin**	**Sanskrit**	**Gothic**	**Old English**	
patēr	**p**ater	**p**itā'	**f**adar	**f**æder	'father'
treîs	**t**rēs	**t**rayas	**þ**reis	**þ**rī	'three'
(he)-**k**atōn	**c**entum	**ś**átām	**h**und	**h**und	'hundred'
déka	**d**ecem	**d**áśa	**t**aihum	**t**ēon	'ten'
geúomai	**g**ustus	**dʒ**ōs-	**k**iusan	**c**eōsan	'taste, test, choose'
phérō	**f**erō	**bh**arāmi	**b**aira	**b**eoru	'I carry'
(é)-**th**ē-ka	**f**ēcī	a-**dh**ām	(ga)-**d**ē-þ-s	**d**ǣd	'put/do; deed'
kheúō	**f**u-n-dō	**h**o-tar	**g**iutan	**g**ēotan	'pour'

Table 2: Indo-European correspondences (adapted from Trask 2007:119)

Repeated, regular sound correspondences are less likely to occur between languages than mere phonological identity, and thus reduce significantly the probability that the similarities observed can be adduced to chance alone; the demonstration of such correspondences is therefore taken as strong evidence to support hypotheses postulating genetic relationships between languages.

In order to reduce still further the probability that observed similarities can be ascribed to chance, only forms which are the same semantically, or follow well-attested paths of semantic change, are typically analysed for correspondences, in order to avoid the identification of correspondences which could be falsely attributed to genetic relationship. An example, shown below in Table 3, demonstrates how a wide semantic latitude in a comparison of Basque and English, two languages not presently considered to be demonstrably genetically related, can present the apparently regular, repeated $h{\sim}h$ and $r{\sim}r$ correspondences, spuriously indicative of a genetic relationship.[5]

Basque		**English**
***h*o*r*i** 'that'	∼	***h*e*r*e**
***h*o*r*ma** 'ice'	∼	***h*a*i*r**
***h*i*r*u** 'three'	∼	***h*und*r*ed**

Table 3: Spurious correspondences between Basque and English

[5] Basque forms from Saltarelli (1988:298).

The wider a semantic latitude one permits, the more forms will be eligible for comparison, and thus the greater the probability that data entering the comparison are similar simply by chance. In addition, one must also bear in mind that any lexical cognates established will have developed from the same word in a common language not only in form but also in meaning; the meaning shift must therefore be plausible. However, precisely what constitutes an expected or plausible semantic shift is wildly subjective, owing to the current lack – and potential impossibility – of a widely accepted model for semantic change.

The comparison of meanings which have a high probability of being represented by forms which tend to be similar cross-linguistically is to be avoided in the initial stages of the comparative method. This helps to ensure that the only reasonable explanation for similarities presented to support a hypothesis of genetic relationship is inheritance from a common ancestor. Similarities between onomatopoeic and sound-symbolic forms, as well as forms which may, cross-linguistically, tend to originate in nursery words (such as Jakobson's (1960) *mama-papa* phenomenon) are all liable to be explained thus. While the definition of what constitutes a meaning which may be represented by an onomatopoeic, sound-symbolic, or nursery form is, at present, woefully subjective (as identified by, for example, Campbell and Poser 2008:197), reasonable care must be taken to ensure that forms for such meanings do not enter the comparanda in the initial stages of the comparative method.

The procedures detailed above – the demonstration of regular, repeated sound correspondences in forms which are similar semantically, or differ only along expected semantic paths, in areas of the vocabulary which are not likely to be similar cross-linguistically – reduce the probability that similarities observed between two languages are present because of chance or cross-linguistic tendencies. However, there is a third non-genetic explanation for similarity we have yet to discuss: that of borrowing, particularly lexical borrowing, between languages. Indeed, a further benefit of the demonstration of repeated, regular sound correspondences is that the probability that the similarities in the data under comparison are a result of borrowing is also reduced. This is because any forms which were borrowed after the phonological shift that resulted in these correspondences will not have been candidates for the shift, and thus will not display the same correspondences. However, the demonstration of sound correspondences is not foolproof means of

excluding borrowed forms from being erroneously identified as inherited forms: forms which were borrowed before a phonological shift will be equally liable to undergo the shift as inherited forms, and thus will display correspondences in forms which are also similar semantically. False evidence for a genetic relationship between two languages can thus be presented.

One way in which those attempting to provide initial proof for a hypothesis of genetic relationship between languages can further reduce the probability that any similarities observed are due to borrowing is to compare forms for only those meanings which are considered to be the most 'basic' or 'culture-free'. Unfortunately, only very loose definitions currently exist to define what a list of such meanings might comprise, as is demonstrated in the literature: "Basic vocabulary [is]...understood intuitively to contain terms for common body parts, close kin, frequently encountered aspects of the natural world, and low numbers. It is assumed that...in general, basic vocabulary is more resistant to borrowing..." (Campbell and Poser 2008:166); "Those words in a language which are of very high frequency, which are learned early by children, and which are supposedly more resistant to lexical replacement than other words" (Trask 2000:39); "There is one aspect of language in which change has been found to move along at an approximately constant rate: that portion of the lexicon which has least to do with cultural advance, and which has been called *basic vocabulary*" (Swadesh 1972:32; emphasis in original); "[basic vocabulary is] usually taken to imply a set of words similar to those on Swadesh's lexicostatistical lists i.e. numerals, pronouns, nouns for body parts, family members, natural phenomena" (Lohr 1999:12).

A brief discussion of the terminology to be employed in this dissertation is here warranted. In the literature to date, by far the most common way of referring to lists of the type under discussion here is by using terms which identify the lexical component, for example 'core vocabulary', 'basic vocabulary', or 'stable words' (e.g. Brown et al. 2008; Campbell 1998:112; Fox 1995:66; Starostin 2000; Swadesh 1950, 1952, 1955, 1972). These terms, however, are not entirely accurate for our purposes; it is not lists of *words* which are under consideration here, but lists of *meanings*, semantic 'slots', represented by word forms in the individual languages. Therefore, following Heggarty (2010), Lohr (1999), and McMahon and McMahon (2005), the way in which these lists will be referred to will be 'meaning lists', lists

comprising 'meanings' or semantic 'items'. This is not to say there is no place for lexically-based terms such as 'vocabulary list' when they are demanded by context (for example, when referring to the list of words drawn up for a particular language by filling in the semantic slots of a meaning list); however, when referring to the lists at the meta-level of cross-linguistic comparison, 'meaning list' is more appropriate.

There does not, at present, exist a widely-used list designed specifically for use in the initial stages of the comparative method; the quotes given above aptly summarise the confusion caused by the lack of a standard definition of what constitutes a language's 'basic vocabulary'. Indeed, the subjectivity with which such a semantic area is defined is lamented by Campbell (1998:314): "Basic vocabulary is usually not defined rigourously..."; Fox (1995:66) similarly protests: "...it is impossible to determine which items of vocabulary are 'basic'". The subjectivity, currently used to determine those meanings which are most appropriate for comparison of form between languages in the initial stages of the comparative method, is clearly undesirable.

This section has identified three areas in the application of the initial stages of the comparative method which lack rigorous definition: what constitutes an 'expected' semantic shift; which meanings are likely to be represented by forms which tend to be similar cross-linguistically; and which meanings are cross-linguistically most 'basic'. In order to narrow the field of investigation, allowing for focussed and effective research, the definition of what would constitute an 'expected' path of semantic change for a form, as well as the determination of which meanings are likely to be represented by forms which tend to be similar cross-linguistically, depends on the definition of which meanings are to be compared in the first place.[6] Our first step, therefore, in untangling these knots is to take a closer look at what the most appropriate meanings for comparison in the initial stages of the comparative method are – which meanings are the most 'basic'.

This dissertation will thus explore the issues concerning the definition of an empirical, objectively derived list of those meanings which are most basic cross-

[6] While the incorporation of the exploration of issues regarding universal tendency towards similarity of form with the definition of a list of maximally 'basic' meanings is possible, and will be discussed in section 5, we will treat the two separately for the purposes of the initial research described in this dissertation.

linguistically, for use in the initial stages of the comparative method. Such a list would reduce the unnecessary subjectivity present in the application of the comparative method in language comparison, and increase the reliability and replicability of comparative linguists' results; it would also allow for further research into other areas of subjectivity in the comparative method. In the remainder of this section, we will identify the differences between the role of standardised meaning lists in lexicostatistical and glottochronological studies and the role of standardised meaning lists in the comparative method, before turning to evaluate the work of others who have addressed the concerns presented here. Our first port of call, however, should be an attempt to define precisely what is meant by the 'basicness' of a meaning in the context of the comparative method.

2.2 Defining 'basicness' in the context of the comparative method

In part, the present confusion as to what would constitute an optimal meaning list for use in the initial stages of the comparative method has arisen because, as the quotes given towards the end of the previous section show, there seems to be some disagreement as to what the *role* is of basic meanings in this context. Campbell and Poser (2008:166) focus on those meanings which are least likely to be represented by a form borrowed from another language, and which are therefore most resistant to replacement by borrowing. Indeed, the term 'culture-free' implies this definition, as it is generally accepted that words which have extensions which are more closely linked to culture-specific phenomena – in the domains of technology, religion, politics, and so on – are more frequently borrowed than those which have less culture-specific extensions (a long-held assumption, demonstrated statistically in Tadmor et al. 2010). Trask (2000:39), meanwhile, suggests that basic meanings equate to the most stable meanings: those that are least likely to be replaced by any means, either endogenous (where the source for the replacement is internal to the language, such as neologism or the semantic shift of another form) or exogenous (where the source of the replacement is external to the language, i.e. a borrowing from another language). Swadesh's definition (1972:32) of basic vocabulary as being those meanings which are replaced in terms of form at the most constant rate in a

language, while a controversial claim, echoes the reasons that originally motivated the construction of his lists: for use in lexicostatistics and glottochronological dating of language splits (to be discussed in more detail below in section 2.3). Indeed, a further consequence of Swadesh's original aim of providing a test list for lexicostatistical and glottochronological work is that, after eliminating all the semantic areas he considered to be unrepresentative of meanings which undergo lexical replacement at a constant rate, "there remained only the vocabulary that is called 'basic', or that of universal and simple things, qualities, and activities, which depend to the least degree possible on the particular environment and cultural state of a group" (1972:275), showing a further assumption that these basic meanings represent universal and simple concepts, common to all languages. Despite Swadesh's original intentions, however, and as indicated by Lohr (1999:12), the term 'basic vocabulary' is generally taken to refer to a Swadesh or Swadesh-type list, whether its application is in lexicostatistics or glottochronology, or in the initial stages of the comparative method.

Laying aside at present the issue of whether or not such meanings are represented by forms which are replaced at a constant rate (a controversial assumption which will be returned to throughout this dissertation), how stable, borrowable, conceptually simple, and universal the data used in the initial stages of language comparison are will have an effect on how successful and reliable such comparison will be. Our definition of a 'basic' meaning list will thus focus on these four criteria. It is worth here entering a discussion of these terms; we will concentrate particularly on the role played by each of these features of an item in comparative linguistics and the facilitation of the comparative method, in order to ensure absolute clarity of the concepts on which our search for an optimal meaning list for use in the initial stages of language comparison will be based.

We have already introduced the idea of the **stability** of a meaning above, defining an item's stability as how likely the form by which it is represented is to be replaced, from either endogenous or exogenous sources. If a meaning is more stable, then the form by which it is represented is less likely to be replaced than a meaning that is less stable. For example, in the history of English, the meanings 'cow' and 'ox' have displayed stability, in that the forms representing these meanings (*cow* and *ox*) have not been replaced since PIE – they have been

inherited directly from the PIE forms of the same meaning, *$gw\bar{o}us$ 'cow' and *$uk(w)s\bar{e}n$ 'ox'.[7] The meaning 'squirrel', on the other hand, has undergone at least one replacement of form in the history of English since PIE (the earlier inherited form PIE *$werwer$- > Old English $\bar{a}c$-$weorna$ was replaced in the 14th century by a borrowing from Anglo-French *esquirel*, ultimately from the Greek neologism *skíouros* 'squirrel', literally 'shadow-tailed'). The meaning 'squirrel' has thus been less stable in the history of English than the meanings 'cow' or 'ox'. While this definition of 'meaning stability' may at first seem counter-intuitive – after all, it is the form representing the meaning which is being replaced, rather than any change in the sphere or scope of the meaning itself – it is important to note that it is the underlying meaning, and the likelihood inherent to this meaning that the form by which it is represented will be replaced, with which we are concerned.

'Stability' can also be used to refer to other lexical and semantic tendencies in diachronic linguistics, and we must be sure to distinguish these uses carefully. On the lexical level, phonological stability refers to the stability of the pronunciation of a form; a form is considered phonologically stable if it has undergone comparatively little phonological change in its history. For example, English *two* has displayed more phonological stability in its history than has the Armenian word for 'two', *erku*, the former being more similar phonologically to the PIE *$dwéh_3(u)$ 'two', from which both forms are inherited. Semantic stability also operates on the lexical level: a form in a language is semantically stable if it has shifted semantically comparatively little in its history. For example, English *five* < PIE *$pénkwe$ 'five' has displayed more semantic stability in its history than has English *dizzy* < PIE *$dhwésmi$ 'breathe'.

On the semantic level operate meaning or item stability (the stability with which we will be primarily concerned in this paper, and the definition of which is given above), and conceptual stability. Influenced more by the socio-cultural than the linguistic situation, conceptual stability is the degree to which a concept, and hence the form used to refer to it, is retained in a culture – for example, the dropping out of use of a technological concept when the technology itself becomes obsolete ('eye', for example, is a meaning with higher level of conceptual stability

[7] The PIE reconstructions used throughout these sections are those of Mallory and Adams (2006).

than the meaning 'hygrodeik', an obsolete instrument once used to measure air humidity). Concepts themselves can be borrowed, and it is important to distinguish between borrowing of form and conceptual borrowing, although the two frequently operate hand-in-hand – the conceptual borrowing of ecclesiastical concepts during the Early Middle Ages in the British Isles and the parallel borrowing of the forms used to refer to them from Latin and Greek into Old English (such as Old English *apostol* 'apostle' < Latin *apostolus*, Old English *créda* 'creed' < Latin *crēdo* 'I believe') are a prime example of this.

Unless otherwise qualified, our definition of stability for the purposes of this paper will be that of meaning stability, the likelihood of a meaning to undergo replacement of form from either endogenous or exogenous sources. Owing to the constant flux of natural language, stability is unlikely to be constant cross-linguistically; an item that is more stable in one language may be less so in other languages. However, studies such as those by Pagel et al. (2007), in which word frequency is shown to correlate with item stability, as well as the results to be presented in this dissertation, suggest that there are tendencies for some items to be more stable cross-linguistically. Any basic meaning list for use in the initial stages of language comparison will thus utilise these tendencies in order to compose a list that maximally represents the most cross-linguistically stable items. The rationale behind the comparison of such items is clear: when looking for initial evidence of genetic relationship, it is preferable to maximise the likelihood that such evidence will be present. As more stable items are less likely to have been replaced in terms of form in the languages' histories, such meanings are more likely to retain inherited forms.

The **borrowability** of an item refers to the likelihood that the form representing the meaning in question will be replaced by an exogenous form, i.e. by a form borrowed from another language. The form representing an item with a high borrowability is more likely to be borrowed; the form representing the item is therefore also more likely to be replaced by the mechanism of borrowing, presuming there is little semantic shift in the borrowing process. A list comprising items which are maximally resistant to borrowing would again be beneficial to our comparative needs; as discussed above, while the demonstration of regular sound correspondences will help eliminate those loans which were borrowed after the

relevant shift, loanwords borrowed before a phonological shift takes place will display the same regular correspondences as inherited forms. Long considered to be the case intuitively by comparative linguists (see Campbell and Poser 2008:1-86 for a history), Tadmor et al. (2010) have statistically shown that semantic spheres exist which are cross-linguistically less likely to be borrowed or replaced by borrowing: those spheres which are less cultural and more universal, such as the semantic fields of sense perception and spatial relations, tend to have lower borrowability rates than other semantic spheres.

Some linguists, notably Dixon (1997), have argued that there is no cross-linguistic tendency for there to be an area of the vocabulary which is less borrowable. He supports his argument with evidence from his own research on Australian languages, in which: "...similar percentages of shared vocabulary are obtained by comparing 100 or 200 or 400 or 2,000 lexemes[8], from adjacent languages" (1997:10), citing Breen (1990:54) in support. In response to Dixon's claim, Bowern et al. (2011) carried out a large-scale study with data for 204-item culturally-appropriate lists from 122 languages spoken by hunter-gatherer/small-scale cultivator societies in California and the Great Basin, Amazonia, and Australia.[9] They demonstrate that the proportion of loans in the languages considered (mean=5.06%) is not only not nearly as high as Dixon's prediction of 50% shared vocabulary between contiguous Australian languages (1997:26-7), but is also considerably lower than that of the languages sampled in World Loanword Database (mean=10.24% of items borrowed in an equivalent list[10]; Haspelmath and Tadmor 2009). Furthermore,

[8] These lists presumably represent larger or smaller sets of 'core' meanings. If it were simply a random sample of 100, 200, 400, and 2 000 items from the languages' lexica, we would of course expect to see little variation in the amount of shared vocabulary.

[9] Hunter-gatherer societies are implicitly typically in equilibrium in Dixon's punctuated equilibrium model, the state in which he argues languages borrow at a much higher rate throughout the vocabulary – the introduction of agriculture is listed as a 'punctuation' event (1997:77), and "long periods of equilibrium" (1997:4) are the norm for most of human history.

[10] Issues regarding the potential bias of the language sample in World Loanword Database, to be discussed in section 2.4.2, must be noted; furthermore, the list used by Bowern et al. (based on that used by Greenhill et al. 2008) was tailored specifically for languages spoken by hunter-gatherer societies and the areas under consideration, and thus is not ideal for cross-linguistic comparison (2011:7). Some of the meanings are certainly restricted culturally, for example item 108, 'mosquito', or linguistically, for example item 188, 'we (exclusive)'. However, the result still rejects the hypothesis that speakers of hunter-gatherer languages borrow vocabulary at a much higher rate into their basic vocabularies than the cross-linguistic average.

they argue that the Australian language groups on which claims for radical levels of borrowing have been based, those in the Victoria River District and the Yolngu languages of Eastern Arnhem Land, while showing higher levels of borrowing in their study, are "atypical of the Australian languages in our large sample" (2011:4). The research by Bowern et al., in conjunction with that of Tadmor et al. (2010) discussed above, indicates that we can speak meaningfully about those items which are cross-linguistically less likely to be lexically replaced by borrowing; our search for a meaning list representative of such items for use in the initial stages of language comparison is thus not a futile one.[11]

The extent to which an item's stability overlaps with its borrowability is one of the main issues to be teased apart in the course of this dissertation. In order to enter a meaningful discussion, it is of critical importance that these concepts are not conflated. Often in the literature, for example, it is assumed that an item's stability and its borrowability are one and the same: see, for example, Wang and Wang (2004), whose summary of research by Chen (1996) belies a conflation of stability and borrowability of a meaning in the two different definitions of Chen's low rank meanings (the 100 meanings of Swadeshs 200-item list not included in Swadesh 1955). In their first definition of the 'low rank', focus is on the feature of borrowability: "the low rank tends to be influenced by more frequent borrowing'; we are then given a definition that focusses on stability more generally: "vocabulary replacement [in the low rank] occurs at a greater rate than for high-ranked words" (2004:644). These definitions are not necessarily in conflict; however, if a productive analysis of the relationship between item stability and item borrowability (and thus, more generally, the optimality of a meaning for a standardised meaning list) is to be entered, it is important to carefully distinguish the two features. In sections 2.4.3 and 2.4.4, we will give further examples of the conflation of item stability and item borrowability in research by Dolgopolsky (1986) and Lohr (1999). While stability refers to the likelihood of the form representing a meaning being replaced by an endogenous *or* exogenous form, borrowability refers to the likelihood of the form representing a meaning being replaced *only* by an exogenous form; stability

[11] It should be noted, however, that no items have been demonstrated to be completely unborrowable or resistant to replacement by borrowing (Campbell and Poser 2008:166; Embleton 1986:67; McMahon and McMahon 2005:90-91; Starostin 2009:160-1; Swadesh 1950:157); counter-examples of suggestions of such items can always be found.

and borrowability are therefore two separate features of an item's retentiveness. Indeed, we will, in the course of this dissertation, have the opportunity to explore the validity of the assumption that item stability and borrowability are significantly related; as will be demonstrated, the frequent conflation is unwarranted.

In the determination of a standard basic meaning list for use in the initial stages of language comparison, the **simplicity** of the concepts on a basic meaning list will be considered, following Swadesh (1972:275). Those concepts which are more complex, such as 'mother-in-law' or 'day after tomorrow', tend to be represented by compounds or phrases, and thus tend to be borrowed less. These items' low borrowability, however, is not to do with the intrinsic low likelihood of a form representing the meaning to be replaced by borrowing, as with the low borrowability of more simple concepts such as 'fire' or 'water'; rather, it is due to the lexical form a representation of the meaning will tend to take, itself reflective of the complex nature of the meaning (Tadmor et al. 2010:236). In order to avoid these items of low borrowability influencing our results, more complex concepts should be avoided. Furthermore, while such constructions may have been inherited, the transparency of such forms usually suggests much more recent development. As our goal is to increase the frequency with which inherited forms are compared using a basic meaning list, only those items which are maximally conceptually simple will be considered as candidates. Insistence on maximal conceptual simplicity of meanings will thereby supplement the identification of those meanings which are maximally resistant to borrowing and maximally stable.

Finally, the **universality** of the items must be considered. Again following Swadesh, who recognised the need for maximally universal items to enable cross-linguistic applicability of a standard list (1972:275), this criterion for an optimal meaning list will be explored carefully throughout this paper. It has been suggested by some, notably Hoijer (1956), that universality in a meaning list is impossible; Hoijer argues that, as all cultures, and thus languages, draw their semantic categories along different lines, the meanings of the forms for one language will never correspond in a one-to-one manner with the meanings of the forms for another. However, the implication of this view, noted by Cowan, should be taken into account: "In its extreme consequence it would mean that *no* two comparative lists are ever valid, whatever their composition and however large the number

of agreements, and this would unsettle the basis for *all* comparative linguistics" (1959:237-8; emphasis in original). This is clearly undesirable. Furthermore, similar to the criteria of stability and borrowability, while it may not be possible for any one list to be completely universal, it cannot be denied that there are concepts which tend to be more universal to the human experience than others – indeed, this has been demonstrated statistically by Tadmor et al. (2010:234-6), whose scoring of items for their representation in a cross-linguistic study of 41 languages showed that, while some items were represented in all the languages studied (they give as examples 'where', 'which', 'there', 'rise', 'stand'), others, such as 'netbag' and 'tumpline', are found in only a few languages. Finally, it should be noted that Hoijer's criticism is of the basic meaning list as used in lexicostatistics, in which forms are compared with only one corresponding form in another language, resulting in a binary 'cognate'/'non-cognate' decision to allow for quantitative analysis (see section 2.3 for a more detailed discussion of lexicostatistics); the use of a basic meaning list in the more qualitative initial stages of language comparison will be more flexible, allowing for the comparison of more than one form per item. The pursuit of those items which tend to be more universal will thus inform our discussion of the optimal list for use in the initial stages of language comparison.

We identified in the previous section that our priority in curtailing the unnecessary subjectivity rife in the establishment and evaluation of hypotheses of genetic relationship between languages should be the determination of an empirically defined meaning list, derived from cross-linguistic data, designed to maximally facilitate the implementation of the initial stages of the comparative method. Such a meaning list would allow for a more meaningful discussion of the other areas of subjectivity in the initial stages of the application of the comparative method. The focus of this dissertation will therefore be on determining what an optimal list of basic meanings for use in the initial stages of the comparative method would comprise.

In this section, we have defined a 'basic' meaning as those meanings which, cross-linguistically, can be said to be not only maximally resistant to borrowing or replacement by borrowing, in order to reduce the likelihood of confounding factors entering the data compared, but also maximally stable, maximally conceptually simple, and maximally universal, in order to increase as far as possible the likelihood

of comparing forms which are likely to display features inherited from a common ancestor, where they exist. After a brief excursion into the need for a distinction between basic meanings as used in the initial stages of language comparison and those used in lexicostatistics and glottochronology, we will embark on an evaluation of the literature which has, to date, explored those issues most pertinent to the present discussion. Through this evaluation, we will determine the optimality of the standardised meaning lists which are currently widely used or recommended for use in the initial stages of the comparative method. This will also enable us to further examine our criteria of stability, borrowability, simplicity, and universality, the extent to which they overlap, and the relative importance of our criteria; furthermore, we will demonstrate the importance of empirical methodology, objectively applied, and the cross-linguistic applicability of results. We will conclude that none of the lists established thus far are optimal for use in the initial stages of the comparative method, either because of flaws in the methodologies used to draw them up, or because of flaws in the methodologies' application.

2.3 Basic meaning lists in lexicostatistics and glottochronology

Before we continue with our investigation of the role played by meaning lists in the initial stages of language comparison, an important distinction must be made between the use of basic meanings in the comparative method and the use of basic meanings in lexicostatistical and glottochronological research. Developed by Swadesh (1950, 1952, 1955), lexicostatistical methods use the percentage of cognate forms between two languages on a standard meaning list to determine the degree to which they are related – the application of the comparative method is thus required before relationships between languages are assessed lexicostatistically, in order to determine which items on the list are indeed cognate. Glottochronological methods attempt to date the separation of two languages from their common ancestor using calculations based on the same percentage of cognates and a hypothesised constant or predictable rate of change.

While the distinction between lexicostatistics and glottochronology is often blurred in the literature, even by Swadesh himself (note, for example, the title

of his 1952 paper 'Lexico-Statistic Dating of Prehistoric Ethnic Contacts'; cf. Campbell 1998:177, Fox 1995:279-80), it is important that a distinction is made, particularly as the hypothesised constant or predictable rate of change necessary for glottochronological work remains so controversial (see, for example, Bergsland and Vogt 1962; Heggarty 2010; McMahon and McMahon 2006). Embleton argues particularly strongly for the precise definition of the difference between lexicostatistics and glottochronology, pointing out that: "This [conflation] leads to confusion – and often debate at cross-purposes – over the goals or achievements of the particular method used in a particular piece of research" (2000:160 n.3); McMahon and McMahon also warn us of the potential consequences of confusion of terminology: "Identifying lexicostatistics with glottochronology, and rejecting both on the basis of problems specific to the latter, runs the risk of throwing out the meaning-list baby with the time-depth bathwater" (2005:34).

It is the basic meaning lists used in lexicostatistical and glottochronological work that constitute our first encounter with a standardised list – particularly, the Swadesh 100- and 200-item lists, to be discussed in detail in section 2.4.1. The Swadesh lists are used most frequently by far for lexicostatistical and glottochronological purposes. Indeed, as indicated above, it is the items on these lists that often constitute the concept of a basic meaning list in many scholars' minds, regardless of whether the basic meaning list is for use in the comparative method, or for use in lexicostatistical or glottochronological work. There are many issues, however, with Swadesh's lists, as we shall see below, especially regarding the levels of empiricism and objectivity with which they were drawn up, which render the conclusion that they are the optimal lists for *any* purposes, comparative or quantitative, questionable.

Lohr (1999:53) stresses the need for emphasis on different characteristics in lists for use in the initial stages of language comparison compared with those for use in lexicostatistics and glottochronology. Stability, borrowability, simplicity, and universality are important features that must be taken into consideration when determining what would constitute an optimal list for use in either the comparative method or lexicostatistical/glottochronological studies. However, she suggests that a list in which universality is emphasised, and which comprises meanings with a relatively similar cross-linguistic rate of retention, would be more appropriate

for use in lexicostatistics; a list which perhaps sacrifices some of the universality of its items, however, for a greater overall cross-linguistic stability, will be more appropriate for use in language comparison, in order to maximise the likelihood of finding evidence for genetic relationship (where it exists). Indeed, a list which is focussed secondarily yet still heavily on universality should have enough meanings that are cross-linguistic to allow the meaning list to be considered a standard one; furthermore, in the initial stages of the comparative method, if an item is not present or does not overlap in a one-to-one manner, there is flexibility to allow comparison of items to which the original form may have conceivably shifted semantically.

The distinction thus made between lists for use in language comparison and those for use in lexicostatistical and glottochronological work, a brief consideration of the glottochronological constant will be entered here. The supposition of a constant rate of change will be returned to throughout this dissertation; indeed, in the course of this research, we will find ourselves in the fortunate position of being able to statistically explore the validity of this hypothesis.

Lees (1953) and Swadesh (1955) suggest, from a study of those languages with long written records (English, Spanish, French, German, Egyptian, Athenian Greek, Cypriot Greek, Mandarin Chinese, Swedish, Italian, Portuguese, Romanian, and Catalan), that r, the rate of lexical retention, is around 86% in Swadesh's 100-item list – in other words, that the forms for 14% of the meanings are replaced over a thousand-year timespan – and is around 80% in Swadesh's 200-item list. These figures are used in glottochronological calculations, which attempt to determine the absolute time depth separating two languages. However, Bergsland and Vogt (1962) demonstrate, in their classic study of the histories of Icelandic, Georgian, and Armenian, that the rate of change is clearly not cross-linguistically constant at these values, calculating for the 100-item list $r = 96\%$ in Icelandic, $r = 97.2$-95% in Georgian, and $r = 97.8$-94.6% in Armenian. Starostin (2000) attempts to explain these discrepancies by removing borrowings from consideration as a means of replacement of form, and by taking into account the age of the individual forms, as well as suggesting an individual constant rate of change for each meaning – thus arguing that the rate of change in basic meaning lists is predictable, rather than constant. Indeed, Kruskal et al. (1973) take as one of their fundamental

assumptions in their maximum-likelihood model calculating retention rates not only that the replacement rate varies for each meaning, but that it also varies between language families; Embleton (1986:74) further suggests that the constants change not only between different meanings and different language families, but even between different languages. Finally, while Lohr (1999:28) recognises this fluctuation in r between languages, she suggests that the clustering distribution of calculated retention rates at around 86% rather than an even spread across the values indicates a tendency for r to be roughly predictable cross-linguistically.

However, these suggested adjustments to the glottochronological calculation potentially betray the futility of the endeavour. As Heggarty argues:

> ...these attempts all still see departures from the assumed basic rate as disruptions to be explained away, betraying a continuing faith that the rate of decay 'should' be underlyingly constant in the long run...The point is that we have no reason to assume a priori that they [languages] should [diverge at a constant rate] in any case, and good reason in principle positively to expect they will not.
>
> (2010:304)

For instance, the argument by Embleton that the rate of change varies from meaning to meaning and from language to language could be seen as a tacit admission that there is not in fact a predictable rate of change. If we follow her reasoning through, we very quickly encounter the eternal problem of how to define a 'language' – not only synchronically, but also diachronically. Should we, for example, consider Present Day English to be the same language as Old English? The two are certainly not mutually intelligible, the most frequently cited criterion in distinguishing languages from dialects, and yet the former has developed from the latter. Have the retention rates of different items also varied in the process? If retention rate varies not only from meaning to meaning, from language to language, but also through time as well, this in effect renders it unpredictable; a conclusion supported, interestingly, by Kruskal et al.'s attempts to calculate retention rates using a maximum-likelihood model, finding that the rates vary in different areas of a family tree, both synchronically (from language to language) and diachronically (1973:52).

Second, the number of languages on which the calculation of r is based is very low; necessarily, written records have been required to determine the precise number of lexical replacements diachronically. However, there are obviously problems with the selection of languages used to calculate this constant. First, there are far too few languages in which meaning retentiveness is traced to make any cross-linguistic generalisations. Furthermore, it has been suggested that literacy, necessary to enable the detailed determination of how many times a meaning has been replaced lexically in a certain time-frame, in fact has an effect – either accelerating or impeding – on meaning stability (e.g. Bergsland and Vogt 1962; O'Neil 1964). The results cannot, therefore, be applied cross-linguistically, where the vast majority of languages throughout time have not had a written tradition.

Indeed, this concern can be taken one step further still; while the development of literacy may well have an effect on meaning stability, the political and societal conditions that enable literacy to develop will undoubtedly do so to a much greater degree. Those societies which have a long written tradition have tended to be larger in terms of population, spread over greater geographical areas, more hierarchical, centralised, and trade directly and engage in political relations over greater distances than those societies which do not; while the development of literacy does not automatically arise out of these conditions, one only has to consider the languages used in Lees' (1953) attempt to calculate the rate of lexical retention and the cultures in which they were spoken to recognise the parallels. All of these factors will have a wide-ranging, complex effect on linguistic communication, even on the most basic vocabulary in a language. Therefore, any calculation of r based solely on data from languages spoken in these conditions simply cannot be generalised cross-linguistically.

The theory of a constant or predictable rate of lexical replacement in basic meanings, and hence glottochonological theory generally, is thus considered in this paper to be at best unproven, and at worst untenable. This conclusion will inform the construction of the hypothesis regarding the rate of change, which will be tested in section 4.3 using the results of the investigation presented in this dissertation; we will show conclusively that there is no statistical support for a constant rate of change.

But what are the implications of this conclusion for our investigation of cross-linguistic stability and borrowability? It could be argued that if the rate of change is not constant, the present study would be rendered futile. However, as discussed above, statistical studies have shown that while a predictable rate of change may not exist in basic meanings, there are tendencies for some items to be generally more stable or less likely to be replaced by borrowing cross-linguistically (respectively, Pagel et al. 2007; Tadmor et al. 2010). We can therefore continue with confidence that, while the absolute rate of replacement may not be predictable, it is not completely random either; it is thus possible to talk meaningfully about those items which are the least liable to be replaced in terms of form by borrowing cross-linguistically.

2.4 Standard meaning lists and their methodologies to date

2.4.1 The Swadesh lists (Swadesh 1950, 1952, 1955, 1972)

While the importance of the comparison of basic meanings in studies regarding linguistic relationship was recognised as early as de Laet's insistence that comparanda should consist of basic vocabulary (1643; cited in Campbell and Poser 2008:17), it was not until the mid-twentieth century and the advent of lexicostatistics and glottochronology that a standard list comprising such meanings was formulated, by the comparativist Morris Swadesh (see Appendix A.1). Swadesh's aim in drawing up his lists, one of 200 items, and one of 100 items, was not, as has often been thought, to produce a list of those items which are the most stable and the most resistant to borrowing (see, e.g. Dixon 1997:10; Embleton 1986:43; Fox 1995:282; Heggarty 2010:308); his explicit aim was to define the area of meanings which cross-linguistically represent a 'lexical chronometer' – in other words, to define the area of meanings which are replaced at a cross-linguistically stable rate, for use as an 'index of time' in glottochronological calculations, whilst being maximally universally applicable (Swadesh 1972:274-277). Laying aside the issues presented above regarding the assumption of a constant rate of change in languages' basic meanings, the Swadesh lists have now come to be most commonly identified as

being representative of the meanings which are most stable and most resistant to borrowing cross-linguistically. Their applications in historical linguistics are wide-ranging, from the lexicostatistical and glottochronological work for which they were designed and use in Oswalt's shift tests (1970, 1991) in determining the factor of chance in the explanation for similarities observed between languages, to studies attempting to distinguish potential cognates from potential loanwords in language comparison (McMahon et al. 2005; Wang and Wang 2004). Notably for this discussion, the 100-item Swadesh list has also been used as the basis for subdivision using the criterion of stability, i.e. in attempts to identify the most stable meanings (Holman et al. 2008; Starostin 2000).

The most important use of the Swadesh lists for our purposes, however, is their potential implementation in the initial stages of the comparative method. While their use is rarely explicitly advocated in introductory textbooks on the subject of historical and comparative linguistics, the general trend instead being simply to suggest that the use of 'basic' or 'core vocabulary' is preferable in the initial stages of comparison (Campbell 1998; Fox 1995; Hock and Joseph 2009), the definition by Lohr given above for what this would comprise is telling and warrants repetition: "[basic vocabulary is] usually taken to imply a set of words similar to those on Swadesh's lexicostatistical lists..." (1999:12). Antilla's suggestion further identifies basic meanings with the Swadesh list: "One way to guard against borrowing is to start the comparative method with vocabulary items that come from semantic spheres not usually borrowed from, that is, basic noncultural vocabulary...the so-called Swadesh list is a handy starting point" (1989:231). That the definition of basic meanings is inextricably linked with the Swadesh lists in many linguists' minds is clear, and indicates that a careful evaluation is required of the level of empiricism underlying the Swadesh lists, as well as the representation of item stability, borrowability, simplicity, and universality, in order to determine whether they represent optimal lists for use in the initial stages of the comparative method.

Originally, in search of an area of meanings considered to be replaced lexically at a constant rate, Swadesh composed a list of 215 items, narrowed to 200 items because some of the items are "unsatisfactory for many language groups" (1952:457). This list was later reduced to 100 items, on the basis that many of the items on the original list had proved to be too cultural, for example 'snow', 'ice', 'sea', 'salt'

(1955:125); or because the semantic divisions varied too much cross-linguistically and thus were not universally applicable – for example, 'wife' was removed because of potential synonymy in some languages with 'woman', or 'river', 'lake', and 'sea' was removed because of potential synonymy with 'water' (1955:125). Items such as 'leg' and 'back' were also removed because their "reference is fairly inclusive and shades with that of other words" in some languages (1955:125); in other words, because of the potential for the form to overlap with other meanings. Today, either the 100-item or 200-item list is implemented in the various domains of historical linguistics, as described above. As neither one is explicitly recommended for use in the initial stages of language comparison, and as the reasons and methods underlying their formulation are the same, both will be evaluated as one for our purposes.

Precisely how these items were chosen is not made clear by Swadesh; we are informed only that, in the creation of the original 200-item list: "It was not difficult to form a list of about two hundred relatively stable lexical items, consisting of body parts, numerals, certain objects of nature, simple universal activities" (1952:455). This lack of transparency is concerning; the creation of the 100-item list by refining the 200-item list is equally so. Hoijer (1956:52), for example, discusses the subjective way in which some items were removed because of the potential for representation by bound morphemes in some languages, or owing to the possibility of duplication (such as 'and', 'because'; and 'wife', 'river', 'lake', 'sea', respectively), while other items which could also potentially be described thus were retained – he indicates, for example, the items in the 100-item list 'who', 'all', 'not' as being potentially represented by bound morphemes, and 'who'∼'what', 'that'∼'this' and 'skin'∼'bark' as potentially being represented by the same root in some languages. Hoijer also draws attention to the issue of the subjective definition of what constitutes a meaning that is likely to be represented by an onomatopoeic form: "there is no way of determining in advance which of the test items will be expressed so frequently by sound-imitative terms as to require elimination" (1956:52).

Indeed, this undefined, unempirical means of identifying the meanings, referred to by Lohr as "intuitive" (1999:53) and Embleton as "problematic" (1986:43), is a fundamental criticism of the Swadesh lists. The lack of an explicit methodology in Swadesh's formulation of the 200-item list means we cannot evaluate the procedures

used in its creation scientifically, nor can we attempt to reproduce the lists by his methods to verify his results; the same criticisms can be made of the subjectivity displayed in the formulation of the 100-item list. As discussed above, the curtailment of subjectivity in the initial stages of language comparison is one of the aims of this paper; any list, therefore, that does not have explicitly defined stages or parameters in its composition must be automatically rejected. That is not to say that the Swadesh lists do not represent those items which are maximally stable, maximally resistant to replacement of form by borrowing, maximally simple, and maximally universal; rather, it means we cannot accept the items presented as such until they have been demonstrated independently by explicit, empirical methodology.

It is clear from Swadesh's description of the means by which he identified those meanings likely to be replaced in terms of form at a constant rate that Swadesh considered items' borrowability and stability to be inextricably linked with their rate of change. Regarding borrowability, he states: "The reason culture terms have to be avoided is that their retention or loss is too closely correlated with fluctuations in the cultural situation to serve as an index of the passage of time" (1955:124), demonstrating that he considers the items least affected by the cultural situation – the least borrowable items – to best represent items subject to lexical replacement at a constant rate. His view of stability of items is similar: "This approach [glottochronological dating using the 'index of time' of basic vocabulary] should prove fairly dependable...because it is a well known fact that certain types of morphemes are relatively stable" (1950:157), again showing how he considered items with high stability to have the most constant rate of lexical replacement.

How this conclusion was reached is understandable. Items which are less subject to the erratic nature of borrowing, so dependent on socio-cultural factors, may intuitively be assumed to have a more stable rate of change than items which are more so (again, for present, laying aside the question of whether this rate of change is constant or predictable). Furthermore, the forms for items which have the highest stability are replaced least frequently; a slower rate of change may appear more constant, as there is less evidence of change. Finally, as discussed above, an item's stability is intuitively linked to its borrowability, in that it is assumed that an item which is replaced more frequently by borrowing will be less stable, and vice versa. However, as was noted, item stability and item borrowability should

not be conflated, as they are two separate features of an item's retentiveness. More importantly, however, the presence of these features is presumed to be a side-effect of identifying those meanings whose forms are replaced at the most constant rate – and the validity of this assumption is far from being demonstrated.

It thus simply cannot be taken at face value that the Swadesh lists, drawn up along non-explicit lines with the intention of representing the semantic sphere which displays the most constant rate of replacement of form, identifies those items which are maximally stable and maximally resistant to replacement by borrowing. The Swadesh lists are therefore not appropriate, according to our criteria of empiricism and representation of maximal stability and borrowability, for use in the initial stages of the comparative method.

This conclusion has many implications for the use of this list as representative of the most basic meanings in historical and, specifically, comparative linguistics, owing to the wide acceptance of the Swadesh lists as lists of those items which are maximally stable and maximally resistant to borrowing. This conclusion also automatically removes from consideration as optimal for use in the initial stages of the comparative method those lists, mentioned above, which take as their input the items on the Swadesh list, and proceed to rank them according to various criteria – namely, the ASJP list (Holman et al. 2008) and the Yakhontov list (Starostin 1991, 2000). As a result of these conclusions, we will therefore proceed to evaluate standardised meaning lists that have been drawn up independently of the Swadesh lists. In the words of Swadesh himself: "It is doubtless possible to devise a better test list than the present one..." (1952:457).

2.4.2 The Leipzig-Jakarta list (Tadmor, Haspelmath and Taylor 2010)

Noting the problem of the potential for borrowed words to enter the data in the initial stages of language comparison, Tadmor et al. (2010) use the data from the World Loanword Database (Haspelmath and Tadmor 2009; henceforth WOLD), itself a product of the Loanword Typology project, to draw up an empirical list of the least borrowable items cross-linguistically. WOLD consists of data from 41 languages, representative of 26 different language families spread across the globe. For each language, a specialist assessed a list of 1460 meanings, based on

the Intercontinental Dictionary Series, which was in turn based on Buck (1949), recording for each meaning how much evidence there is for it being represented by a borrowed form in that language, by assigning it one of five statuses on a scale from 'clearly borrowed', with a score of 0, to 'no evidence for borrowing', scored 1.00.[12] Each meaning was also assessed and scored in a similar manner for its age, on the premise that a form that has an earlier attestation or reconstruction is more likely to represent a meaning with a lower borrowability, and thus will help reduce the influence of unidentified loans; its representation throughout the languages, to avoid the borrowed score of those meanings that are only represented in very few languages being weighted equally with those that are attested in all the languages considered; and its analyzability, or the likelihood that a meaning is represented more frequently by complex words or compounds, such as the meanings 'younger sister' or 'the day after tomorrow', in order to account for the probability that the forms representing these meanings will not be borrowed not because of any intrinsic resistance to borrowing, but because of the increased likelihood that their representation will tend to reflect a use of language-internal mechanisms such as compounding. The scores given for each of these criteria were multiplied together to form a composite score, and the items considered were ranked accordingly to give the Leipzig-Jakarta list of basic vocabulary (Appendix A.2). Tadmor et al. report that verbs and adjectives are borrowed less frequently than nouns, that function words are borrowed less frequently than content words, and, supporting that intuition of linguists present since the earliest days of language comparison, that vocabulary representing less cultural concepts is borrowed less frequently than vocabulary representing more cultural concepts (2010:231-233).

The empiricism of the methodology used to formulate this list is admirable. While Tadmor et al. note that the languages considered are not a random sample, and are likely to be biased in favour of those languages which have higher levels of borrowing, owing to the likelihood that those experts engaged in research with such languages would have been more interested in volunteering their time for the project

[12] N.B. The scale runs in the opposite direction on the WOLD website (http://wold.livingsources.org/, accessed 21/6/12); that is, 'clearly borrowed' is given the highest score of 1.00.

(2010:229), WOLD represents the most cross-linguistic and statistically-based study of levels of borrowability to date.

Tadmor et al. suggest that the Leipzig-Jakarta list also takes into account the criteria of stability, simplicity, and universality of the items, and thus represents a list suitable for use in the initial stages of language comparison: "...it is a basic vocabulary list that takes into consideration the features normally associated with basic vocabulary in historical and comparative linguistics: resistance to borrowing (the borrowed score), universality (the representation score), simplicity (the analyzability score), and stability (the age score)" (2010:238). Borrowability is calculated in a manner suitable for our purposes, as is universality, to some extent, although the use of only 41 languages in the project, while these languages are drawn from a cross-cultural and cross-linguistic spread, means we should be wary of making any generalisations about item universality based on these data alone. The determination of how conceptually simple an item is rests on the assumption by Tadmor et al. that conceptual complexity correlates positively and significantly with the analyzability of an item's form, i.e. the more complex an item is conceptually, the more analyzable the form by which it is represented will tend to be, and *vice versa*. This assumption will be discussed in some detail in section 5; for now, however, we will accept its validity.

Their treatment of item stability, however, is not satisfactory for our purposes. The language experts were asked to give a score of 1.00 to those words which were attested or can be reconstructed to before 1000, 0.90 for those earlier than 1500, 0.80 for those earlier than 1800, 0.70 for those earlier than 1900, 0.60 for those earlier than 1950, and 0.50 for those earlier than 2007. As can be seen from the spread of date cut-offs, the detail with which item stability is recorded is not sufficient to accurately determine which items are most stable. We would expect a large proportion of the forms representing meanings which we were investigating as candidates for maximal cross-linguistic stability to be older than a millennium; for all forms older than this to be given equal weighting removes a lot of information necessary to calculate the items' relative stability. Indeed, the rationale behind the inclusion of a quantification of item stability in the formulation of the Leipzig-Jakarta list was not to include a complete representation of which items are most stable cross-linguistically, but to reduce the probability of undetected or misassigned

loanwords skewing the data by giving more weight to those forms which are older (2010:236-238).

That the Leipzig-Jakarta list does not represent a detailed picture of item stability is, in fact, tacitly admitted by Tadmor et al., in the comparison of the Leipzig-Jakarta list with Dolgopolsky's list (1986) and Lohr's List 1 (1999), two lists which, as will be discussed below, place more weight on the criterion of item stability: "It thus seems that some meanings may be subject to change (e.g., semantic change, or replacement by novel formations), but not so much subject to borrowing" (2010:242). It would, therefore, be preferable for a list for use in the initial stages of language comparison to give more focus to the stability of items than in Tadmor et al., in order to maximise the likelihood of the presence of evidence for genetic relationship, where it exists.

Nonetheless, borrowability, simplicity, and universality are very important criteria for our purposes, and the ability of the Leipzig-Jakarta list to represent these features should be taken into consideration. While the list itself is not appropriate for our purposes, owing to the lack of detailed information on item stability, the methods used in its formulation, as well as the borrowed, analyzability, and representation scores, will be very useful tools in determining a methodology to define an optimal list for use in the initial stages of the comparative method. If we were to develop a means of determining item stability in more detail than presented here, the results could be integrated with the scores for the other three features presented in Tadmor et al. (2010), in order to create an improved item composite score, fully reflective of all four of our criteria. We now turn to the evaluation of lists by Dolgopolsky (1986) and Lohr (1999), in which the criterion of item stability is considered in more detail.

2.4.3 The Dolgopolsky list (1986)

Dolgopolsky's list was drawn up for much the same reason that concerns us in this paper: to create an empirically defined meaning list for use in the "preliminary evaluation of the advisability of comparing certain languages" (1986:27); in other words, for use in the initial stages of the comparative method. Dolgopolsky's methodology and explicit definition of the parameters of his list clearly emphasise

item stability – "only those morphemes are considered that are characterized by a high degree of stability in any or all of the languages compared" (1986:30). However, in his reasoning for the creation of his list, Dolgopolsky demonstrates the conflation of stability and borrowability discussed above: he claims that, if similarities can be found between the forms representing the same semantic item in different languages in comparisons based on this list, and such forms can be shown to have a very low probability of independent innovation, then these forms must be cognate, owing to the very low likelihood of such forms having been borrowed (1986:30). Dolgopolsky's mistaken claims, however, about the justification for his list invalidate neither the methodology used to define it, nor the list itself *a priori*; we shall thus proceed to evaluate this list based solely on the methodology used to compose it, and the application of this methodology, both of which focus on item stability.

Noting the difficulties in defining a list of the most stable items cross-linguistically without the existence of written records for the vast majority of the world's languages, Dolgopolsky presents a novel approach to quantifying item stability. Each meaning considered is studied in many different language families. For a single meaning in a single language family, the number of different etymological roots is determined. If all the forms are cognate in all the languages of the language family, this suggests the form representing the meaning has not been replaced at all in the family's history; if there are forms from two different etymological roots, this suggests the form representing the meaning has been replaced at least once; if there are forms from n different etymological roots representing the same meaning in the same language family, there have been at least $n - 1$ replacements of form.

Dolgopolsky exemplifies this means of determining item stability with the meanings 'star' and 'lightning' in Romance. There is only one extant etymological root for the meaning 'star' in Romance, represented by, e.g., Latin *stella*, French *etoille*, Sardinian *àstru*, and so on, suggesting that this semantic item has not undergone any replacement of form since the break-up of the most recent common ancestor. However, there are four extant roots for the meaning 'lightning': 1) Romanian *fulger* < Latin *fulgur*; 2) Spanish *relampago*, Portuguese *relampago*, Catalan *llampec*, Italian *lampo*; 3) French *éclaire*; 4) Rhaeto-Romance *sajetta*. This suggests a minimum of three lexical replacements in the history of Romance

(1986:30-1). For each meaning under investigation, the minimum number of replacements of form throughout all of the language families is totalled, and the items are then ranked for stability; those with the lowest number of minimum replacements are ranked the most stable. The average number of years a form can be expected to be retained is calculated by dividing the number of observed replacements for a meaning across the language families by the total time depth of the languages looked at.

Dolgopolsky draws attention to the inability of this method to represent the absolute number of lexical replacements that have occurred in the history of a meaning in a language family. One replacement of form obscures any unattested replacements of form which may have occurred earlier in a language's history and which have not been inherited into any extant varieties. However, he suggests that the figures gained for each meaning across language families represent instead "the indices of the degree of propensity of a specific semantic value..." (1986:31). A further problem with the methodology is that of mis- or unidentified cognates, which will contribute to an item appearing more or less stable respectively. However, by not restricting any such investigation of stability to those languages with written records alone, this methodology not only allows for many more data to be considered in the determination of an item's stability, but also means that any information garnered will be far more cross-linguistic in nature. Using this methodology, with a few 'tweaks', Dolgopolsky composes his list of items which are maximally stable; the initial ranking of items' stability, as well as Dolgopolsky's adjusted list, are provided in Appendix A.3.

It is these 'tweaks', however, which cause the most concern in the evaluation of Dolgopolsky's list as an optimal list for use in the initial stages of language comparison. As has been stressed throughout this dissertation, an objective, empirically-defined initial list for comparison is paramount, if we are to reduce the level of subjectivity currently present in the comparative method. While the methodology presented above is empirical in nature, Dolgopolsky's selection of the data to which it is applied and his subsequent adjustment of the list drawn up is worryingly subjective.

First, and most concerning, is Dolgopolsky's selection of the language families looked at to determine item stability. He states that data from language families of

Australia, Africa, and the Americas were not included in the initial ranking, because "semantic data significant for our purposes here are lacking for these languages (e.g., terms for 'heart', 'louse', 'horn', and so on)" (1986:33). Not only is this not the case[13], but it is notable that all of the semantic items provided as examples of being "significant for our purposes here" are on the initial list, and two are on Dolgopolsky's adjusted list. This *a priori* elimination of data, apparently because they do not adhere to expected or desired results, undermines Dolgopolsky's list in terms of our criterion of objectivity.

This subjective selection of data is compounded in the second stage of the formulation of Dolgopolsky's list; his 'adjusted' list provides his final ranking (1986:34-5). At this stage, Dolgopolsky *does* decide to use data from Australian and American language families. For example, the items 'new moon', ranked tenth on his initial list, and 'winter', apparently not ranked but discussed anyway, are removed, on the grounds that "they are lacking in many languages of the world"; the numbers '3' to '100' because "they do not have specific roots in many Australian languages"; and the number '1', as "its stability is rather weak in many Australian languages" (1986:34). One suspects that, had these languages been considered in the initial formulation of the list, the original ranking would have reflected these issues. More seriously, however, once one decides not to include data, for whatever reason, their later inclusion to adjust an original result not only undermines that original decision, but also throws doubt on the objective intentions of the whole enterprise, in that it will appear to have been manipulated to adhere to the researcher's expectations. Dolgopolsky's decision to pick and choose which data influence his item stability ranking is a serious enough transgression to render his whole list very questionable indeed.

Second, and again in the adjustment of the initial ranking, is the issue of subjective decisions in the removal of some items on the grounds of reduplication,

[13] But a few counter-examples include: Mangaray *ḋulu* 'heart', *gudaṙu* 'horn', *magery* 'louse' (Merlon 1982:232-3) and Wailibri *mantulka* 'heart' (Reece 1970:43) in Australian languages; Maidu *hòní* 'heart', *pedési* 'louse' (Shipley 1963:230, 285) and Nisean *hon*, *honni* 'heart', *di* 'louse' (Uldall and Shipley 1966:261, 265) in American languages; and Ila *mozo* 'heart', *lwiya* 'horn', *injina* 'louse' (Smith 1964 [1907]:304, 305, 314), and Nkore-Kiga *omutima* 'heart', *eihembe* 'horn', *enda* 'louse' (Taylor 1985:230) in African languages. It should also be noted that all three meanings are fully represented in WOLD (Haspelmath and Tadmor 2009).

those semantic items which are commonly represented by a form derived from the same root as a form representing another semantic item; this avoidance of reduplication is necessary, Dolgopolsky states, to allow for the accurate calculation of the probability of a form having arisen, fundamental to his determination of the likelihood of genetic relationship between languages. As examples, Dolgopolsky suggests 'we', as the form is often derived from the form used to represent the item 'I'; 'nit' and 'louse' are often represented by forms derived from the same root; and '20' is potentially derived from the form for the meaning '2'; 'we', 'nit', and '20' are thus removed from the list. 'Who' and 'what' are also considered to be represented by the same form in enough languages to warrant their consideration as a single semantic item, as are the prohibitive and verbal negative forms (1986:34). First, it should be noted that reduplication of results, while having an effect on calculations of the probability that two languages are related, should not be an issue in Dolgopolsky's calculations of stability, as what is being investigated is the individual stability of each individual item; it should not, therefore, matter that some of the items are represented by forms derived from the same root. However, what is most concerning about this adjustment of Dolgopolsky's original list is that no definition or set number of observances of what would constitute a potential reduplication is given, nor are examples provided – we are told simply: "...those semantic values that are often represented by derivative from roots which express or represent other semantic values also present in this list must be eliminated from the above inventory" (1986:34). We therefore simply have to place our trust in Dolgopolsky's decisions, rendering his selections unevaluable in terms of objectivity, and certainly unreplicable, and thus unscientific.

It should again be noted that the major criticisms of Dolgopolsky's list are not of the methodology itself, but of his choice of data for and subjective adjustment of the list drawn up using the methodology. While there are some issues with the methodology itself – the inability to give an absolute figure for item stability, unlike those lists based on languages with written records, or the problem of mis- or unidentified cognates within a language family – it is able to take into account a wide cross-linguistic spread of data in the determination of item stability. However, while the methodology is worth further consideration, Dolgopolsky's subjective application of this methodology to the data renders his list unscientific; it must

therefore be avoided for use in the initial stages of the comparative method, if our goal is to reduce the subjectivity with which it is applied.

2.4.4 The Lohr lists (1999)

Expressing similar concerns to those discussed throughout this paper regarding the need for a less intuitive basic meaning list methodology, the lists presented by Lohr (1999) are an attempt to empirically determine those meanings which are maximally stable and universal.

Lohr stresses that different criteria should be emphasised for lists to be used in different ways. First, she suggests a list of items which are maximally stable would be optimal for determining genetic relationship, in order to maximise the likelihood of inherited similarities (where they exist) entering the data. In another example of the conflation between item stability and borrowability discussed in section 2.2, she proposes that maximal stability in a list will also reduce the number of confounding factors in the form of borrowings entering the data. Second, she suggests a list of items with *similar* levels of stability (but not necessarily a particularly high level) for use in lexicostatistical studies, to avoid skewing effects of different levels of retentiveness on results at greater time depths. Finally, she suggests a list avoiding forms which are similar due to borrowing in lists for use in shift tests such as those of Oswalt (1970, 1991) – in these tests, forms from a standard list for two languages are compared for phonetic similarity with the semantics matched, then shifted so that the semantics are mismatched to gain a background chance score of similarity, against which the initial matched score is compared in order to determine whether the similarities observed when the semantics are matched are at a level above that of chance. The preference for the minimalisation of the number of borrowed forms in a standardised list for use in a shift test is due to the inability of the method to distinguish inherited similarities from borrowed similarities. She thus draws up two lists, one more representative of universality, the other more representative of item stability, and then cross-references them, in order to test her hypotheses that the factors discussed should be given more consideration in the various implementations of a basic meaning list (1999:53-84).

Lohr constructs her first preliminary list by assembling those items which have reconstructed forms in at least two of four proto-languages – Proto-Indo-European (Buck 1949), Proto-Afro-Asiatic (Ehret 1995), Proto-Austronesian (Zorc 1995), and Proto-Sino-Tibetan (Luce 1981) – on the grounds: "Such meanings are likely to be relatively basic, universal, and [conceptually] stable, since they reflect cultures of several millennia ago, cross at least two cultures, and were able to be reconstructed from descendant languages" (1999:54). For interest's sake, she also includes any items from the 200-item Swadesh list and Greenberg's Amerind list (Greenberg 1987) that were not included in her results.

Lohr's second preliminary list focusses more on item stability. The lexical retentiveness of all the items on the previous list were calculated in Indo-European by tracing each item onomasiologically, determining by how many different forms they have been recorded to have been represented, in Buck's etymological Indo-European dictionary (1949). The total time-depth for all the languages looked at – 31.3 millennia – was divided by the total number of visible replacements for each meaning, in order to obtain the mean number of years each meaning can be expected to retain its form before it is replaced.

Essentially, this procedure is the same as that used by Dolgopolsky to determine item stability; however, the information used to calculate stability in Lohr (1999) is limited to one language family. The information available for this language family, however, is detailed enough, owing to the long written traditions of many of the Indo-European languages, to allow earlier recorded replacements of form, which may subsequently have been replaced again, to be included in the calculation of an item's stability. Nonetheless, not all replacements will have been recorded: some of the languages in Buck (1949) have had a comparatively short written tradition, such as Lithuanian, and none of the languages, of course, have a written tradition stretching back to Proto-Indo-European. Lohr, therefore, like Dolgopolsky, notes that not all replacements may be visible, stating: "The number of visible replacements of each meaning was counted...a visible replacement might conceal a large number of intervening replacements" (1999:58). While some earlier replacements are visible with the detailed etymological information provided in Buck's dictionary, a full picture of item stability cannot be presented for this reason.

Lohr cross-references the two preliminary lists, to create two lists with different emphases – one emphasising stability, the other universality – while still permitting the less important criterion of each to have some influence (see Appendix A.4). For her first list, List 1, she cross-references all of the meanings collected in the first preliminary stage with only those items which she calculates as having had an average retentiveness of 10 000 years or more in Indo-European; List 1 thus emphasises stability, sacrificing the criterion of maximal universality. For List 2, she cross-references only those items that she found to be reconstructed in three or more proto-languages with those items calculated to have had an average retentiveness of 5 000 years or more in Indo-European; List 2 thus emphasises universality of meaning, sacrificing maximal stability.

To test her lists, Lohr uses each of them as the input basic meaning list for lexicostatistical classification, and in a replication of Oswalt's shift test (1970, 1991). She found that results that agreed better with expert classification were attained using List 2, the list more representative of item universality, as an input list for lexicostatistics, and List 1, more representative of item stability, for the shift test. Unfortunately, no attempt is made, presumably owing to the lack of resources, to test her third suggestion, that of a maximally stable list being optimal for use in the initial stages of the comparative method.

Lohr's objective and transparent application of the methodology are admirable; however, it is the methodology itself and the input data used to draw up each final list with which there is a problem which casts doubt on the optimality of either of these lists for use in the initial stages of the comparative method. The criteria of both item stability and the cross-linguistic applicability of results, as discussed above, are very important in our search for an optimal list; either item stability or cross-linguistic applicability of results, however, is sacrificed to an unacceptable extent in each of Lohr's lists.

List 2, created on the basis of cross-referencing those items which are maximally universal with those that are somewhat stable, sacrifices maximal stability in favour of universality. Universality is an important criterion to consider a list optimal for use in the initial stages of the comparative method, in order to render a standard list applicable cross-linguistically; however, while a list should be as universal as possible, it should not be so at the expense of stability. A list that

maximises item universality at the expense of item stability reduces the number of similarities that will potentially enter the comparanda in the comparative method; how universal the items on a list are thus becomes moot. A list which maintains maximal stability while maximising item universality is preferable; such a list will allow for the greatest number of inherited similarities, where they exist, to enter the data, while increasing the applicability of the list cross-linguistically. List 2 is thus not appropriate for use in the initial stages of language comparison.

On the other hand, List 1, created on the basis of cross-referencing those items which are maximally stable with those which are somewhat universal, places emphasis on her second preliminary list; the second preliminary list, however, considers only Indo-European data, thus rendering any results ungeneralisable cross-linguistically. As noted, the procedure used to draw up this second preliminary list is, in effect, the same as that used by Dolgopolsky, discussed above; however, using Buck's dictionary at times presents a slightly more detailed history of item stability, in that the written records of many of the Indo-European languages render visible any earlier replacements of form which are subsequently themselves replaced. Nonetheless, written records do not record all replacements of form of an item, nor are written records available for all of the history of all of the languages; there may still be obscured changes, and this list therefore suffers from the same inability to fully represent an item's stability. However, by considering only Indo-European data, it also sacrifices cross-linguistic applicability of results to a much greater extent than Dolgopolsky's application of the methodology.

Neither of Lohr's lists, therefore, are optimal for use in the initial stages of the comparative method; while List 2 sacrifices the weight placed on item stability to an undesirable extent, the items identified in List 1 as more stable are not necessarily so cross-linguistically. Nevertheless, the methodology used to determine item stability will be useful to us in our pursuit of the optimal list for use in the initial stages of the comparative method; the fundamental problem in the determination of item stability in Lohr (1999) is not one of the methodology itself, but one of the data used as input. We shall therefore discuss the methodology used by Lohr and Dolgopolsky in more detail below.

2.5 Discussion

This section has explored the methodologies underlying a number of standardised meaning lists which have been drawn up to date, as well as the application of these methodologies, and has evaluated their optimality for use in the initial stages of the comparative method using our criteria of empiricism, cross-linguistic applicability of results, and adequate representation of item stability, borrowability, conceptual simplicity, and universality.

While the Swadesh lists were not formulated with the intention of listing those meanings optimal for use in the initial stages of the comparative method, we have shown that they have come to be viewed as such by many linguists. However, we have demonstrated that, according to our criteria of maximal stability and maximal resistance to borrowing, as well as our criteria of empiricism of methodology and transparency in its application, the Swadesh lists are not appropriate for use in the initial stages of the comparative method, unless those items listed were to be independently, statistically, and objectively confirmed as optimal. The Swadesh lists, and therefore also those lists which have been formulated on the ranking of these lists (namely, the ASJP list (Holman et al., 2008) and the Yakhontov list (Starostin, 1991, 2000)) will therefore inform our search for the optimal meaning list for use in the initial stages of the comparative method no further, either in terms of methodology, or as an input for any aspect of items stability, borrowability, conceptual simplicity, or universality.

The Leipzig-Jakarta list (Tadmor et al. 2010) has been shown to be founded on replicable, scientific methodology, and to have been formulated using quantifications of item borrowability, conceptual simplicity, and universality appropriate for our purposes. While the sample of only 41 languages means that there is much scope for further research regarding item universality, the analysis of a wider range of languages to determine the presence or absence of a meaning would be a very large-scale task, and is beyond the scope of this dissertation. The quantification of item universality used in the formulation of the Leipzig-Jakarta list is therefore, for the purposes of this study, taken to be adequate. Similarly, the validification of the assumption on which the determination of conceptual simplicity is based – that analyzability and conceptual simplicity correlate significantly in a positive

relationship – is also beyond the scope of this dissertation, and will therefore be taken to be accurate. Item stability is also taken into account in the formulation of the Leipzig-Jakarta list; however, a detailed picture of item stability of the most stable items, those with which we would be most concerned as optimal for comparison in the initial stages of the comparative method, is not given, owing to an effective cut-off in the measurement of item stability at a 1000-year time depth. The borrowability, simplicity, and universality scores calculated in the formulation of the Leipzig-Jakarta list will certainly be of further use to us in our pursuit of the optimal meaning list; however, a more effective means of quantifying item stability of those meanings whose forms potentially have very large time depth is required.

The procedures used by Dolgopolsky (1986) and Lohr (1999) to determine item stability have been shown to be, in principle, very similar. This methodology is founded on the principle that, the larger the number of extant cognate sets there are within a language family for a meaning, the more replacements of form will have taken place, and thus the less stable the item is within the language family. The methodology operates by determining the number of extant cognate sets within a language family and the number of years total time depth of the language family, in order to calculate the mean number of years the form representing an item is expected to be retained without replacement. However, the unscientific application of the methodology by Dolgopolsky and the limitation of the input data to the Indo-European language family by Lohr mean that the lists themselves are unempirical and not cross-linguistically representative, respectively. The methodology, however, while it has the flaw of not being able to identify those replacements of form which are obscured by later replacements, has the benefit of being able to take into account input from large amounts of cross-linguistic data, as well as representing a detailed picture of item stability for more stable items than the methodology used by Tadmor et al. (2010).

It is therefore proposed that a maximally transparent and cross-linguistic application of the methodology used by Dolgopolsky (1986) and Lohr (1999) be implemented in the remainder of this dissertation, in order to quantify cross-linguistic item stability, thus enabling the ranking of our input items by how stable they are. The representation of item stability presented will then be combined with the representations of item borrowability, conceptual simplicity, and universality

as determined in Tadmor et al. (2010), in order to fully integrate the four criteria identified here as facilitating the initial stages of language comparison; the input items can therefore be ranked by their optimality for use in the application of the comparative method.

The results of this investigation will put us in the fortunate position of being able to statistically explore other issues pertinent to the basicness of an item. The question of whether more stable items vary less in their stability cross-linguistically will be addressed. In addition, in response to the frequent conflation of these four features of a meaning, highlighted throughout this section, the relationship between item stability and item borrowability, conceptual simplicity, and universality will be assessed. Finally, the results gained will enable the statistical exploration of the validity of the glottochronological hypothesis; we will take the opportunity to do so, thereby presenting significant results for historical and comparative linguistics.

3 Determining item stability: Method

As was demonstrated in the previous section, an optimal meaning list for use in the initial stages of the comparative method would comprise items which are maximally stable, maximally resistant to replacement by borrowing, maximally conceptually simple, and maximally universal. While the Leipzig-Jakarta list (Tadmor et al. 2010) quantifies item borrowability, potentially conceptual simplicity, and, to a lesser extent, item universality, in means which are appropriate for our purposes, it was demonstrated that the quantification of item stability was not adequate to determine the relative stability of the most stable items, those with which we are most concerned if we are to maximise the likelihood that evidence for genetic similarity, where it exists, is present in comparison. While the list determining item stability presented by Dolgopolsky (1986) is based on unscientific, opaque processes, and that presented by Lohr (1999) is based on information from only one language family, and thus is not cross-linguistically representative, the methodology used by both has the potential to be transparently applied to a range of language families, taking into account large amounts of quantifiable, detailed, cross-linguistic information from which to determine item stability.

This section thus details the procedures in the application of this methodology, discussing in turn the metric of item stability to be employed in this dissertation; the determination of which meanings to investigate; the determination of the language families and languages within these families from which to draw data; the collection of the primary extant form(s) for each meaning, to enable the determination of the number of cognate sets apparent within each language family for each meaning; the determination of the minimum number of replacements of form to have taken place for the cognacy patterns observed to have occurred; and, ultimately, using this information, the quantification of item stability. The results of this investigation, and the combination of the item stability score with the scores used in Tadmor et al. (2010) to determine the relative optimality of individual meanings for use in the initial stages of the comparative method, will be presented in section 4.

3.1 The metric of item stability

In order to facilitate the integration of the stability scores with the borrowed, analyzability, and representation scores given in Tadmor et al. (2010), the representation of item stability will not be as the mean number of years a form representing an item is expected to be retained without replacement by another form, as it is in Lohr (1999) and Dolgopolsky (1986). Rather, it will be in terms of the *probability* that a form representing an item will be retained for a thousand-year timespan. This novel approach to the quantification of item stability has the added benefit of enabling meaningful discussion about the likelihood that a form representing a meaning would be retained for, for example, 3000 years, and thus will help the analysis of proposals of genetic relationship on more statistical grounds.[14]

A further benefit of using probability of retainment of form as the metric for item stability is that it does *not* carry any implication that item stability is constant or predictable. As was discussed in depth in section 2.3, there is simply no empirical linguistic evidence to support this theory; indeed, we will present further evidence to support the rejection of this assumption in section 4.3. As was also noted, however, this does not render any discussion of rates of change meaningless, as long as the discussion is based on observable evidence. Imagine, for example, a group of six sprinters competing in a 100m race. If we count the number of times the competitors trip and fall in, say, a thousand races, we can determine the probability that a competitor will trip and fall in one race. Crucially, this calculation does not depend on the competitors running at a constant or predictable speed; neither does the calculation of the probability of a form representing an item being retained for a thousand years depend on a constant or predictable rate of change.

3.2 The input meanings for investigation

Ideally, the meanings that would be assessed for relative stability in the determination of an optimal meaning list would be those looked at in WOLD: 1460 meanings, ultimately based on Buck (1949). Not only would this facilitate the combination of the newly-calculated item stability scores with the scores representing item borrowability, simplicity, and universality presented in WOLD

[14] I am indebted to Prof Ronnie Cann for first suggesting this means of representing stability.

(Haspelmath and Tadmor 2009), but would create a wide catchment of meanings, from which a shorter list of those meanings which are optimal for use in the initial stages of the comparative method could be refined. However, owing to the timescale of this project, such a wide range of meanings could not be taken into consideration; furthermore, as will be discussed below, some of the etymological resources available for the language families, essential for determining cognacy of extant forms, are quite limited in scope, severely restricting the number of items whose cognacy can be considered in this way.

The decision was thus made to take, as a starting point, the intersection of the Leipzig-Jakarta list with one of the Lohr lists; these meaning lists are those identified in section 2.4 as having been formulated using appropriate methodology, empirically applied. However, as it is item stability that is under scrutiny, in order to avoid circularity, the input of item stability into these lists must be kept to a minimum. While item stability is indeed a criterion in the Leipzig-Jakarta list, the cut-off point of 1000 years at which item stability is no longer measured in detail means that stability only contributes to the composite score by which the items are ranked to a certain extent; the Leipzig-Jakarta list as it stands was thus used as input for intersection.

Of more concern perhaps is the determination of how to use Lohr's research to avoid circularity. In the course of her investigation, Lohr drew up four lists: her first preliminary list of those items reconstructable in two or more of four proto-languages (thereby giving weight to universality, and, to some extent, stability[15] and borrowability); her second preliminary list, based on item stability in Indo-European (thereby giving weight to item stability, but none to universality); and her List 1 and List 2, both of which were created using intersections of the two preliminary lists. In order to minimise circularity, the input from her second preliminary list, that which emphasises item stability in Indo-European, must be kept to a minimum. The only list which has no input from the second preliminary

[15] While it is true that reconstructability does to a certain extent depend on an item's stability – a form representing an item must have been widely retained throughout the language family in order to provide evidence of an ancestral form, and therefore items which have been reconstructed are liable to be more stable than those that have not been – note that items do not have to have been maximally stable in order to be reconstructable. Such items will of course be relatively conceptually stable; however, as discussed in section 2.2, this definition of stability is not that which is our priority.

list is the first preliminary list; it was therefore the first preliminary list that was intersected with the Leipzig-Jakarta list.

The implication of the decision to intersect the Leipzig-Jakarta list and Lohr's first preliminary list, in effect working from the bottom up, rather than taking a wider range of meanings and discarding those meanings which are less stable, is that the result will not necessarily in itself be a list of those items which are most stable cross-linguistically. The resultant ranking will simply determine the relative stability of the input items to one another. The research here presented, therefore, will outline the viability of statistically determining item stability, and indicate preliminary results; a more comprehensive study of item stability, culminating in a list of the most stable items cross-linguistically, refined from a much wider range of meanings, must await future research.

The intersection resulted in a list of 74 items. Those meanings which were considered to have a high potential for being represented by onomatopoeic forms were removed, in order to reduce the likelihood that any similarities identified using the list in the initial stages of the comparative method were due to the non-genetic cross-linguistic tendency for these meanings to be represented by such forms. While this is necessarily a subjective procedure until the nature of the representation of meanings by onomatopoeic forms is more thoroughly understood, a conservative stance was taken, in order to take reasonable care to ensure that such meanings are not advocated for use in the initial stages of the comparative method.[16] The preposition 'in' was also removed from further consideration during the course of

[16] The meanings removed were 'to blow', 'to cry', 'to laugh', and 'to suck'. During the course of data collection, the decision was taken to remove 'to hit', owing to the overlap of the form in some languages (such as the Dravidian languages Pengo and Brahui) with the meaning 'to beat', which it was felt was more likely to be represented by an onomatopoeic form.

It should be noted that the removal of meanings which are considered likely to be represented by onomatopoeic forms at this stage is somewhat circular – it was stated in section 2.1 that, in order to determine which meanings should be discarded as having a high likelihood of being represented by onomatopoeic forms, we must first determine which meanings are optimal for use in the initial stages of the comparative method. However, as there is as yet no statistical means of determining the likelihood that an item be represented by an onomatopoeic form, the decision was taken, for the purposes of this study, to remove these items *a priori*; a reasonable picture of the relative optimality of the input items for use in the initial stages of the comparative method can thus be given. Means of incorporating the likelihood that an item will be represented by an onomatopoeic or sound symbolic form into the determination of which items are optimal for use in the initial stages of the comparative method will be discussed in section 5.

data collection, owing to the difficulty in accurately identifying the primary form in many languages. This resulted in a final list of 67 items, presented below in Table 4.

The precise definition of the extension of each meaning was necessary for many of the items, in order to ensure that the forms for the meanings with the same extension were being compared between languages; as no further information was provided in Lohr (1999), any ambiguity was resolved using the descriptions and typical contexts in which the items are used, given in WOLD (Haspelmath and Tadmor 2009). Further information relevant to each meaning is given in brackets.[17] Notable was the potential for overlap of a form with a meaning with another extension, for example 'mouth (and lips)', 'shade (and shadow)', the more common of which are indicated. However, as long as the form also represents the indicated meaning, this was not considered to be an issue; it is the domain of the meaning, rather than the domain of the form, with which we are concerned.[18]

3.3 The language families and the languages from which to draw data

In order to accurately determine the number of cognate sets present in a language family for a particular meaning, for use in the calculation of the probability that a form representing an item will be retained for a thousand years in that language family, etymological resources are required that provide the cognacy not only of the forms inherited from the proto-language, but of all extant forms in the language family. Such resources were found for Austronesian (Greenhill et al. 2008[19]),

[17] This information is by no means exhaustive; it is merely the result of the demands of the particular languages looked at in this study.

[18] It will be noted that there is the potential for overlap between items on the list: specifically, 21. 'foot (and leg)' and 34. 'leg (and foot)'. However, as discussed in our evaluation of Dolgopolsky's list in section 2.4.3, while this kind of overlap should be avoided in the pursuit of lexicostatistical or glottochronological calculations, as we are only concerned with the calculation of the probability for *each individual meaning* of its form being retained for a thousand-year timespan (rather than the total of binary cognacy judgements, as in lexicostatistics or glottochronology), reduplication will not affect the results. In other words, it does not matter for our purposes if the same form or root represents the meanings 'leg' and 'foot' in any of our sample languages; we are concerned only with the likelihood that the form representing each meaning will be replaced.

[19] While the cognacy judgements in the Austronesian Basic Vocabulary Database are impressively comprehensive and adhere strictly to the comparative method, they are constantly

1. big
2. to bite (tr.)
3. bitter
4. black
5. blood
6. bone
7. to burn (intr.)
8. to carry
9. child (kinship term)
10. to come
11. to do (tr.)
12. dog
13. ear
14. to eat (tr.)
15. egg
16. eye
17. to fall ('to drop')
18. fire
19. fish
20. fly
21. foot (and 'leg')
22. to give
23. to go
24. good
25. to grind (dry substances, finely)
26. hair (of the human head)
27. hard (vs. 'soft')
28. to hear
29. to hide (tr.)
30. house (traditional)
31. I (1st.sg.subject)
32. it ('he', 'she'; 3rd. sg.neutral.subject. proximal.definite)
33. to know ('savoir')
34. leg (and 'foot')
35. liver
36. long
37. meat (and 'flesh')
38. mouth (and 'lips')
39. name
40. new
41. night
42. nose
43. not
44. old (vs. 'young', not vs. 'new')
45. one (adj.)
46. rain
47. root
48. salt
49. sand
50. to say (intr.)
51. to see
52. shade (and 'shadow')
53. small
54. to stand (animate subject; 'to be erect' rather than 'to stand up')
55. star
56. to take
57. thick (dimension, not density)
58. thigh
59. this (adj.)
60. to tie ('to bind')
61. tongue
62. tooth
63. (fresh) water
64. wide
65. wind
66. wood (material)
67. you (2nd.sg.subject)

Table 4: The 67 meanings for investigation, resulting from the intersection of the Leipzig-Jakarta list with Lohr's first preliminary list.

Dravidian (Burrow and Emeneau 1984), Indo-European (Buck 1949, supplemented by Dyen et al. 1992), and Uto-Aztecan (Stubbs 2011).[20] The language families from which information could be drawn to calculate item stability was thus limited to these four languages families. The language families represent a pleasing spread of languages spoken in various linguistic, geographic, socio-cultural, economic, political, and literary contexts for our purposes, allowing us to capture many of the factors indicated in section 2.3 as potential influences on the rate with which forms are replaced. In addition, a wider range of languages is represented in this sample than in Lohr (1999), increasing the cross-linguistic applicability of the results presented herein. However, it should be noted that, with the advent of more comprehensive resources, there will be much scope for improving the cross-linguistic relevance of the results presented in this dissertation by including quantifications of item stability from a greater number of language families.

From within each language family, a range of languages from which to draw the primary form(s) for each meaning were selected. This mainly depended, once again, on the resources that were readily available. As even some glottochronological work, so dependent on a predictable rate of item replacement, suggests that the rate of item replacement may vary not only from language family to language family, but from language to language (Embleton 1986:74; Kruskal et al. 1973:52), an effort was made to represent as many branches of the language family as possible, in order to capture as much potential variation as possible. This decision also had the benefit of further increasing the range of those contexts discussed above which may affect the rate of item replacement.

The languages from which information on the cognacy of the primary form(s) representing an item were drawn, along with the primary resources used to determine the primary form(s) of a meaning, are as follows:

---

under improvement and correction, and cognacy judgements given should not be taken as final. The preliminary nature of this database thus warrants its use with caution; the implications of the conservative attitude that must be taken are discussed in detail in section 3.5.

[20] Such a resource was also found for the proposed 'Altaic' language family (Starostin et al. 2003); however, considering not only the indeterminate nature of the Altaic hypothesis (see, for example, Campbell and Poser 2008:235-341; Comrie 1993), but also the critical reviews the dictionary received from both non-Altaicists (Georg 2004) and former Altaicists (Vovin 2005) alike, it was decided to hold this particular resource at arm's length.

- **Austronesian:** Atayal, Tsou, Rukai, Paiwan, Yami, Isnag, Kalinga Limos, Tagalog, Aklanon, Kagayen, Sarngari Blaan, Murut, Malagasay, Aceh, Toba Batak, Minangkabau, Indonesian, Sundanese, Javanese, Madurese, Balinese, Sasak, Gorontalo, Bugis, Wolio, Manggarai, Ngada, Sika, Roti, Buru, Irarutu, Manam, Takia, Yabem, Kaulong, Tolia, Kilivila, Tawala, Motu, Mekeo, Roviana, Maringe, Lau, Kwaio, Raga, Paamese, Kwamera, Xârâcùù, Nengone, Kiribati, Marshallese, Ponapean, Woleanian, East Fijian, West Fijian, Rotuman, Tongan, Samoan, Tahitian, and Rapanui (Tryon 1995, cross-referenced with Greenhill et al. 2008, used to determine primary forms for all Austronesian languages).

- **Dravidian:** Tamil (Asher 1982; Fabricus 1972; Mahadevan 2011; Tamil-English-Tamil Dictionary, accessed 10/2/12), Malayalam (Asher and Kumari 1997; Zilva Wickremasinghe and Menon 2005), Kannada (Rao 1967; Sridhar 1990), Telugu (Gwynn 1991; Telugu-English-Telugu Dictionary, accessed 16/2/12), Kolami (Emeneau 1961), Pengo (Burrow and Bhattacharya 1970), Parji (Burrow and Bhattacharya 1953), Ollari (Bhattacharya 1957), Kui (Letchmajee 1902), and Brahui (Bray 1934).

- **Indo-European:** Ancient Greek, Present Day Greek, Latin, Italian, French, Spanish, Romanian, Old Irish, Present Day Irish, Welsh, Breton, Gothic, Old Norse, Danish, Swedish, Old English, Middle English, Present Day English, Dutch, Old High German, Middle High German, Present Day German, Lithuanian, Latvian, Church Slavonic, Serbo-Croat, Czech, Polish, Russian, Sanskrit, and Avestan (Buck 1949, supplemented by Dyen et al. 1992, used to determine primary forms for all Indo-European languages).

- **Uto-Aztecan:** Northern Paiute (Yerington Paiute Tribe 1987), Shoshoni (Dictionary, accessed 3/4/12), Comanche (Canonge 1958), Chemehuevi (Press 1979), Southern Paiute (Sapir 1931), Luiseño (Kroeber and Grace 1960), Tübatülabal (Voegelin 1958), Hopi (Hopi Dictionary Project 1998), Pima de Yepachec (Shaul 1994), Tarahumara (Thord-Gray 1955), Yaqui (Fernández 2009), Mayo (Collard and Collard 1962), Cora (McMahon and

de McMahon 1959), and Nahuatl (Herrara 2004; http://whp.uoregon.edu/dictionaries/nahuatl/index.lasso, accessed 2/3/12).

In her application of this methodology, Lohr (1999:58) does not recommend collecting data from languages with a deep individual time depth (i.e. those which have developed independently from any of the other languages which have been chosen from which to collect data for this study). In these cases, earlier replacements of form may be obscured, thus leading to more inaccurate determinations of item stability. Lohr cites this as the reason she did not measure item stability in Armenian or Albanian. However, it was felt that the inclusion of data from such languages could also have benefits. First, the overall time depth of the languages studied would be increased. According to the law of large numbers, the more times the same experiment is repeated, the closer the average of the results will come to the expected value. In effect, our 'experiment' is determining whether the form representing an item has been replaced in a thousand-year timespan in a language family; the more times this experiment can be repeated, therefore, by looking at item replacement over as great a time-depth as possible while maintaining accuracy, the better. Furthermore, as has been stressed throughout this section, the inclusion of as many languages from as many different linguistic, geographic, socio-cultural, etc. contexts as possible is vital to maximise cross-linguistic applicability of results. Finally, there are only five examples in our large sample of languages identified as having a time depth of around 5000 years (Atayal, Paiwan (Austronesian); Cora, Nahuatl, and Pima de Yepachec (Uto-Aztecan)), the point at which Lohr suggests one should become wary of the results. The potential benefits, therefore, outweighed the potential drawbacks, thus informing our decision to collect data from languages with a deep individual time depth.

As was mentioned above, the calculation of the total time depth of the languages from which data were collected was required for each language family, in order to calculate the probability for each meaning that the form by which it is represented will be retained for a thousand years in each language family. This involved determining the most current, most widely accepted sub-groupings within the family. The resources used to determine both the language family sub-grouping

and estimates of the dates at which the languages are thought to split from one another were:

- **Austronesian:** Greenhill et al. (2008).
- **Dravidian:** Krishnamurti (2003).
- **Indo-European:** Lohr (1999:57-8), Mallory and Adams (2006:15-26; 103-4).
- **Uto-Aztecan:** Campbell (1997:134, 137), McLaughlin (1992:158-9), Miller (1983:118, 121), Miller (1986:100), Mithun (1999:539-540), Stubbs (2011:3).

The family trees used for each of the language families are given in Appendix B. Included on these trees are conservative best-judgements of the split dates of the languages, as well as, along each branch, the time depth between splits.

3.4 The collection of the primary extant forms

With the determination both of the meanings for which to collect data, and the language families and languages from which to collect data, the primary forms representing each meaning in each language were assembled, in order to determine how many cognate sets can be identified in each language family, and thus calculate items' relative stability within each language family. For the purposes of this dissertation, a primary form is (or multiple primary forms are) defined as the most common form(s) used by the majority of the speaker population for a meaning, unmarked in terms of stylistic domain (i.e. not restricted to use in, for example, religious or poetic contexts). The form(s) should be either:

1. the singular subject form (if a noun)
2. the form which interacts syntactically with singular nouns (if a verb)
3. the form which interacts syntactically with singular subject nouns of the least marked nominal class (if an adjective).

Emphasis was placed on the collection of primary forms so that we can ensure that we are assessing the cognacy of the forms representing like meanings with

like meanings, i.e. so that the meanings are standardised across the languages. In addition, as was discussed in section 2, it is often hypothesised that it is the most basic meanings, and it has been statistically shown that it is the most frequently used meanings (Pagel et al. 2007), which are the most stable. As the frequency with which forms are used is identified as being a feature of a primary form in our definition, it is precisely these meanings, and the cognacy of the forms used to represent them, with which we are most concerned.[21]

The forms were initially collected using the primary sources referenced above in section 3.3. However, as Slaska (2005) discusses in detail, this approach can be very problematic: it can lead to unbalanced and unrepresentative data being collected and analysed, and, in the worst cases, without input from native speakers, mistakes can easily creep in to the data. Therefore, where possible, specialists in the languages in question were asked to verify the primacy of the forms collected for each meaning. Slaska suggests that the primary forms of a meaning should be determined based on the input from several native speakers – however, while this would of course be a preferable means of gathering data, the large number of languages and meanings being handled in the present study unfortunately meant that such thoroughness was simply not viable.

If a datum was not present in the primary source, or if it was unclear as to whether the form listed represented precisely the required meaning, the datum was not collected for that language for the meaning, and the individual time depth of the language (i.e. the length of time it has developed independently from any of the other languages from which data were being collected) was subtracted from the total time depth of the languages in question for that meaning alone. The former is true, for example, for the meaning 'fly' in Gothic, 'blood' in Comanche (Uto-Aztecan), and 'new' in Kolami (Dravidian); the latter is true for the meaning 'to carry' in Cora (Uto-Aztecan) and the meaning 'old' in Parji (Dravidian). This removal of data, while hopefully improving the accuracy of results compared with if

[21] Whether a meaning can be translated precisely across languages' shifting and unmappable conceptual domains – whether meaning exists external to language – as well as whether a form can be identified with a meaning when used out of context, were, of course, theoretical and philosophical issues which were given due consideration in the process of data collection. Unfortunately, limitations on space preclude giving these questions thorough treatment here.

the data had been left in, is a further demonstration of how valuable native speaker and language specialist input are in determining the primary forms of a meaning.

3.5 The determination of the number of extant cognate sets within a language family for a meaning

Once the data were collected from the languages, each form for each meaning was analysed for cognacy within the language family using the etymological resources listed above. Each cognate set was given an identifying number. Sample cognacy judgements given for the primary forms of a selection of meanings in Dravidian are given below in Table 5.[22]

	this		**to tie**		**tongue**		**tooth**	
Tamil	inta	1	kaṭṭu	1	nākku	1	pal	1
Malayalam	ī	1	keṭṭuka	1	nākkə	1	pallə	1
Kannaḍa	ī	1	kaṭṭu	1	nālage	1	pal	1
Telugu	ī	1	kaTTu	1	nālika	1	[dantam] (<Skt.)	2
Kolami	i·	1	kaṭ-	1	na·lka	1	pal	1
Pengo	ī	1	gac-	2	-	-	-	-
Parji	i	1	kaṭṭ-	1	nevãḍ	1	pel	1
Ollari	-	-	kaṭ-	1	nāŋ	1	pal	1
Kui	-	-	-	-	-	-	paḍu	1
Brahui	dā	2	tafing	3	dūī	2	[dandān] (<Pers.)	3

Table 5: Cognacy judgements for the primary forms of a selection of meanings in Dravidian

[22] The primary forms collected for each meaning in each language, along with the cognacy judgements given, are available on request from the author. While every effort was made to ensure the accuracy of the data used in this dissertation, all corrections are sincerely welcomed.

At this stage, there were three circumstances under which data were removed from further consideration in the calculation of an item's stability:

1. If the only form representing the meaning in the etymological dictionary is obviously a different form to that identified in the initial data collection stage as the primary form. For example, the meaning 'to bite' in Kagayen (Austronesian) is identified as being represented by the form *maŋagat* in Tryon (1995:535), but *kitkít* in Greenhill et al. (2008); we are clearly dealing here with two unrelated extant forms, and to assess the former form's cognacy using the latter source would be meaningless. Slight deviation in terms of orthographic representation, dialectal variation, and differences in morphological segmentation were, however, allowed for. For example, the Atayal (Austronesian) form *ḳə-zəmə-zəməc* 'night' was maintained, despite the difference in terms of orthographic representation and morphological segmentation from the etymological source, where the form is given as *qəzəməzəmət*. The Sasak form for the same meaning, on the other hand, *kələm*, was removed due to the perceived discrepancy in the representation of the form in Greenhill et al. (2008) as *maləm*, despite three of the five segments being identical. As resources were limited, and as an incorrect judgement as to whether the cognacy of the form identified as the primary form for the meaning is that represented in the etymological source would have resulted in a skewing of the results, the side of conservatism was strictly erred on; still, the unfortunate subjectivity of the process must be noted.

2. If the cognacy of the form with other extant forms under comparison in the language family was marked as doubtful in the etymological source. The decision to remove such data was taken so as not to bias the results one way or another (as, for example, a decision to consistently mark forms of doubtful cognacy as not cognate would have done).[23] This includes the removal of

[23] Note that this did not mean the removal of all forms with doubtful cognacy in the etymological source. If the cognate set to which the form may or may not belong was not represented by any other forms in the languages under comparison, i.e. if the form would be of an individual cognate set in the data collected whether it were cognate with the forms suggested in the etymological source or not, the datum was retained and an individual cognacy score given. For example, while Rukai (Austronesian) *mā-roḍaŋa*'old' is given a doubtful cognacy judgement of 2 in Greenhill et al.

those items whose loan status is doubtful – if the form were a loan from a related language, this would represent an additional replacement of form than if it were a cognate form. For example, the Ancient Irish form *long* 'long' is marked in Buck (1949:882) as either cognate with Latin *longus*, or a borrowing of the form; as an incorrect decision of cognacy either way would skew the results, the form was removed.

3. If the cognacy judgements given (or, more often, not given) appeared questionable.[24] Again, this was, unfortunately, necessarily a subjective process; again, however, the removal of such data was approached with a strictly conservative attitude. A prime example of questionable cognacy judgement can be found for the meaning 'eye' in Austronesian. Reconstructed as **maCa* in Proto-Austronesian (Blust 1999, cited in Greenhill et al. 2008), the form representing this meaning has apparently been rather phonologically stable, with descendent forms such as Tsou *mɛ̆ō*, Rukai *maɛ̆a*, Indonesian *mata*, Mekeo *maa*, and Rapanui *mata*. The forms in Paiwan and Kagayen are *maɛ̆a* and *mata* respectively; however, neither of these forms has been given a cognacy judgement, nor are they coded as loanwords, i.e. the forms appear as independent innovations in the languages (Greenhill et al. 2008). While this is of course possible, it seems highly unlikely, given the similarity in form, coupled with the apparent phonological stability of the inherited form, that the Paiwan and Kagayen forms are not related in some manner, most likely through direct inheritance, but possibly through borrowing, to the Proto-Austronesian form. The decision was thus taken to remove the Paiwan and Kagayen data from the calculation of the probability that the form representing 'eye' would be retained for a thousand years in Austronesian, due to the questionable cognacy judgement given.

(2008), no other forms collected from the selected Austronesian languages for the same meaning were given a positive cognacy judgement of 2. This datum was therefore retained, as whether or not it belonged to the cognate set identified as 2 would not affect the form's relationship to the other forms collected; it is unrelated either way.

[24] While this was a particular issue with the Austronesian Basic Vocabulary Database (Greenhill et al. 2008) owing to the nature of the resource, identified in section 3.3, it should be noted that none of the etymological resources were free from this problem.

3.6 The determination of the minimum number of replacements of form

Once the number of visible cognate sets in the languages from which data were drawn was determined for each meaning in each language family, the minimum number of replacements of form required for the patterns of cognacy to be observed in each language family could then be inferred. Simply put, this was generally one fewer than the number of cognate sets identified, as described by Dolgopolsky (1986), except in cases where there were two or more extant primary forms for a meaning for one or more of the languages within a language family.

An item was considered to have been replaced in terms of form when the form by which it is primarily represented at time T_2 is not a direct lexical inheritance from the form by which it is primarily represented at time T_1. For example, the English word *leg*, today representing the primary form for the meaning 'leg', is not descended etymologically from the Old English form for the same meaning, *sceanca*; at least one replacement of form has occurred in English since the twelfth century for the meaning 'leg'.

Echoing our definition of item stability, the likelihood that the form representing a meaning will be replaced by either an endogenous or an exogenous form, the *means* by which a replacement occurs is not relevant for the purposes of determining the minimum number of replacements. The replacing form could have its origins in the language's own internal word-formation processes, or it could be a form which underwent semantic shift to become the primary form of the meaning, or a borrowed form from another language. In all of these situations, a visible replacement is considered to have taken place. To return to our metaphorical sprinters of section 3.1: precisely what causes a competitor to trip and fall – be it collision with another sprinter, uneven terrain, or the runner tripping over his or her own feet – is not relevant to the calculation of the probability that a sprinter will trip up in a race.

This judgement, that the origin of the replacing form is irrelevant, holds true even if a replacing form belongs to the same cognate set as the form it is representing, but has had a separate lexical history. For example, the Present Day Irish forms for 'to hear', *cluinim* and *cloisim*, are ultimately both reflexes of PIE *$\hat{k}leu$-(s-)* (Mallory and Adams 2006:335). As Buck (1949:1036) indicates, however, the latter

form is a re-formation of the verbal noun *clos* (itself derived from *cluinim*). As the creation of the new form represents a break in the direct lexical inheritance of the form *cluinim* – the origin of *cloisim* is the word-formation processes of Irish, rather than direct inheritance from *cluinim* – the form for the meaning 'to hear' is therefore considered to have undergone a minimum of half a replacement in Irish (see below for a discussion of fractional replacements).

The judgement that a visible replacement has taken place also holds true if a borrowed form is etymologically related to the form it is replacing, i.e. if the form is being borrowed from a related language and is a reflex of the same ancestral proto-form as the form it is replacing. For example, the Middle English form for 'egg', *ey*, was fully replaced during the Middle Ages throughout the speech community by an earlier borrowing into some dialects of English from Old Norse, *egg*. *Egg* and *ey* are both ultimately descended from PIE $*h_a\bar{o}(w)i\text{-}om$ (Mallory and Adams 2006:143); however, this lateral transmission of the Old Norse word represents an additional replacement of form in English.

Where there are two or more extant primary forms representing a meaning in a language, a minimum replacement of the relevant fraction is considered to have taken place – a half a replacement if there are two primary forms, a third for each replacing form if there are three, and so on. For example, there are two primary forms in Indonesian for 'ear': *təliŋa*, a reflex of the Proto-Austronesian **Caliŋa* (Blust, 1999, cited in Greenhill et al., 2008), and *kupiŋ*, a borrowing from either Javanese or Balinese (Tadmor 2009). The form representing the meaning 'ear' can be considered to have been halfway replaced; there is therefore inferred to have been a minimum of half a replacement of form for the item 'ear' in the transmission of Proto-Austronesian to present-day Indonesian. Another example is that of the Kaulong forms for 'house', *maŋ* and *mok*. As neither of these forms are reflexes of the Proto-Austronesian **ʀumaq*, the minimum number of replacements of the form for 'house' was determined to have been one and a half – once when the form inherited from the proto-form was replaced by one of these forms, and a half a replacement when the other of the forms came to share the status of primary form. These guidelines were consistently adhered to throughout the process of determining the minimum number of replacements of form in each language family.

Gothic	*triu*	9		
Old Norse	*trē*	9	*viðr*	6
Danish	*træ*	9	*ved*	6
Swedish	*trä*	9	*ved*	6
Old English	*trēow*	9	*wudu*	6
Middle English	*tre*	9	*wode*	6
PD English			*wood*	6

Table 6: The extant primary forms for 'wood' in a selection of Germanic languages

Thanks to the extensive literary tradition of many of the languages, more extant forms, and thus more replacements, are visible in Indo-European. This means we are frequently in the fortunate position of being able to identify replacements of form in progress. For example, the meaning 'tooth' in Old Irish was represented by the forms *dēt* (a reflex of PIE **h*$_1$*dónt*; Mallory and Adams 2006:174) and *fiacail* (of obscure origin; Buck 1949:231); the latter of the two forms came to fully replace the former as primary for the meaning 'tooth' (Present Day Irish *fiacal*). In this case, one full replacement was considered to have taken place; while, in Old Irish, *fiacail* was only one of two forms representing the meaning 'tooth', this was a 'snapshot' of the replacement of form that was taking place. The language(s) in which a form from a particular cognate set most fully represents the meaning in a language family is taken as the language representative of the extent of that cognate set's replacement of form in the language family.

A slightly different example will perhaps clarify the implications of this approach further. Consider the extant forms for 'wood' in the following Germanic languages, along with the identifying number the two different cognate sets were designated, given in Table 6. In this table, we can see that most of the languages have two primary forms meaning 'wood', belonging to cognate sets 9 and 6. Gothic, however, has only *triu* (cognate set 9), whereas Present Day English has only *wood* (cognate set 6). Cognate sets 9 and 6 were therefore considered to be representative of a full visible replacement of form each (neither of them being descended from PIE **dóru*; Mallory and Adams 2006:157), despite the forms from one of the cognate sets necessarily having only half replaced the form representing 'wood' in Old Norse,

Danish, Swedish, Old English and Middle English. Similar to the Irish case given above, the presence of two primary forms in these languages can be considered to be a synchronic 'snapshot' of a diachronic replacement; one, however, which has in this case been fossilised, thus giving the patterns of cognacy observed.[25] This explains the decision to use, for each cognate set, the extent of replacement in the language(s) in which a form from the cognate set has visibly most fully replaced the earlier form as the extent of replacement that the cognate set represents in the language family as a whole. If a form from a cognate set is the only primary form for a meaning in Language X, and yet is one of two primary forms in Language Y, the form has come to most fully represent the meaning in Language X; it is therefore the extent to which the form represents the meaning in Language X, i.e. fully, which is taken as the extent of replacement the cognate set represents in the language family, i.e. one full replacement.

There were circumstances under which the inference given above, that the minimum number of visible replacements is equal to one fewer than the number of extant cognate sets, may not or does not appear to hold. For example, there are two cognate sets present in the Dravidian languages looked at for the meaning 'to eat': one represented in Tamil and Brahui, reconstructed to Proto-Dravidian as **tiHn-*, the other represented in the remaining languages from which data were drawn, reconstructed as **uHṉ-/ūṉ-* (Krishnamurti, 2003:528, 524). This patterning is shown in Figure 1.

As the last common ancestor of Tamil and Brahui was Proto-Dravidian, that the two languages share a cognate indicates that the form was present in the proto-language. The pattern of cognacy thus observed has, hypothetically, several possible explanations:

1. There were two primary forms for 'to eat' in Proto-Dravidian (either co-existing throughout the speech community, or as the forms used by two equally prestigious or widely-used dialects), both of which were variously inherited through the tree (suggesting a minimum of one and a half replacements of form in the tree: two half-replacements as the reflexes of **tiHn-* came to fully represent 'to eat' in Tamil and Brahui, and at least one half-replacement

[25] This should not be taken to imply that the replacement process is considered to be inexorable or unidirectional.

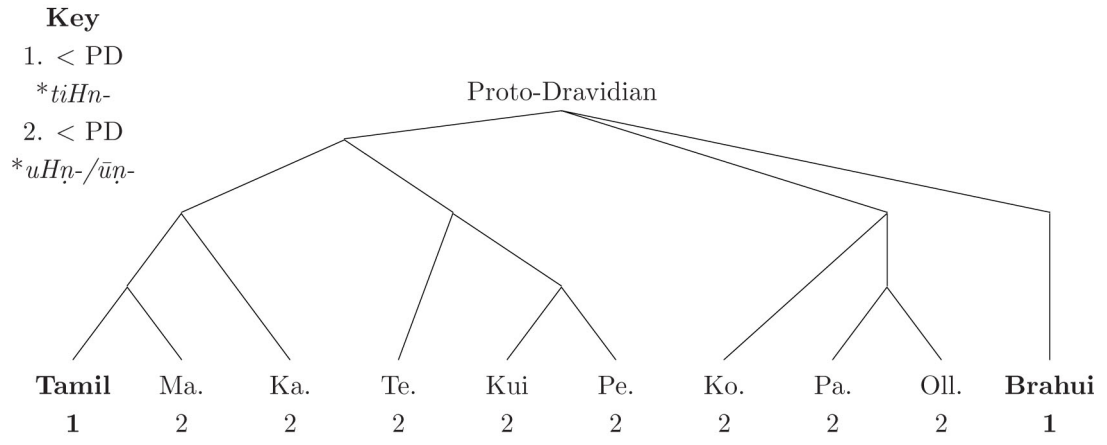


Figure 1: The cognacy of 'to eat' in Dravidian (Tamil and Brahui highlighted)

as the reflexes of *$uHṇ$-/$ūṇ$- came to fully represent the meaning in the remaining languages).

2. There was one primary form for 'to eat' in Proto-Dravidian, *$uHṇ$-/$ūṇ$-, and a semantically related or stylistically distinct form, *$tiHn$-. *$uHṇ$-/$ūṇ$- was the primary form inherited through most of the tree, but in Tamil and Brahui, two independent semantic shifts occurred, resulting in the form inherited from *$tiHn$- coming to be the primary form (suggesting a minimum of two replacements of form in the tree).

3. The *$tiHn$- form was shifted semantically in either Tamil or Brahui to come to mean 'to eat', and was then borrowed into the other language (suggesting a minimum of two replacements of form in the tree: the first being the semantic shift, the second being the unidentified borrowing).

4. There was an unidentified period of shared history between Tamil and Brahui, resulting in a sub-group innovation of the semantic shift of *$tiHn$- to represent the primary form 'to eat' (suggesting a minimum of one replacement of form in the tree).

Some of these possibilities are more likely than others – the final explanation is, for example, nigh on impossible. Such a shared innovation would have to exclude

Malayalam from the proposed sub-group, or we would expect to see a form of the same cognate set representing 'to eat' here too (unless an additional semantic shift were to have taken place following the Tamil-Malayalam split to reinstate the reflex of *$uHn̩$-/$ūn̩$- as the sole primary form); the comparatively recent split of Tamil and Malayalam, given in Krishnamurti (2003:22) as around the 9th century CE, therefore precludes a shared period of history of Tamil and Brahui before this point.

This example does, however, indicate the difficulties with determining the minimum number of replacements of form when a cognate set transcends sub-groups, and thus cannot be easily ascribed to shared innovation without considering alternative possibilities. When looking at language families for which the history is even less well-known than Dravidian, any one of the four hypothetical scenarios given above could explain apparent shared cognacy. Indeed, even the precise history of the well-studied Indo-European languages is not known – the origin of the *centum-satem* split and the verification or rejection of the Italo-Celtic hypothesis are just two issues which remain unresolved (see, e.g. Andersen 2009; Fortson 2004:53-4; McMahon and McMahon 2005:74-5). Given this uncertainty, it was decided that consistency, whilst maintaining conservatism of judgement, was paramount in the determination of the minimum number of replacements of form required for the patterns of cognacy observed for a meaning in a language family to arise. Therefore, the *minimum* number of replacements of form was determined as being one fewer than the total number of cognate sets represented by forms which fully represent the item in question, plus any fractional replacements identified.

3.7 The translation of the minimum number of replacements of form into a probability score

Once the minimum number of replacements of form necessary for the extant cognacy patterns observed to occur was determined, the probability that the form representing an item would be retained for a thousand years could be calculated for each language family, using the following formula:

$$p = 1 - \frac{\text{minimum number of changes}}{(\text{total number of years}/1000)}$$

It should again be stressed that the total number of years used in the calculation takes into account the languages for which data were missing or removed under the circumstances discussed above, i.e. if a datum was missing for a language or languages in the language family for a meaning, the total individual time depth of this language or these languages was subtracted from the total time depth for the language family.

The result is the probability that a form representing the meaning would be retained for a thousand years without replacement in the language family. For each meaning, the mean of these results was calculated across the language families[26]: this mean will henceforth be referred to as the **stability score**.

3.8 Discussion

This section has outlined the procedures used in the cross-linguistic and transparent implementation of the methodology used by Dolgopolsky (1986) and Lohr (1999). This methodology quantifies item stability by using the minimum number of replacements of form necessary for the patterns of cognacy observed within a language family to have occurred, along with the total number of years of the languages looked at from that language family, to calculate the probability of the form representing a meaning being retained for a thousand years, i.e. without replacement by another form which is not a direct lexical inheritance. Many of the limitations and potential pitfalls of the method have been discussed in detail: the impossibility of using anything other than the minimum number of visible replacements in the calculation using this methodology, running the risk of a less accurate calculation of stability, particularly for less stable items; the dependence of the method on others' cognacy judgements, including some judgements considered to be questionable; the determination of the primary form(s) for a meaning in the various languages; and the determination of the minimum number of replacements of form necessary for the extant cognacy patterns to be observed, particularly in

[26] Those meanings represented by data from only two language families were removed at this stage, *viz.* 'not', 'thigh', and 'this'.

cases where a form does not appear to be inherited from the primary form in the proto-language for the meaning, and yet still transcends subgroups.

This section has also outlined the precautions taken to keep inaccuracies arising from these potential hazards to a minimum. For example, if the cognacy judgement was given as doubtful or was conservatively judged to be doubtful, if the form provided in the etymological dictionary was conservatively considered to be a different word from that determined to be the primary form, or if the meaning could not be positively determined to represent the precise extension required, the datum was removed in order to remove any source of potential bias. Wherever possible, specialists in the languages from which data were collected were consulted to verify that the forms collected for each meaning are indeed the primary forms. The *minimum* number of replacements of form required was determined throughout, in order to maintain consistency where the deeper history of the language family may be unknown. The reasoning behind these precautions has been thoroughly and transparently discussed, enabling objective and scientific evaluation of the methodology and its application. However, before the results of this research are presented in section 4, there remain a couple of issues related to the method used and the representation of item stability requiring clarification.

First, it should be stressed that, while the stability scores calculated for each item may be more or less accurate at predicting the number of years the form representing a meaning may be retained before it is replaced, the scores are essentially descriptive, based on observations from a comparatively small window of human language. While we can, using these scores, make statements regarding how probable it is, based on observed events, that a form will be replaced in 10 000 years, or about the 'linguistic half-life' of an item – how many years it takes the probability that an item will have been replaced to reach 0.5 (Pagel et al. 2007) – the scores do not predict that the form representing an item *cannot* be replaced after 100 years, or 100 000 years. Rather, the scores comment on the *probability* of such an occurrence.

Second, the representation of item stability as the probability that the form representing the meaning being retained for a thousand years within a language family should not be taken to imply that this rate of change is consistent throughout the language family. Again, as was discussed above in section 2.3, the rate of

change has been shown to vary not only from language to language (Embleton 1986), but also diachronically within the same language (Kruskal et al. 1973); each meaning's stability score in each language family is thus based on fluctuating rates of change throughout the language family.

Finally, the model of replacement of form adopted for the purposes of this paper is, unfortunately, necessarily simplistic. It works at times with an idealised theory of replacement of form in which the forms representing meanings are clearly inherited through the branches of a family tree, until they are cleanly replaced overnight. This is particularly true for those language families in which there has not been a long written tradition, which would allow us to see a diachronic picture of replacement of form in progress, in a large number of the languages which have been incorporated into an etymological resource. However, while this approach does not represent the subtleties and fluidity of language use, variation, and change, and is therefore a rather blunt instrument, it does allow for a meaningful overview of the relative stability of the sample meanings with the tools we currently have available to us. If we are to develop our understanding of the processes of language change, as well as facilitate the use of the comparative method in the initial stages of determining genetic relationship between languages, the items on which we are most likely to focus, the most stable items cross-linguistically, must be identified. This study represents the next steps in this process.

4 Results

In this section, we will present the results of our quantification of the stability of the meanings given in Table 4. We will first present the items ranked by their stability score, in order to identify which of the meanings looked at are, on average, most likely to be retained for a thousand-year timespan – which are the most stable. We will use these figures to test a hypothesis regarding the relationship between item stability and the variation in this stability between language families.

We then turn to the main purpose of this investigation: the integration of a more accurate quantification of stability than used in Tadmor et al. (2010) with the borrowed, analyzability, and representation scores reported therein. The modified composite scores will be used to adapt the ranking of the meanings under investigation from that given in the Leipzig-Jakarta list, thereby taking a step towards the definition of the optimal meaning list for use in the initial stages of the comparative method. Using the stability scores in conjunction with Tadmor et al.'s scores, we will also explore the relationship between item stability and item borrowability, simplicity, and universality. Finally, we will take the opportunity offered by the results presented in this dissertation to statistically assess the validity of the glottochronological tenet that the forms representing basic items are replaced at a constant rate.

4.1 Item stability

Once the stability score had been calculated for each of the meanings under investigation, the items were ranked by this score, in order to identify which of the meanings are, on average, most stable. This ranking is presented below in Table 7, with the most stable items ranked highest. The standard deviation for each of the stability scores is also given.[27]

[27] Both the stability scores and the standard deviations of these scores are rounded to three decimal places.

Table 7: The meanings ranked by stability

Rank	Meaning	Stability score	Standard deviation
1	I (1st.sg pro)	0.983	0.034
2	eye	0.958	0.041
3	you (2nd.sg pro)	0.956	0.043
4	name	0.942	0.066
5	tongue	0.938	0.014
6	ear	0.932	0.06
6	tooth	0.932	0.062
8	salt	0.916	0.093
9	nose	0.911	0.051
10	shade	0.907	0.02
11	one	0.906	0.114
12	to hear	0.9	0.048
12	water	0.9	0.057
14	new	0.897	0.07
15	liver	0.89	0.079
16	to stand	0.887	0.112
17	to give	0.884	0.06
18	fly	0.88	0.071
19	to bite	0.877	0.067
19	to eat	0.877	0.046
21	root	0.876	0.026
22	to come	0.872	0.094
23	bitter	0.871	0.01
24	bone	0.865	0.055
25	wind	0.864	0.088
26	house	0.859	0.065
27	foot	0.856	0.071
28	to go	0.851	0.036
29	night	0.848	0.143

Continued on next page...

The meanings ranked by stability – Continued

Rank	Meaning	Stability score	Standard deviation
30	mouth	0.847	0.063
31	blood	0.846	0.044
31	star	0.846	0.062
33	it (3rd.sg. neutral pro)	0.843	0.059
34	hard	0.842	0.052
35	fish	0.84	0.05
36	fire	0.837	0.071
37	rain	0.835	0.03
38	leg	0.833	0.075
39	dog	0.832	0.077
39	sand	0.832	0.069
41	to know	0.823	0.082
42	to tie	0.82	0.081
43	egg	0.818	0.089
44	to take	0.808	0.053
45	to see	0.803	0.043
46	wood	0.803	0.081
47	black	0.797	0.109
48	old	0.792	0.037
49	to do	0.791	0.081
50	to grind	0.784	0.106
51	to fall	0.779	0.075
52	long	0.776	0.067
53	big	0.774	0.087
54	to burn (intr.)	0.765	0.11
55	hair	0.761	0.066
56	thick	0.759	0.053
57	wide	0.756	0.079
58	small	0.741	0.038

Continued on next page...

The meanings ranked by stability – Continued

Rank	Meaning	Stability score	Standard deviation
59	to hide (tr.)	0.74	0.026
60	meat	0.739	0.083
61	to carry	0.734	0.08
62	child	0.733	0.074
63	to say	0.721	0.06
64	good	0.696	0.088

4.1.1 The relationship between item stability and the variation in stability between language families

As the data were being collected, it appeared that the more stable an item was, the less this stability seemed to vary between the language families looked at, and *vice versa*. In order to test this theory, a Pearson product-moment correlation coefficient was calculated between the mean probabilities of retainment for a thousand years (the stability scores) and the standard deviations of these means (the extent to which the stability varied between the language families), in order to assess the relationship between item stability and variation in stability. There was found to be a significant negative correlation between the mean probability that an item would be retained for a thousand years and the variation of the mean, $r = -0.238$, $n=64$, $p < 0.05$. This result supports the hypothesis that the more stable an item is, the less its stability will vary between language families. The results are summarised in Figure 2.

4.2 The integration of the stability scores with the Leipzig-Jakarta list

This research has been carried out, as discussed in section 2.5, in order to determine a more accurate means of quantifying stability than that used in Tadmor et al.

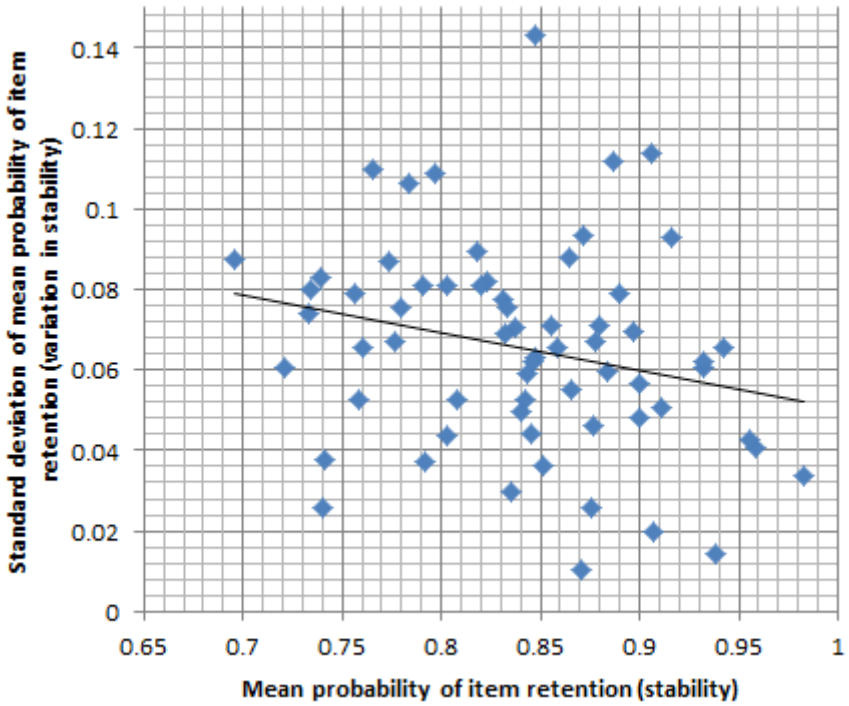


Figure 2: The relationship between item stability and the variation in stability between language families, r= −0.238, n=64, $p < 0.05$

(2010), while maintaining empiricism and cross-linguistic applicability of results. The goal of this quantification has been to incorporate a more detailed picture of item stability into a composite score with Tadmor et al.'s quantifications of item borrowability, simplicity, and universality, thereby determining a figure fully representative of all the criteria we established in section 2.2 as necessary to identify those items which will maximally facilitate the implementation of the initial stages of the comparative method – maximal stability, maximal resistance to replacement by borrowing, maximal conceptual simplicity, and maximal universality.

This ranking is not representative of the optimal meaning list for use in the initial stages of language comparison, owing to the limited number of meanings for which stability was determined (working, as discussed above, from the bottom-up, rather than the top-down). Rather, the ranking determines the input items'

relative optimality for use in the initial stages of the comparative method, thus allowing for cross-linguistic comparison of the optimality of the input meanings. This integration also allows a more meaningful discussion regarding the relationship between item stability and item borrowability, simplicity, and universality to now be entered.

The stability scores, integrated with Tadmor et al.'s item borrowed, analyzability, and representation scores, are given in Table 8. The composite score, by which the items are ranked, is the product of the four scores.

Table 8: The integration of the stability scores with Tadmor et al.'s (2010) item borrowed, analyzability, and representation scores

Rank	**Meaning**	**Borrowed score**	**Analyzability score**	**Representation score**	**Stability score**	**Composite score**
1	I	0.97	0.936	0.976	0.983	0.871
2	nose	0.973	0.98	1	0.911	0.869
3	you	0.958	0.933	1	0.956	0.854
4	tongue	0.934	0.954	1	0.938	0.836
5	name	0.915	0.955	1	0.942	0.823
6	water	0.909	0.987	1	0.9	0.807
7	root	0.944	0.973	1	0.876	0.805
8	fire	0.965	0.995	1	0.837	0.804
9	ear	0.896	0.961	1	0.932	0.803
10	tooth	0.882	0.975	1	0.932	0.801
11	to go	0.963	0.974	1	0.851	0.798
12	to come	0.968	0.94	1	0.872	0.793
13	fly	0.948	0.942	1	0.88	0.786
14	eye	0.904	0.904	1	0.958	0.783
15	to stand	0.981	0.889	1	0.887	0.774
16	bone	0.918	0.971	1	0.865	0.771
17	to hear	0.953	0.895	1	0.9	0.768
18	mouth	0.92	0.982	1	0.847	0.765

Continued on next page...

The integration of the stability scores with Tadmor et al.'s (2010) item borrowed, analyzability, and representation scores – Continued

Rank	Meaning	Borrowed score	Analyzability score	Representation score	Stability score	Composite score
18	blood	0.904	1	1	0.846	0.765
20	one	0.87	0.969	1	0.906	0.764
21	new	0.92	0.92	1	0.897	0.759
22	salt	0.848	0.976	1	0.916	0.758
23	bitter	0.975	0.889	1	0.871	0.755
24	to bite	0.964	0.887	1	0.877	0.75
25	shade	0.887	0.931	1	0.907	0.749
26	liver	0.869	0.967	1	0.89	0.748
27	to eat	0.92	0.925	1	0.877	0.746
28	house	0.893	0.969	1	0.859	0.743
29	night	0.931	0.934	1	0.848	0.737
30	to give	0.913	0.907	1	0.884	0.732
31	rain	0.916	0.95	1	0.835	0.727
32	foot	0.856	0.972	1	0.856	0.712
33	it	1	0.955	0.878	0.843	0.707
33	fish	0.855	0.984	1	0.84	0.707
35	wind	0.828	0.987	1	0.864	0.706
36	egg	0.91	0.945	1	0.818	0.703
37	hard	0.918	0.903	1	0.842	0.698
38	to know	0.933	0.908	1	0.823	0.697
39	sand	0.901	0.928	1	0.832	0.696
40	leg	0.856	0.972	1	0.833	0.693
41	to do	0.947	0.914	1	0.791	0.685
42	to tie	0.879	0.948	1	0.82	0.683
43	black	0.951	0.899	1	0.797	0.681
43	star	0.83	0.97	1	0.846	0.681
45	big	0.889	0.98	1	0.774	0.674

Continued on next page...

The integration of the stability scores with Tadmor et al.'s (2010) item borrowed, analyzability, and representation scores – Continued

Rank	Meaning	Borrowed score	Analyzability score	Representation score	Stability score	Composite score
46	dog	0.838	0.96	1	0.832	0.669
47	long	0.956	0.898	1	0.776	0.666
48	to fall	0.946	0.903	1	0.779	0.665
49	to see	0.918	0.9	1	0.803	0.663
50	hair	0.944	0.917	1	0.761	0.659
51	thick	0.95	0.906	1	0.759	0.653
51	old	0.896	0.92	1	0.792	0.653
53	small	0.909	0.966	1	0.741	0.651
54	to say	0.972	0.928	1	0.721	0.65
55	wood	0.86	0.94	1	0.803	0.649
56	to burn	0.951	0.889	1	0.765	0.647
57	to take	0.9	0.887	1	0.808	0.645
58	meat	0.877	0.986	1	0.739	0.639
58	wide	0.955	0.885	1	0.756	0.639
60	to grind	0.919	0.886	1	0.784	0.638
61	to carry	0.919	0.953	0.976	0.734	0.627
61	to hide	0.928	0.913	1	0.74	0.627
63	child	0.929	0.93	0.976	0.733	0.618
64	good	0.893	0.945	1	0.696	0.587

In order to determine whether the inclusion of the more detailed quantification of item stability as described here significantly altered the quantification of the optimality of an item for use in the initial stages of the comparative method, a paired-samples t-test was conducted to compare the composite scores presented in Tadmor et al. (2010) with those presented here. A significant difference was found between the means reported in Tadmor et al. (mean=0.745, standard

deviation=0.045) and those given here (mean=0.722, standard deviation=0.068); $t(63)=2.497$, $p < 0.01$. This indicates that the integration of the more detailed quantification of item stability which has been discussed here has a significant effect on the quantification of the optimality of items for use in the initial stages of the comparative method.

4.2.1 The relationships between item stability and item borrowability, simplicity, and universality

As was discussed throughout section 2, the relationships between item stability and item borrowability, conceptual simplicity, and universality are often intuitively or implicitly assumed, without recourse to statistical measures of these relationships. However, with the detailed quantification of item stability, borrowability, conceptual simplicity, and universality now readily available to us, we are in a position to be able to determine whether the correlations so often taken for granted are indeed supported by the figures.

4.2.1.1 Item stability and item borrowability

In section 2.2, we highlighted the subtle but crucial difference between item stability and item borrowability: while the latter refers to the likelihood that the form representing an item is replaced a form borrowed from another language, the former refers to the likelihood that the form representing an item is replaced by any means, either by borrowing, or by language-internal processes such as the semantic shift of a form or neologism. Frequently in the literature, however, item stability and borrowability are conflated (e.g. Dolgopolsky 1986; Lohr 1999; Wang and Wang 2004). While complete conflation of the two features is unwarranted, it does nonetheless seem intuitively a valid assumption that item stability and item borrowability should correlate positively at statistically significant level – as an increase in the number of replacements by borrowing will lead to an increase in overall replacements, our hypothesis is that the more likely the form representing an item is to be replaced by borrowing, the more likely the form is to be replaced by any means, i.e. those items which are less resistant to replacement of form

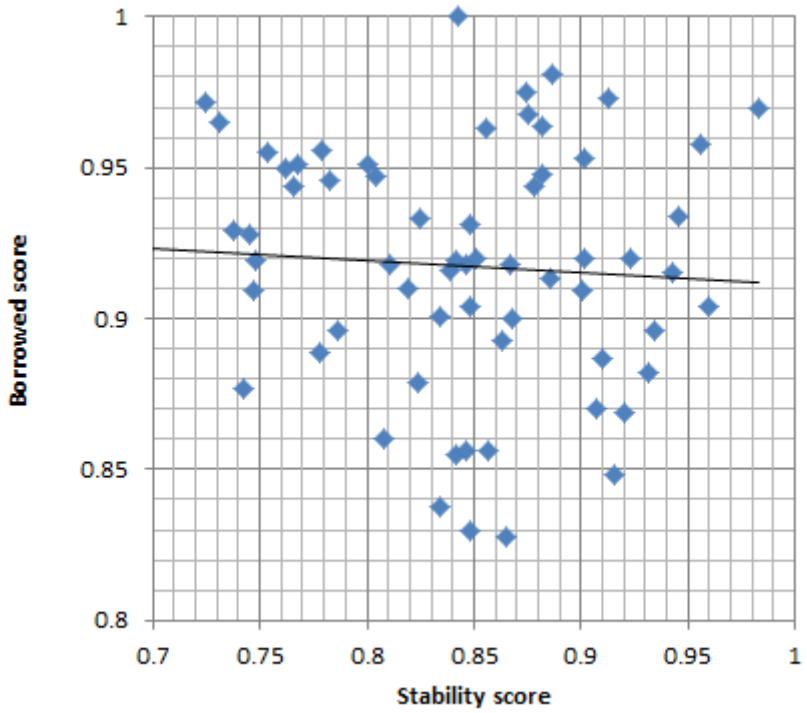


Figure 3: The relationship between the stability score and Tadmor et al.'s borrowed score (2010), r= −0.026, n=64, p > 0.1

by borrowing are also less stable, and therefore those that are more resistant to replacement of form by borrowing are more stable.

Therefore, in response to the conflation of item stability and item borrowability in the literature, as well as the reasonable intuition that there is a significant positive relationship between item borrowability and stability, a Pearson product-moment correlation coefficient was calculated between item stability as presented here, and item borrowability as presented in Tadmor et al. (2010). There was found to be no significant correlation between the stability scores and the borrowed scores, r= − 0.026, n=64, p > 0.1. This result indicates that we cannot reject the null hypothesis that there is no significant relationship between item stability and item borrowability; there is no support for the hypothesis that there is a significant

positive relationship between item stability and borrowability. The results are summarised in Figure 3.

4.2.1.2 Item stability and item simplicity

Tadmor et al. (2010) quantify an item's simplicity by determining how analyzable the forms by which it is represented are cross-linguistically. They assume a link between item simplicity and the analyzability of a meaning's form (2010:238); those items which are more conceptually complex, such as 'the native country', tend to be represented by more analyzable forms (analyzability score 0.58) than those items which are less conceptually complex, such as 'the country' (analyzability score 0.91; Haspelmath and Tadmor 2009). As the link between the analyzability and the age of a form is often made in introductory textbooks to historical and comparative linguistics (forms which are less analyzable tend to be older, while more transparent, analyzable forms tend to be more recent innovations – see, for example, Campbell 1998:276) we would expect to see a significant positive correlation between items' stability scores and their analyzability scores – the more stable an item is, the more unanalyzable the form by which it is represented it will tend to be.

To explore the relationship between item stability and analyzability of form, a Pearson product-moment correlation coefficient was calculated between the stability scores and Tadmor et al.'s analyzability scores. There was found to be a marginally significant positive correlation, r= 0.203, n=64, $p < 0.1$. The results are summarised in Figure 4. This result somewhat supports the hypothesis that the more stable an item is, the more analyzable the form by which it is represented will tend to be. If, as Tadmor et al. (2010) assume, there is a link between forms' analyzability and their counterpart items' simplicity, this would also suggest a somewhat significant positive correlation between item stability and item simplicity. We will, however, return to discuss the legitimacy of this assumption in section 5.

4.2.1.3 Item stability and item universality

The assumed link between item stability and item universality may be demonstrated by two quotes from Swadesh: "Gradually the evidence accumulated showing that *universal* everyday vocabulary... changes at a roughly constant rate"

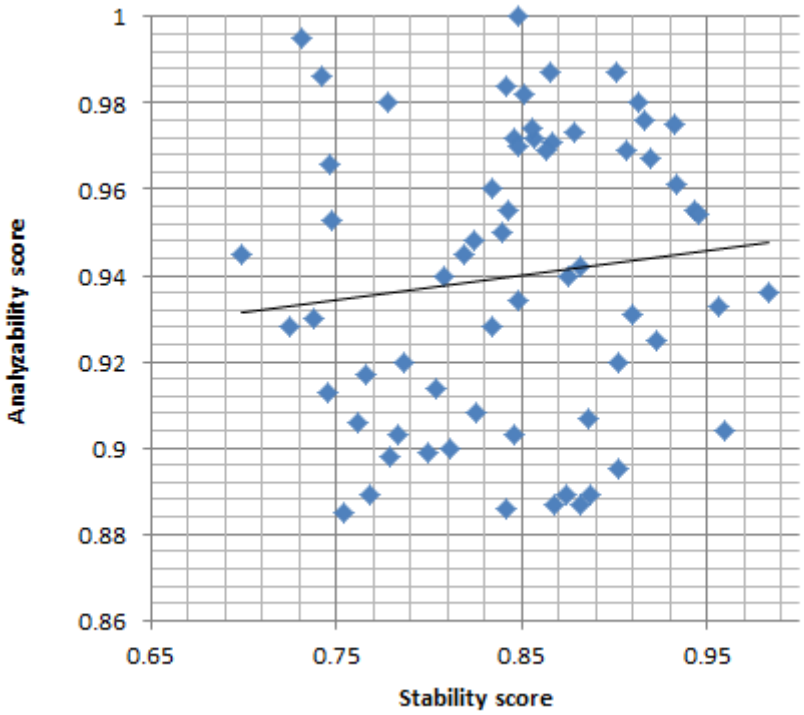


Figure 4: The relationship between the stability score and Tadmor et al.'s analyzability score (2010), r= 0.203, n=64, $p < 0.1$

(1952:455); "This approach [glottochronological dating using the 'index of time' of basic vocabulary] should prove fairly dependable... because it is a well known fact that certain types of morphemes are relatively *stable*" (1950:157; emphasis added to both quotes). These quotes demonstrate Swadesh's assumption that universal meanings change at a constant rate, and that meanings that change at a constant rate are also the most stable items; logically, therefore, we can infer that he considered more universal items to be more stable.

This assumption can be tested statistically, by calculating a Pearson product-moment correlation coefficient between the stability scores and Tadmor et al.'s representation scores. There was found to be a no significant correlation, r= 0.021, n=64, $p > 0.1$. The results are summarised in Figure 5. This outcome does not support the hypothesis that the more stable an item is, the more widely the item will

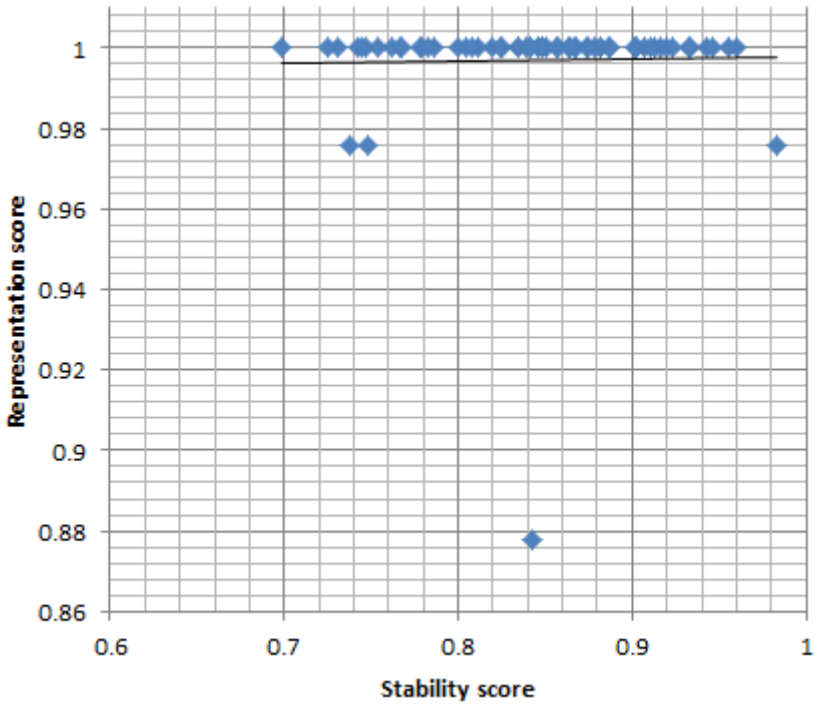


Figure 5: The relationship between the stability score and Tadmor et al.'s representation score (2010), r= 0.021, n=64, p > 0.1

be represented cross-linguistically. However, it should be noted that the majority of the items under consideration here (93.75%) were maximally represented in the language sample in Haspelmath and Tadmor (2009); a selection of items with a wider spread of representation scores might reveal a significant relationship.

4.3 Coda: On the glottochronological constant

With the detailed quantification of item stability, used in the determination of our input items' relative optimality for use in the initial stages of the comparative method, now available to us, it would be amiss not to take this opportunity to use the stability scores to statistically explore the validity of the fundamental assumption of glottochronological theory: that the rate at which the forms for

items are replaced is constant (Gudschinksy 1956; Lees 1953; Swadesh 1950, 1952, 1955).

Using a two-way analysis of variation without replication, we can examine whether there is a significant difference in the stability scores both between meanings and between language families. A significant difference in the stability scores from language family to language family would indicate that there is neither one constant rate of change for all languages and all meanings, nor an individual rate of change for each meaning which is constant across languages (as suggested by, for example, Starostin 2000). A significant difference in the stability scores from meaning to meaning would indicate that there is not an individual rate of change for each language family which is constant for all the basic meanings.[28]

The results of the two-way analysis of variation showed that there is a very significant difference between the stability scores for each language family, $F(3, 50)=15.462$, mean square error=0.004, $p < 1 \times 10^{-8}$, and that there is an extremely significant difference between the stability scores for each meaning, $F(3, 50)=4.866$, mean square error=0.004, $p < 1 \times 10^{-13}$. We can therefore confidently reject the hypotheses that there is an overall constant rate of change, that there is an individual rate of change for each meaning which is constant across languages, and that there is an individual rate of change for each language family which is constant across meanings. The results show that there is a significant difference between the stability scores, both from language family to language family, and from meaning to meaning. Our results simply do not support the assumption that the forms by which basic items are represented are replaced at constant rates.

[28] It was necessary to remove those meanings represented in only three of the four language families to carry out the two-way analysis of variance; this calculation is thus based on the data for 51 meanings. The items removed were 'bitter', 'to carry', 'fly', 'foot', 'to give', 'to go', 'to grind', 'hard', 'it', 'shade', 'to take', and 'wood'.

It should once again be noted that the meanings used in this investigation are not necessarily representative of the definitive list of basic meanings; the glottochronological hypothesis states that it is only in the most basic meanings that the forms are replaced at a constant rate. Nevertheless, the meanings used here were independently identified as basic by both Lohr (1999) and Tadmor et al. (2010); we can therefore be confident that they are reasonably representative.

4.4 Summary of results

This section has presented the stability scores of our input meanings, ranked from most stable to least stable, based on the mean probability across the four language families that the form by which the item is represented will be retained for a thousand years. Using these results, the relationship between item stability and the variation in stability between language families could be explored: it was shown that there is a significant negative correlation between the stability scores and the standard deviation of these scores, i.e. that the more stable an item is, the less this stability varies across the four language families.

Once item stability was calculated for the input meanings, the scores were incorporated with Tadmor et al.'s (2010) scores quantifying item borrowability, conceptual simplicity, and universality, to determine a composite score. This composite score is similar to that used in the formation of the Leipzig-Jakarta list, but takes a more detailed picture of item stability into account – it is therefore maximally founded on the four criteria identified in section 2.2 as necessary to facilitate the implementation of the initial stages of the comparative method. While the list of the input meanings ranked by their composite scores presented in Table 8 does not necessarily represent the meanings which are definitively optimal for use in the initial stages of the comparative method, owing to the bottom-up approach adopted here, it does rank the items investigated by their relative optimality, thereby demonstrating the potential fruitfulness of a more comprehensive investigation into item stability to combine with the borrowed, analyzability, and representation scores of Tadmor et al. (2010).

With the detailed scores for item stability, borrowability, simplicity, and universality available, the relationship between item stability and the three remaining features was analyzed. It was shown that, contrary to expectations, there is no significant relationship between item stability and item borrowability or item representation, and only a marginally significant positive correlation between item stability and analyzability of form. Finally, these results were used to assess the glottochronological tenet that the forms representing items are replaced at a constant rate; our results strongly support the hypothesis that there is a significant difference in the rates of replacement of both meanings and language families, thus

rejecting this tenet. In order to fully interpret these results, the possible reasons underlying them must be thoroughly examined; it is to this discussion that we now turn.

5 Discussion

This dissertation began by identifying areas of concern in the implementation of the initial stages of the comparative method. The lack of a widely-used, objectively-derived standardised meaning list of items was highlighted; it was argued that, in order to facilitate the implementation of the comparative method and produce more reliable results, such a list should comprise those items which are maximally stable, resistant to borrowing, conceptually simple, and universal. In our pursuit of a means of determining the optimal meaning list, we have had the opportunity to quantify the stability of 67 meanings, by determining the probability that the form by which they are represented will be retained for a thousand year timespan.

The careful separation out of the feature of item stability has allowed us to explore questions regarding the relationship between item stability and the extent to which this stability varies in different language families; the relationship between item stability and the features of borrowability, simplicity, and universality; and the glottochronological assumption of a constant rate of change. More importantly, it has also allowed us to integrate a more detailed quantification of item stability into the composite scores given in Tadmor et al. (2010), thus ranking the input meanings for their optimality for use in the initial stages of language comparison; this methodology has therefore been demonstrated as a viable means of identifying those meanings which are most basic in the context of the comparative method. The results of these investigations, presented in the previous section, have highlighted several areas for discussion, particularly regarding the lack of support for a strongly significant relationship between item stability and item borrowability, simplicity, and universality, as well as the rejection of the hypothesis of the glottochronological constant. As this is the first time that the issues discussed here have been investigated statistically in this way, this section will explore in detail the possible interpretations and implications of the results of this dissertation.

The initial ranking of the input items by stability alone allows for a comparison of the items' stability relative to one another. However, as been stressed throughout this dissertation, owing to the bottom-up determination of item stability, these items are not necessarily representative of those items which are most stable cross-linguistically. Despite this, however, a review of the ranking of items by their

stability given in Table 7 reveals that those items which are ranked highest are those which are frequently identified in textbooks (such as Campbell 1998:112 or Trask 2000:39) as those which are most resistant to replacement – pronouns ('I', 'you (sg.)') and terms for body parts ('eye', 'tongue', 'ear', 'tooth', 'nose'), for example, feature heavily in the ten most stable of the 64 meanings. Unfortunately, however, the bottom-up means of determining item stability precludes any meaningful discussion about the composition of the list in terms of semantic fields or parts of speech represented. No meanings have been excluded from our final ranking of stability because of low stability, i.e. all the input meanings, aside from those excluded owing to the potential for representation by an onomatopoeic form, or because of lack of representation in the language families, are present in Table 7. A discussion regarding the representation of semantic fields or parts of speech in this ranking would therefore be exploring nothing more than the composition of the list of input meanings.

Nonetheless, the quantification of stability has allowed us to explore important questions regarding the probability that a form representing a meaning will be retained. The relationship between item stability and the extent to which this stability varies between language families was examined, and a significant negative correlation was calculated, supporting our hypothesis that the more stable an item is, the less this stability varies. However, it should be noted that this result is not necessarily surprising; as detailed in Hedges (1981), means and standard deviations – representing here the stability score and the variation in stability respectively – will tend to correlate in this way in some circumstances, owing to the arithmetic involved.[29] Nevertheless, that the values here followed this tendency, rather than showing no significant correlation or a significant correlation in the reverse, indicates that this result is still meaningful.

The integration of the stability scores with Tadmor et al.'s (2010) borrowed, analyzability, and representation scores to give a composite score founded on the detailed quantification of the four criteria identified as necessary to determine the optimal meanings for use in the initial stages of the comparative method, like the ranking of items by stability, gives a picture of the input items' optimality relative to one another. Once again, however, like the ranking of items by stability,

[29] I am grateful to Dr. Quentin Atkinson for pointing this out to me.

this list does not necessarily definitively identify the optimal meanings for use in the initial stages of the comparative method; nor is it appropriate to explore the semantic fields and parts of speech represented in the list, as these will again be equal to the input meanings. Such investigations will have to await future, more comprehensive research. However, the input meanings are those which had already been independently, empirically assessed by both Lohr (1999) and Tadmor et al. (2010) to be at least among the most basic; the ranking of these meanings, therefore, should give a reasonable idea of which meanings should be expected to be maximally optimal for use in the initial stages of the comparative method.

Of our assessments of the relationships between item stability and item borrowability, simplicity, and universality, perhaps the most surprising and counter-intuitive result is that there appears to be no significant positive correlation between item stability and item borrowability. As discussed in section 4.2.1.1, as replacement of an item's form by borrowing increases the frequency with which the item's form is replaced by any means, we would expect to see an increase in items' stability scores as the borrowability scores increase. There is, however, no evidence to support this expectation. There are three interpretations (or some combination of these three) of this result: 1) The quantification of item stability presented here produces inaccurate results; 2) The quantification of item borrowability presented in Tadmor et al. (2010) produces inaccurate results; 3) There is simply no correlation between item stability and item borrowability, at least among those items identified by Tadmor et al. (2010) and Lohr (1999) as among the most basic.

There is no reason to accept either of the first two interpretations as the underlying explanation of the results presented in this dissertation. Both the study described here, and that described in Tadmor et al. (2010), have been carried out transparently, based on cross-linguistic data, using methodologies fully appropriate to quantify item stability and borrowability respectively. We therefore must accept the third interpretation: There is no significant relationship between item stability and item borrowability. Indeed, counter-intuitive as this conclusion may be, it certainly underlines the necessity of the distinction made between stability and borrowability in section 2.2; borrowability and stability are clearly two separate features of an item. Tadmor et al. in fact reach the same conclusion, stating: 'It thus seems that some meanings may be subject to change... but not so much subject

to borrowing' (2010:242). The implication of this conclusion is that any future conflation of stability and borrowability should be viewed as not only unwarranted, but careless; we can no longer simply assume that those items which are most stable are also maximally resistant to replacement of form by borrowing, as have, for example, Dolgopolsky (1986) and Lohr (1999). The careful separation out of these two features of an item is clearly necessary, not only when considering those items optimal for use in the initial stages of the comparative method, but in all considerations of criteria for creating a standard meaning list in historical and comparative linguistics, lexicostatistics included.

As was shown in section 4.2.1.2, there is a somewhat significant positive correlation between the stability scores and Tadmor et al.'s (2010) analyzability scores. The relationship between item stability and item simplicity, however, rests on the validity of the assumption made by Tadmor et al. that those items which are more conceptually simple tend to be represented by forms which have a higher analyzability score (i.e. are less analyzable), and *vice versa*. For example, Tadmor et al. cite 'younger sister' as a prototypical example of a meaning which tends to be represented by more analyzable forms (with an analyzability score of 0.68; 2010:236); while the link is not discussed explicitly, their indication that the analyzability score is representative of the simplicity of an item (2010:238) suggests they consider those items which tend to be represented by more complex forms to also be more complex conceptually.

While this may well be a tendency, it is important to note that analyzability of form is not equal to simplicity of concept; furthermore, this assumption is apparently based on intuition. Indeed, counter-examples of the assumption can be found in Haspelmath and Tadmor (2009), the database on which the Leipzig-Jakarta list is based. In direct contradiction to their 'younger sister' example, for instance, we find the meanings 'younger sibling', 'older sibling', and 'sibling-in-law' all tend to be represented by forms which are less analyzable (with analyzability scores of 0.81, 0.77, and 0.85 respectively) than the conceptually simpler 'sibling' (with an analyzability score of 0.66). Further counter-examples can be found: for example, while Tadmor et al. also cite 'day after tomorrow' (0.66) as a meaning which tends to be represented by analyzable forms (2010:236), and thus presumably that they consider it to be more conceptually complex than, for example, 'today' (0.78), the

meanings ‘yesterday’ and ‘tomorrow’ (0.92 and 0.89 respectively), both by the same logic more conceptually complex than ‘today’, have higher analyzability scores, i.e. tend to be represented by forms which are less analyzable.

The link between item stability and analyzability of form is not surprising, for the reasons discussed in section 4.2.1.2: the older a form is, the more time there has been for phonological and morphological change, coupled with semantic shift, to act and interact on the lexeme, resulting in a less analyzable form – and a more stable item is by its very nature more likely to be represented by an older form. What is questionable, however, is whether this correlation between item stability and analyzability of form can be mapped wholesale onto the relationship between item stability and item simplicity. Indeed, a means of cross-linguistically quantifying conceptual simplicity across the fluid, ever-changing and, in many (if not all) cases, untransposable conceptual domains from one language to the next may potentially prove impossible. The determination of the validity of the assumption that analyzability of form and item simplicity relate in a significant manner will have to await future research; the results presented here, however, regarding the relationship between item stability and item borrowability should serve as a warning against simply assuming the existence of a relationship. For now, therefore, we can safely say no more than there exists a somewhat significant positive correlation between item stability and the analyzability of form of an item between those meanings identified by both Lohr (1999) and Tadmor et al. (2010) as basic.

The final relationship investigated, between the stability scores and Tadmor et al.’s representation scores (2010), again revealed there to be no significant correlation, again contrary to assumptions made in the literature. There are two points to note here, however. First, as discussed in section 2.4.2, the representation of item universality is based on data drawn from 41 languages from 26 different language families; while these languages are drawn from a cross-linguistic spread, a larger number of input languages would allow for a more confident generalisation of results regarding universality. Second, we indicated in section 4.2.1.3 that almost all of the meanings assessed for stability in this dissertation were fully represented in Haspelmath and Tadmor (2009), and therefore are maximally scored for representation in Tadmor et al. (2010). Only considering those meanings which have apparently both a very high level of stability and universality may perhaps

obscure what is actually a significant relationship between item stability and item universality; in order to fully accept the conclusion reached here that there is no significant relationship, it is necessary to calculate the correlation between items with a wider range of stability and representation scores.

Indeed, these two points apply equally to the relationship between item stability and borrowability and item stability and analyzability. More information is, of course, always preferable when calculating the mean average of a dataset; the law of large numbers states that the more data we include, the closer our average will be to the expected value. If we wished to improve the accuracy of the results presented here, future research could focus on expanding the number of languages in which the set of 1460 meanings are assessed for stability, borrowability, analyzability (and thus perhaps simplicity), and universality – however, the "considerable amount of time and effort" required by a specialist in each language to complete this task reported in Tadmor et al. (2010:229) should be noted. More importantly, however, in order to gain a more realistic picture of the relationships between item stability and item borrowability, analyzability of form, and item universality, items at all levels of the lexicon, more basic and less basic, should be investigated. While there appears to be no significant correlation between item stability and item borrowability and universality in the meanings identified by Lohr (1999) and Tadmor et al. (2010) as basic, this may not be the case across wider span of the lexicon.

The restriction of the determination of the relationships between item stability and item borrowability, analyzability, and universality to those items independently identified by Lohr (1999) and Tadmor et al. (2010) as among the most basic is, as has been emphasised throughout this dissertation, a consequence of the bottom-up means of list-formation employed here, itself a result of the timescale of and resources available for this project. Future research must therefore concentrate on determining item stability in a top-down manner, starting with a wide range of meanings – for example, the 1460 used in Haspelmath and Tadmor (2009) – and calculating the item stability for each. This would allow for both the incorporation of a detailed quantification of item stability in a composite score, and thus the refinement of these meanings to those which are optimal for use in the initial stages of the comparative method, as well as the assessment of the relationship between

the four features across a wider variation of basicness of meaning. The scale of such a project, however, should not be underestimated. In addition, the determination of stability of all of the meanings looked at in Haspelmath and Tadmor (2009) may pose problems in collecting a cross-representative spread of etymological data necessary to determine item stability in the means presented here – particularly as Greenhill et al. (2008), the Austronesian resource used in this project, contains etymological information for only 210 meanings, thus effectively eliminating data from Austronesian for 1250 of the meanings. This project, therefore, to maintain cross-linguistic applicability of results, must await a wider range of more detailed etymological dictionaries of the kind described here to become publicly available.

It was suggested in section 2.1 that the determination of those items optimal for use in the initial stages of the comparative method would reduce the number of meanings to be considered in explorations concerning other areas of unnecessary subjectivity in the initial stages of the comparative method, thus narrowing the field of investigation and allowing for a more focussed, precise examination. The other areas of subjectivity identified in the application of the initial stages of the comparative method are the determination of what should be considered an 'expected' semantic shift, and the determination and exclusion of those meanings which are most likely to be represented by forms which are similar cross-linguistically, due to onomatopoeia, sound symbolism, or a tendency towards forms based on nursery formations. While the investigation of semantic shift in the context of the comparative method would represent a radical departure in methodology from that discussed here, the investigation of universal similarities of form could, in fact, utilise further the concept of the composite score, used to rank the meanings in this dissertation for their optimality for use in the comparative method. Similar to the way in which items were scored for age, likelihood of having been borrowed, analyzability, and representation in Haspelmath and Tadmor (2009), language specialists could score the form(s) of each item by the extent to which they consider it to have an onomatopoeic, sound symbolic, or nursery formation origin. This score can then be incorporated into the composite score with ease; therefore, less weight will be given to those items which tend to be represented by forms which are universally similar, and thus these items are less likely to be identified as optimal for use in the initial stages of the comparative method. Such research would allow

for the empirical determination of meanings likely to be represented by such forms, thus removing the subjectivity with which such meanings are currently identified.

A potential confounding variable which has not hitherto been mentioned warrants discussion: the role played by the input languages in the determination and comparison of the stability, borrowability, analyzability, and representation scores. The input languages used here to determine item stability are largely different from those used in Haspelmath and Tadmor (2009) to determine item borrowability, analyzability, and item universality. It is therefore feasible that the relationships observed and reported here are simply a result of the relationships between item stability and item borrowability, analyzability of form, and item universality of the particular languages from which data were drawn; should this be the case, the relationships observed cannot be extended cross-linguistically. It would of course be desirable to eliminate this confounding variable in future research, by either expanding the determination of item stability to the language families whose inclusion will encompass all the languages represented in Haspelmath and Tadmor (2009) (although the lack of the etymological resources currently available for many of the language families concerned currently precludes this path), or by determining item borrowability, analyzability of form, and item universality for the meanings in all the languages used to gather data to assess items for stability. However, while the potential for this confounding variable having affected the results is certainly a concern, it should be noted that it would not have been possible to analyse item stability in the detailed means presented here for the 41 languages used in Haspelmath and Tadmor (2009), as this method is dependent on data coming from a range of language families, rather than the individual languages used therein. More importantly, there is a large enough spread of languages used in both studies to suggest that the results can be applied cross-linguistically, and should be at least moving towards the expected values in both cases. Thus, while the potential presence of this confounding variable must be acknowledged, we would not expect it to have had too significant an effect on the results presented here.

Our final investigation, using the stability scores calculated in the process of determining the input items' relative optimality for use in the initial stages of the comparative method, assessed the validity of the glottochronological hypothesis that the forms representing items are replaced at a constant rate. This investigation

proved to produce results which are extremely significant for glottochronological theory, as well as historical and comparative linguistics more generally: no evidence whatsoever was found to support the hypothesis of there being a constant rate of change. Furthermore, no support was found for the supposition that each language family has an individual rate of change which is constant across all basic meanings, nor that each meaning has an individual rate of change which is constant across all language families. The evidence presented here is simply incompatible with both the theory that the forms for two different meanings within a language family have the same probability of being replaced, and the theory that the forms for the same meaning in two language families have the same probability of being replaced. It should be noted, however, that these results are not incompatible with Embleton's suggestion (1986) that the rate of replacement varies both from meaning to meaning *and* from language to language; nor does it contradict Kruskal et al.'s findings (1973) that the replacement rates vary in different areas of a family tree, both synchronically and diachronically. However, such apparent variability in the rate of replacement clearly undermines the very foundations of traditional glottochronology; the results of this dissertation strongly recommend that all dates of language separation based on traditional glottochronological methods be treated with extreme caution until independently verified by extra-linguistic evidence, or unless a novel model of dating language splits is developed.

In this section, we have shown how the ranking of the input items by their stability scores sheds interesting light on the items' stability relative to one another. However, with the results presented in this dissertation, we can neither confirm that these items are maximally stable cross-linguistically, nor enter a meaningful discussion regarding the semantic fields or parts of speech represented; we have identified the same issues with regard to the ranking of the input items by their composite score, representative of their optimality for use in the initial stages of the comparative method. Nevertheless, we have demonstrated the viability of this method for quantifying item stability, and thus, with the integration of the stability scores with Tadmor et al.'s (2010) borrowed, analyzability, and representation scores, for determining how basic an item is in the context of the comparative method. Its application to other meanings is strongly encouraged; such an application will allow for the assessment of a wider range of meanings, and thus, ultimately, the

cross-linguistic definition of those meanings are optimal for use in the initial stages of the comparative method. This definition will represent a substantial step towards curtailing the unnecessary areas of subjectivity currently present in the application of the comparative method.

In addition to ranking these items by their relative stability and optimality for use in the initial stages of the comparative method, we have used the stability scores to statistically explore questions relevant item basicness. We have shown that there is a significant relationship between item stability and the variation in this stability between language families, and have concluded that, while the correlation between means and standard deviations presented as evidence to support this is a mathematical tendency, this does not wholly detract from the meaningfulness of the result presented. The relationships between item stability and item borrowability, analyzability of form, and item universality have been explored; no strongly significant relationship was found between item stability and any of the other three features of an item. While there is no reason to believe that the lack of a significant correlation is due to inadequacies in the methodologies used to calculate the scores, it was noted that we are only considering here those items which are candidates for maximal basicness. The restriction of the level of basicness from which the input meanings were drawn in this dissertation was presented as a further argument to recommend future research considering the stability of a wider range of input meanings, in order to test the conclusions reached here against items from a broader spectrum of basicness. We have also identified the lack of support for the conflation of item stability and item borrowability, suggesting that such unwarranted identification of one with the other consequently be avoided. Tadmor et al.'s (2010) assumption that item simplicity and analyzability of form are analogous has been explored; it was found to be in need of further investigation. Finally, we have discussed the implications of the results of our exploration of the glottochronological constant for both glottochronological theory and historical and comparative linguistics more generally: as our results provide no support whatsoever for the constant rate of replacement in basic meanings, it is the strong recommendation of this dissertation that any results reached using traditional glottochronological methodologies be rejected, unless triangulated with evidence from other disciplines.

6 Conclusion

The purpose of the research described in this dissertation has been to take a step towards the definition of the optimal standard meaning list for use in the initial stages of the comparative method. This research has been embarked upon in order to curtail one of the areas of unnecessary subjectivity currently present in the only methodology widely accepted as capable of providing support for hypotheses of genetic relationship between languages. In the course of this research, we have quantified the stability of 67 items, using a methodology which has produced more detailed, objective, and cross-linguistically applicable results than any other study hitherto carried out. This has enabled the incorporation of a more accurate picture of item stability with Tadmor et al.'s (2010) borrowed, analyzability, and representation scores, thus allowing us to rank the input items by a composite score fully representative of the four features identified as necessary for a meaning to be considered optimal for use in the initial stages of language comparison: maximal item stability, maximal resistance to replacement of form by borrowing, maximal conceptual simplicity, and maximal universality.

The stability scores gained in the process of this dissertation have also allowed us to explore other issues related to the basicness of meanings: the relationship between item stability and the variation in stability between language families; the relationship between item stability and item borrowability, analyzability of form, and item universality; and the determination of the validity of the glottochronological hypothesis of a constant rate of replacement of form in basic meanings. We have demonstrated that the more stable an item is, the less the item will vary in stability from language family to language family. It has also been shown that, contrary to expectations and conflations evident in the literature to date, there is no significant relationship among the meanings considered between item stability and item borrowability, nor between item stability and item universality. There is evidence to suggest a marginally significant positive relationship between item stability and analyzability of form; should Tadmor et al.'s (2010) assumption regarding the relationship between analyzability of form and conceptual simplicity prove correct, this would presumably indicate some kind of positive relationship between item stability and conceptual simplicity. Finally, we presented evidence to

conclusively reject the glottochronological hypothesis that the forms representing basic tmeanings are replaced at a constant rate, as well as the hypotheses of an individual rate of change for each meaning which is constant across languages and an individual rate of change for each language family which is constant for all the basic meanings. The significance of these results cannot be overestimated; using replicable and transparent means, we have highlighted confusions, conflations, and assumptions present in the literature regarding basic meanings, and have statistically demonstrated them to be unfounded.

We introduced this dissertation by stating that, if historical and comparative linguistics is to progress as a scientific discipline, the areas of unnecessary subjectivity in the comparative method must be identified, explored, and ultimately resolved. Regarding the area of subjectivity we suggested should be our priority, the lack of a standardised list of meanings selected to maximally facilitate the implementation of the comparative method, we can certainly say that the first two stages, identification and exploration, have indeed been thoroughly addressed in this dissertation; nonetheless, the optimal meaning list for use in the initial stages of the comparative method has yet to be defined. While the issues regarding the lack of a basic meaning list designed specifically for use in the comparison of languages for evidence of a genetic relationship have been scrutinised to the full extent of this dissertation, and the methodology required to achieve such a task has been clearly and transparently identified, described, and demonstrated, the full resolution of this area of subjectivity must await future research.

In particular, the application of this means of quantifying item stability to a wider range of meanings – preferably all of the 1460 meanings considered in Haspelmath and Tadmor (2009) – is strongly encouraged. This will allow for the presentation of the definitive list of meanings optimal for use in the initial stages of the comparative method: as the detailed quantifications of stability are incorporated into items' composite scores, those meanings with the highest composite score can be refined, thus identifying those meanings which are maximally stable, maximally resistant to borrowing, maximally conceptually simple, and maximally universal, and are therefore optimal for use in the initial stages of the comparative method. It would also enable the determination of whether the results presented here regarding the relationship between item stability and the variation in this stability, as well as

the relationships between item stability and item borrowability, analyzability of form, and item universality, are supported with evidence from less basic items as well as more basic items. The likelihood that a meaning will be represented by a form which tends to have cross-linguistically similar origins such as onomatopoeia or nursery formation must also be investigated; we have suggested the potential for the incorporation of this feature of an item into its composite score. Finally, with the identification of those meanings which are optimal for use in the initial stages of the comparative method, the specific area for focussed research exploring attested and reconstructed semantic shifts of forms will be defined, allowing for the rigorous examination of the final area we identified as unnecessarily subjective in the implementation of the initial stages of the comparative method.

While, as has been stated, the scale of such future research should not be underestimated, our responsibilities as linguists, and therefore as scientists, cannot be shirked. One of the main themes of this dissertation has been the demonstration of the inadequacy of simply assuming the existence of phenomena, characteristics, and relationships. The results presented here should warn against any such unwarranted conjecture in future research, in both historical and comparative linguistics, and linguistics more generally. To continue to push back the boundary between the known and currently unknown, the objective assessment of all well-founded speculation is to be encouraged; only in this way will we refine and develop our understanding of the world in which we live. It is hoped that this dissertation represents a positive step in this direction.

A Standardised meaning lists

A.1 The Swadesh lists

A.1.1 100-item Swadesh list (1972:285)

1. I
2. you
3. we
4. this
5. that
6. who
7. what
8. not
9. all
10. many
11. one
12. two
13. big
14. long
15. small
16. woman
17. man
18. person
19. fish
20. bird
21. dog
22. louse
23. tree
24. seed
25. leaf
26. root
27. bark
28. skin
29. flesh
30. blood
31. bone
32. grease
33. egg
34. horn
35. tail
36. feather
37. hair
38. head
39. ear
40. eye
41. nose
42. mouth
43. tooth
44. tongue
45. claw
46. foot
47. knee
48. hand
49. belly
50. neck
51. breasts
52. heart
53. liver
54. drink
55. eat
56. bite
57. see
58. hear
59. know
60. sleep
61. die
62. kill
63. swim
64. fly
65. walk
66. come
67. lie
68. sit
69. stand
70. give
71. say
72. sun
73. moon
74. star
75. water
76. rain
77. stone
78. sand
79. earth
80. cloud
81. smoke
82. fire
83. ash
84. burn
85. path
86. mountain
87. red
88. green
89. yellow
90. white
91. black
92. night
93. hot
94. cold
95. full
96. new
97. good
98. round
99. dry
100. name

A.1.2 200-item Swadesh list (adapted from Swadesh 1952:456-7)

all
and
animal
ashes
at
back (body part)
bad
bark (of a tree)
because
belly
big
bird
to bite
black
blood
to blow
bone
to breathe
burn
child
cloud
cold
to come
to count
to cut
day
to die
to dig
dirty
dog
to drink
dry
dull (of a knife)
dust
ear
earth
to eat
egg
eye
to fall
far
fat (grease)
father
to fear
feather
few
to fight
fire
fish
five
to float
to flow
flower
to fly
fog
foot
four
to freeze
fruit
to give
good
grass
green
guts
hair
hand
he
head
to hear
heart
heavy
here
to hit
to hold
how
to hunt
husband
I
ice
if
in
to kill
to know
lake
to laugh
leaf
left (side)
leg
to lie (down)
to live
liver
long
louse
man
many
meat
mother
mountain
mouth
name
narrow
near
neck
new
night
nose
not
old
one
other
person
to play
to pull
to push
to rain
red
right
right (correct)
river
road
root
rope
rotten
to rub
salt
sand
to say
to scratch
sea
to see
seed
to sew
sharp
short
to sing
to sit
skin
sky
to sleep
small
to smell
to smoke (fire)
smooth
snake
snow
some
to spit
to split
to squeeze
to stab
to stand
star
stick
stone
straight
to suck
sun
to swell
to swim
tail
that
there
they
thick
thin
to think
this
thou
three
to throw
to tie
tongue
tooth
tree
to turn
two
to vomit
to walk
warm
to wash
water
we
wet
what
when
where
white
who
wide
wife
wind
wing
to wipe
with
woman
woods
worm
ye
year
yellow

A.2 The Leipzig-Jakarta list

The Leipzig-Jakarta List of Basic Vocabulary (Tadmor, Haspelmath and Taylor 2010:239-41).

Rank	Meaning	Borrowed score	Age score	Analyzability score	Representation score	Composite score
1	fire	0.965	0.939	0.995	1.000	0.901
2	nose	0.973	0.906	0.980	1.000	0.864
3	to go	0.963	0.887	0.974	1.000	0.832
4	water	0.909	0.926	0.987	1.000	0.831
5	mouth	0.920	0.904	0.982	1.000	0.817
6	tongue	0.934	0.908	0.954	1.000	0.808
7	blood	0.904	0.890	1.000	1.000	0.805
7	bone	0.918	0.904	0.971	1.000	0.805
9	2sg pro	0.958	0.893	0.933	1.000	0.798
9	root	0.944	0.869	0.973	1.000	0.798
11	to come	0.968	0.876	0.940	1.000	0.796
12	breast	0.947	0.856	0.967	1.000	0.783
13	rain	0.916	0.898	0.950	1.000	0.782
14	1sg pro	0.970	0.875	0.936	0.976	0.776
15	name	0.915	0.886	0.955	1.000	0.774
15	louse	0.950	0.861	0.946	1.000	0.774
17	wing	0.884	0.904	0.968	1.000	0.773
18	flesh/meat	0.877	0.892	0.986	1.000	0.771
19	arm/hand	0.881	0.903	0.966	1.000	0.768
20	fly	0.948	0.858	0.942	1.000	0.766
20	night	0.931	0.880	0.934	1.000	0.766
22	ear	0.896	0.888	0.961	1.000	0.764
23	neck	0.895	0.881	0.964	1.000	0.760
23	far	0.944	0.850	0.948	1.000	0.760
25	to do/make	0.947	0.877	0.914	1.000	0.759
26	house	0.893	0.876	0.969	1.000	0.758
27	stone/rock	0.895	0.882	0.958	1.000	0.756
28	bitter	0.975	0.872	0.889	1.000	0.755
28	to say	0.972	0.837	0.928	1.000	0.755
28	tooth	0.882	0.877	0.975	1.000	0.755
31	hair	0.944	0.871	0.917	1.000	0.754
32	big	0.889	0.864	0.980	1.000	0.753
32	one	0.870	0.893	0.969	1.000	0.753
34	who?	0.968	0.838	0.924	1.000	0.749

Continued on Next Page...

The Leipzig-Jakarta List – Continued

Rank	Meaning	Borrowed score	Age score	Analyzability score	Representation score	Composite score
34	3sg pro	1.000	0.893	0.955	0.878	0.749
36	to hit/beat	0.955	0.827	0.947	1.000	0.748
37	leg/foot	0.856	0.897	0.972	1.000	0.747
38	horn	0.840	0.898	0.987	1.000	0.745
38	this	1.000	0.851	0.897	0.976	0.745
38	fish	0.855	0.885	0.984	1.000	0.745
41	yesterday	0.958	0.843	0.922	1.000	0.744
42	to drink	0.904	0.877	0.934	1.000	0.741
42	black	0.951	0.866	0.899	1.000	0.741
42	navel	0.878	0.860	0.982	1.000	0.741
45	to stand	0.981	0.847	0.889	1.000	0.738
46	to bite	0.964	0.861	0.887	1.000	0.736
46	back	0.918	0.868	0.924	1.000	0.736
48	wind	0.828	0.900	0.987	1.000	0.736
49	smoke	0.916	0.863	0.929	1.000	0.734
50	what?	0.971	0.804	0.939	1.000	0.732
51	child (kin term)	0.929	0.866	0.930	0.976	0.730
52	egg	0.910	0.846	0.945	1.000	0.728
53	to give	0.913	0.878	0.907	1.000	0.727
53	new	0.920	0.860	0.920	1.000	0.727
53	to burn (intr.)	0.951	0.860	0.889	1.000	0.727
56	not	0.965	0.880	0.974	0.878	0.726
56	good	0.893	0.860	0.945	1.000	0.726
58	to know	0.933	0.856	0.908	1.000	0.725
59	knee	0.911	0.862	0.922	1.000	0.724
59	sand	0.901	0.866	0.928	1.000	0.724
61	to laugh	0.942	0.844	0.910	1.000	0.723
61	to hear	0.953	0.848	0.895	1.000	0.723
63	soil	0.900	0.883	0.954	0.951	0.722
64	leaf	0.897	0.823	0.977	1.000	0.721
64	red	0.926	0.864	0.900	1.000	0.721
66	liver	0.869	0.857	0.967	1.000	0.720
67	to hide	0.928	0.847	0.913	1.000	0.718
67	skin/hide	0.889	0.875	0.924	1.000	0.718
67	to suck	0.940	0.860	0.888	1.000	0.718
70	to carry	0.919	0.838	0.953	0.976	0.717

Continued on Next Page...

The Leipzig-Jakarta List – Continued

Rank	Meaning	Borrowed score	Age score	Analyzability score	Representation score	Composite score
71	ant	0.865	0.850	0.975	1.000	0.716
71	heavy	0.911	0.874	0.901	1.000	0.716
71	to take	0.900	0.898	0.887	1.000	0.716
74	old	0.896	0.867	0.920	1.000	0.715
75	to eat	0.920	0.840	0.925	1.000	0.714
76	thigh	0.906	0.856	0.918	1.000	0.712
76	thick	0.950	0.827	0.906	1.000	0.712
78	long	0.956	0.824	0.898	1.000	0.707
79	to blow	0.962	0.857	0.878	0.976	0.706
80	wood	0.860	0.871	0.940	1.000	0.705
81	to run	0.976	0.833	0.867	1.000	0.704
81	to fall	0.946	0.825	0.903	1.000	0.704
83	eye	0.904	0.847	0.918	1.000	0.703
84	ash	0.853	0.891	0.921	1.000	0.699
84	tail	0.883	0.813	0.973	1.000	0.699
84	dog	0.838	0.869	0.960	1.000	0.699
87	to cry/ weep	0.871	0.871	0.921	1.000	0.698
88	to tie	0.879	0.836	0.948	1.000	0.697
89	to see	0.918	0.842	0.900	1.000	0.695
89	sweet	0.914	0.857	0.887	1.000	0.695
91	rope	0.848	0.824	0.993	1.000	0.694
91	shade/ shadow	0.887	0.840	0.931	1.000	0.694
91	bird	0.842	0.857	0.962	1.000	0.694
91	salt	0.848	0.838	0.976	1.000	0.694
91	small	0.909	0.790	0.966	1.000	0.694
96	wide	0.955	0.819	0.885	1.000	0.692
97	star	0.830	0.859	0.970	1.000	0.691
97	in	0.948	0.856	0.943	0.902	0.691
99	hard	0.918	0.833	0.903	1.000	0.690
100	to crush/grind	0.919	0.845	0.886	1.000	0.688

A.3 Dolgopolsky's Lists

A.3.1 Dolgopolsky's initial list

(Adapted from Dolgopolsky 1986:33-4)

1.	no replacements :	'five', 'three', 'four', 'six', 1st.sg pronoun
2.	1-1½ replacements :	'two', 'seven', 'eight'
3.	2-2½ replacements:	2nd.sg pronoun
4.	3-3½ replacements:	'who'
5.	4-4½ replacements:	'ten', 1st.pl pronoun, 'one', 'tongue', 2nd.pl pronoun, 'nine'
6.	5-5½ replacements:	'name'
7.	6-6½ replacements:	'100', 'what'
8.	7-7½ replacements:	'eye', 'twenty', 'heart'
9.	8-8½ replacements:	'tooth', prohibitive (imperative) NEG ('no', 'non-', 'not'), verbal NEG, 'nit'
10.	9-9½ replacements:	'finger/toenail', 'louse', 'new moon, crescent of a moon', 'tear' (n.)
11.	10-10½ replacements:	'water', 'dead', 'hand'
12.	11-11½ replacements:	'night', 'blood'
13.	12-12½ replacements:	'horn', 'full', 'sun', 'ear', 'salt'

A.3.2 Dolgopolsky's final list

(1986:34-5)

1. first person marker
2. 'two'
3. second person marker
4. 'who', 'what'
5. 'tongue'
6. 'name'
7. 'eye'
8. 'heart'
9. 'tooth'
10. verbal NEG (both negative proper and prohibitive)
11. 'finger/toenail'
12. 'louse'
13. 'tear' (n.)
14. 'water'
15. 'dead'

A.4 Lohr's lists

A.4.1 List 1

Intersection of items with a retentiveness > 10 000 years in IE and items from all initial lists (1999:65)

one	full	feather	shadow	to be (3rd.sg. pres)
two	green	fly	silver	to come
three	high	foot	sister	to drink
four	light (in weight)	gold	sleep	to eat
five	long	goose	smoke	to fly
six	middle	hand	snow	to give
seven	new	heart	son	to lick
eight	thin	horn	star	to listen
nine	young	knee	sun	to live
ten	bear	lamb	sweat	to measure
hundred	brother	light	tear	to milk
first	cow	month	thunder	to sew
all	daughter	mother	tongue	to sit
I	day	mouse	tooth	to spit
not	ear	nail	water	to stand
other	egg	name	wind	to stretch
over	elbow	navel	wolf	to weave
we	evening	night	worm	to yawn
what?	eye	nose	yesterday	
you (pl.)	father	root	yoke	
you (sg.)		salt		

A.4.2 List 2

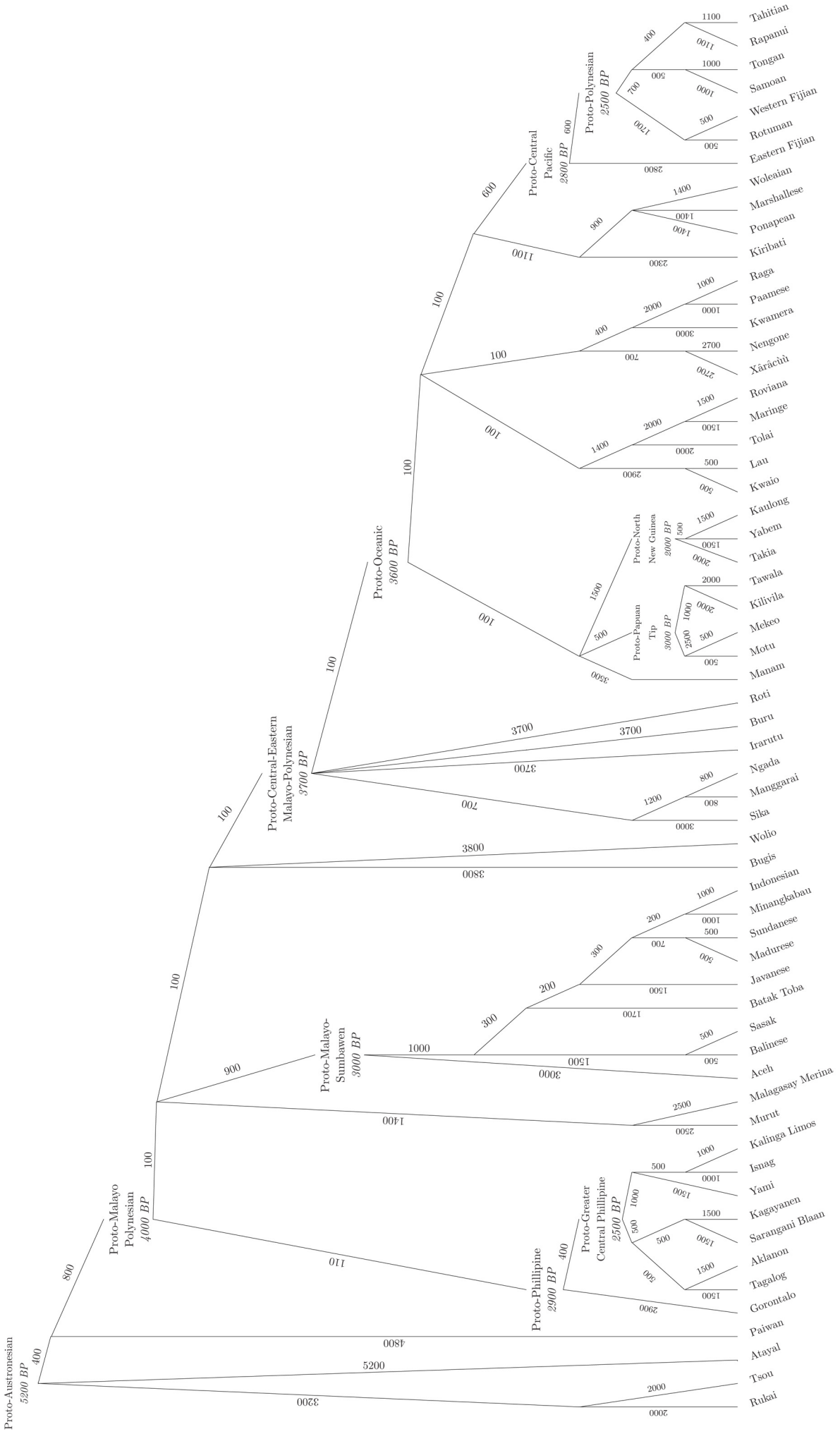
Intersection of items with retentiveness > 5 000 years in IE and forms which were reconstructed for three or more proto-languages only (1999:66)

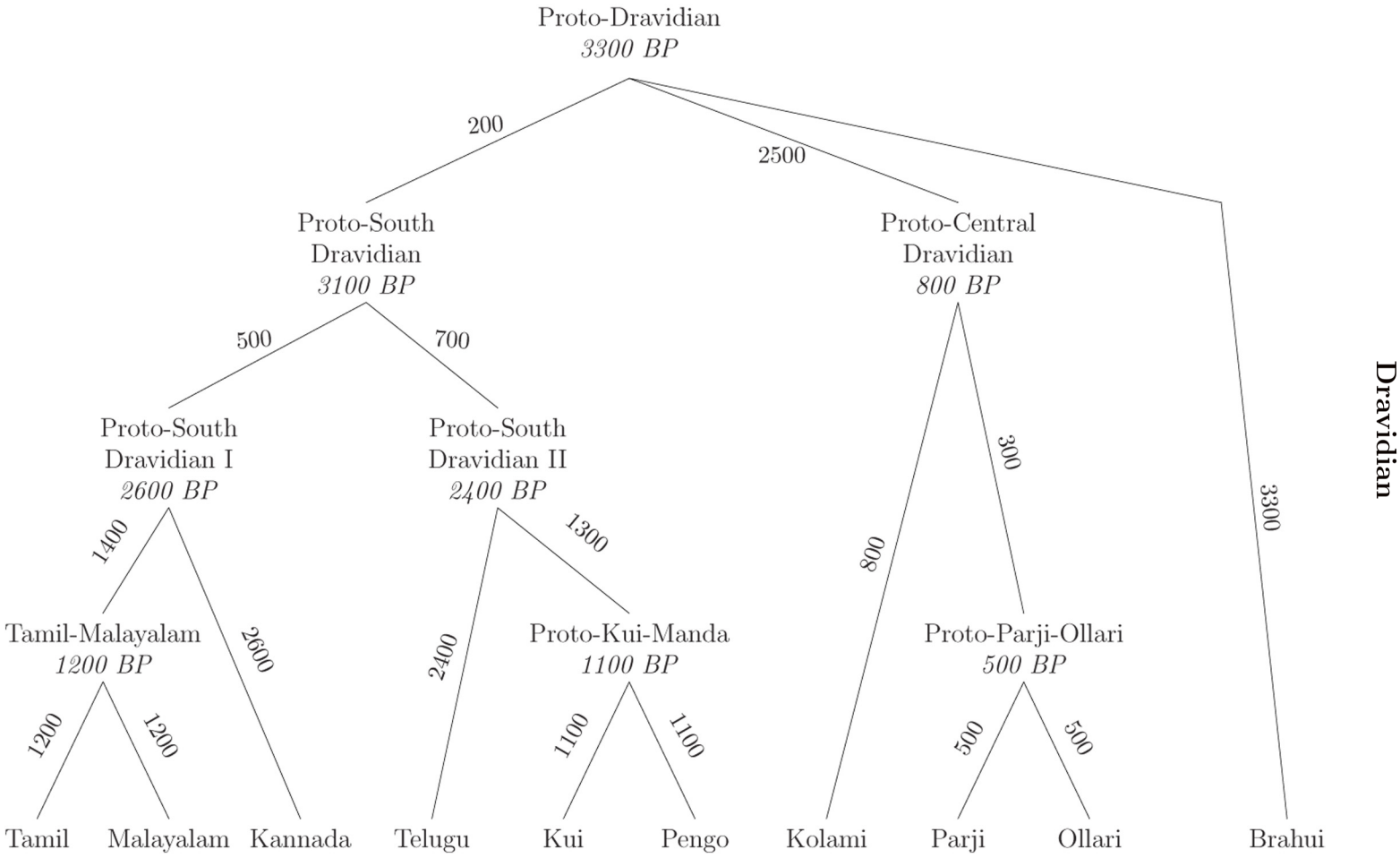
one	hard	fly	sky	to go
two	high	foot	sleep	to grind
three	long	ground	snake	to grow
four	narrow	hair	star	to hew
five	new	head	sun	to laugh
six	right (side)	heart	thread	to leave
seven	sharp	honey	tongue	to lick
eight	thin	light	tooth	to lie
nine	true	liver	tree	to listen
ten	wide	meat	voice	to live
first	yellow	moon	water	to pour
and	young	mother	wind	to put
half	bood	mouth	worm	to rub
I	bone	nail	to breathe	to see
many	brother	name	to burn	to sell
not	cloud	night	to carry	to sew
other	darkness	nose	to change	to shine
over	day	path	to come	to show
that	dog	person	to cover	to sing
we	ear	pig	to defecate	to spit
you (pl.)	egg	rain	to die	to stand
you (sg.)	evening	root	to eat	to suck
bitter	eye	salt	to fall	to tie
dry	father	sand	to float	to weave
full	fish	seed	to fly	
green	flower	shadow	to give	

B Trees

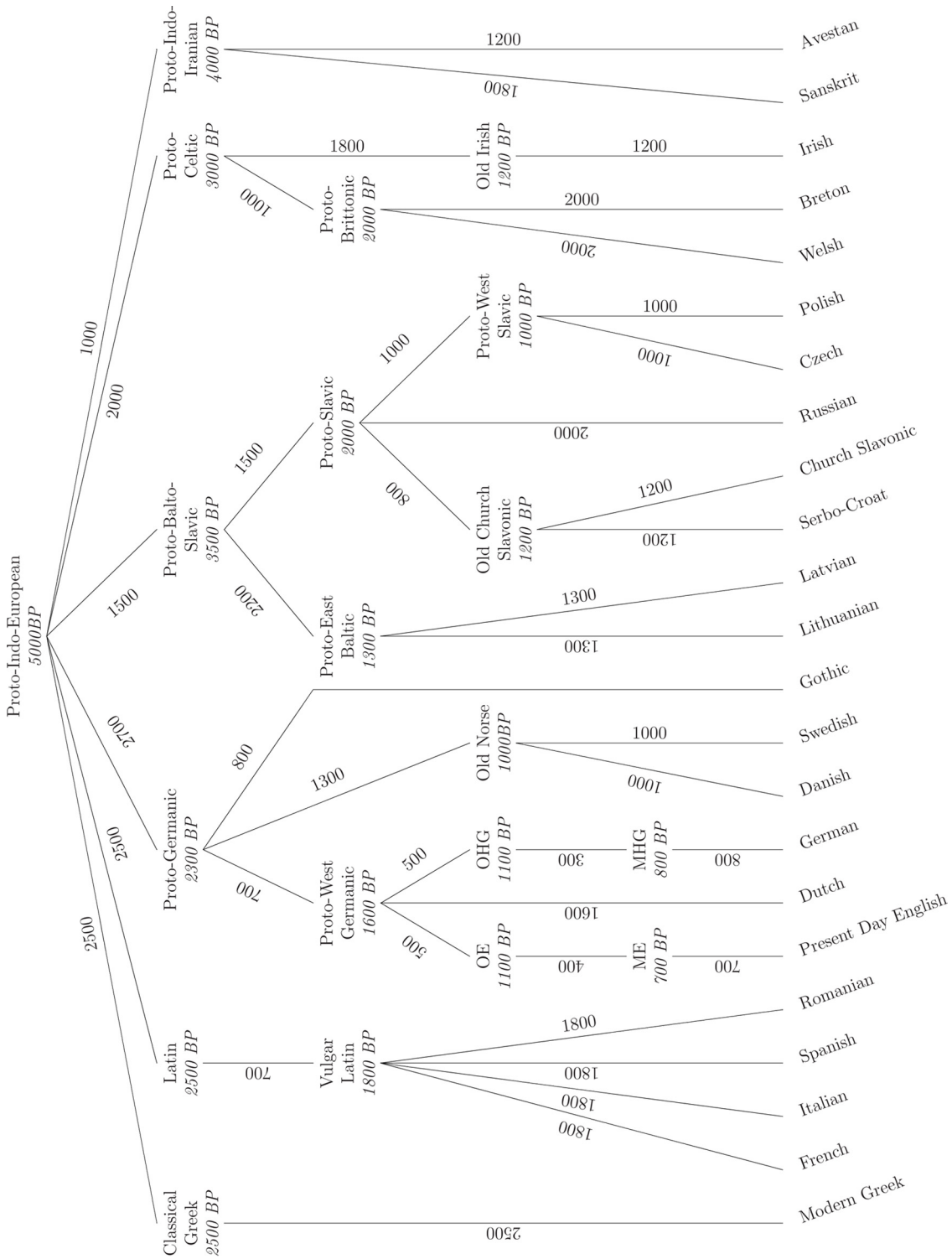
Each node is dated in years before present (BP), and the time depth between nodes is given alongside each branch in years. The sources used to obtain these conservative best-judgements for split dates are cited in section 3.3.

Austronesian

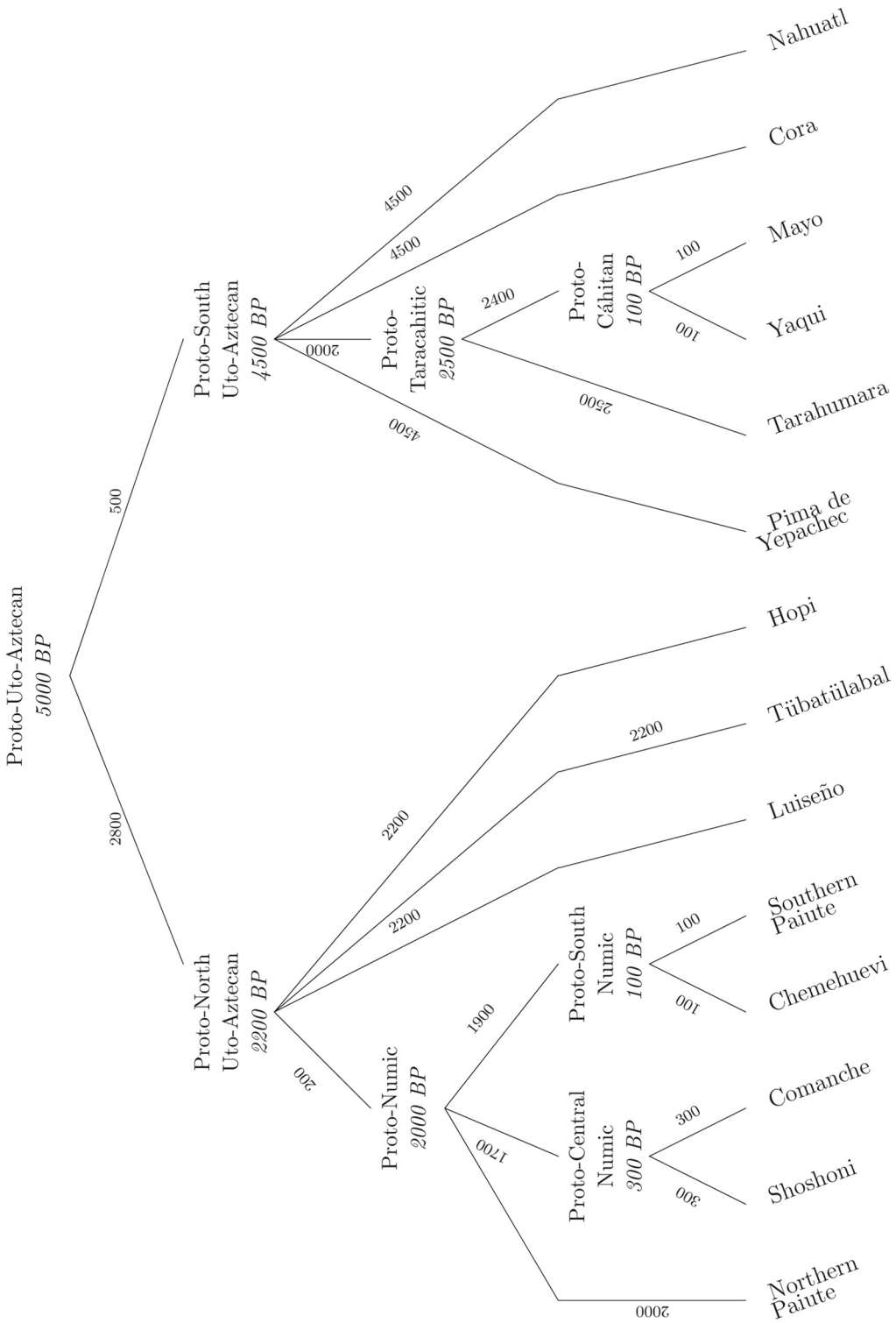



Proto-Dravidian
3300 BP
200
2500
Proto-South Dravidian
3100 BP
Proto-Central Dravidian
800 BP
500
700
Proto-South Dravidian I
2600 BP
Proto-South Dravidian II
2400 BP
1400
1300
800
300
3300
Tamil-Malayalam
1200 BP
Proto-Kui-Manda
1100 BP
Proto-Parji-Ollari
500 BP
2600
2400
1200
1200
1100
1100
500
500
Tamil
Malayalam
Kannada
Telugu
Kui
Pengo
Kolami
Parji
Ollari
Brahui
Dravidian

Indo-European



Uto-Aztecan


Proto-Uto-Aztecan
5000 BP
500
2800
Proto-South
Uto-Aztecan
4500 BP
Proto-North
Uto-Aztecan
2200 BP
4500
4500
2000
4500
Proto-
Taracahitic
2500 BP
2400
2500
Proto-
Cáhitan
100 BP
100
100
Nahuatl
Cora
Mayo
Yaqui
Tarahumara
Pima de
Yepachec
2200
2200
2200
200
Proto-Numic
2000 BP
1900
1700
Proto-South
Numic
100 BP
100
100
Proto-Central
Numic
300 BP
300
300
2000
Hopi
Tübatülabal
Luiseño
Southern
Paiute
Chemehuevi
Comanche
Shoshoni
Northern
Paiute

References

Andersen, H. (2009). The satem languages of the Indo-European Northwest. First contacts? In *The Indo-European Family: Questions about its Status*, number 55 in Journal of Indo-European Studies Monograph Series. Washington, DC.: Institute for the Study of Man. [62]

Antilla, R. (1989). *Historical and Comparative Linguistics*. Amsterdam: John Benjamins, 2nd revised edition. [4, 25]

Asher, R. E. (1982). *Tamil*. Amsterdam: Lingua Descriptive Studies. [50]

Asher, R. E. and Kumari, T. C. (1997). *Malayalam*. London: Routledge. [50]

Beekes, R. S. P. (2011 [1995]). *Comparative Indo-European Linguistics*. Amsterdam: John Benjamins, 2nd. revised edition. [6]

Bergsland, K. and Vogt, H. (1962). On the validity of glottochronology. *Current Anthropology*, 3:111–53. [20, 21, 23]

Bhattacharya, S. (1957). *Ollari: a Dravidian speech*. Delhi: Manager of Publications. [50]

Blust, R. (1999). Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In *Selected Papers From the 8th International Conference on Austronesian Linguistics*. Taipei, Taiwan: Academica Sinica. [56, 58]

Bowern, C., Epps, P., Gray, R., Hill, J., Hunley, K., McConvell, P., and Zentz, J. (2011). Does lateral transmission obscure inheritance in hunter-gatherer languages? *PLoS ONE*, 6(9):e25195. doi:10.1371/journal.pone.0025195. [15, 16]

Bray, D. (1934). *The Brāhūī language. Part II. The Brāhūī problem; Part III. Etymological vocabulary*, volume 2. Delhi: Manager of Publications. [50]

Breen, G. (1990). *Salvage studies of western Queensland Aboriginal languages*. Canberra: Pacific Linguistics. [15]

Brown, C. H., Holman, E. W., ren Wichmann, S., and Velupillai, V. (2008). Automated classification of the world's languages: a description of the method and preliminary results. *STUF: Language Typology and Universals*, 61(4):285–308. [9]

Buck, C. D. (1949). *A Dictionary of Selected Synonyms in the Principal Indo-European Languages: A contribution to the history of ideas*. Chicago: University of Chicago Press. [29, 37, 44, 49, 50, 56, 57, 59]

Burrow, T. and Bhattacharya, S. (1953). *The Parji Language: A Dravidian Language of Bastar*. Hertford: Stephen Austin and sons, ltd. [50]

Burrow, T. and Bhattacharya, S. (1970). *The Pengo Language: Grammar, Texts, and Vocabulary*. Oxford: Clarendon Press. [50]

Burrow, T. and Emeneau, M. B. (1984). *A Dravidian etymological dictionary*. Oxford: Clarendon Press, 2nd edition. [49]

Campbell, L. (1988). Review of *Language in the Americas*, Greenberg 1987. *Language*, 64:591–615. [1]

Campbell, L. (1997). *American Indian Languages*. Oxford: Oxford University Press. [52]

Campbell, L. (1998). *Historical Linguistics*. Edinburgh: Edinburgh University Press. [1, 9, 10, 20, 25, 76, 83]

Campbell, L. and Poser, W. (2008). *Language Classification*. Cambridge: Cambridge University Press. [1, 8, 9, 11, 15, 16, 24, 49]

Canonge, E. (1958). *Comanche Texts*. Norman: Summer Institute of Linguistics of the University of Oklahoma. [50]

Chen, B. (1996). *Lun Yuyan Jiechu Yu Yuyan Lianmeng*. Beijing: Yuwen Chubanshe. [16]

Collard, H. and Collard, E. S. (1962). *Castellano-Mayo, Mayo-Castellano*. México: Instituto Lingüístico de Verano en cooperación con la Dirección General de Asuntos Indígenas de la Secretaría de Educación Pública. [50]

Comrie, B. (1993). Review of *Altajskaja problema i proisxozdenie japonskogo jazyka* [The Altaic problem and the origin of the Japanese language], Starostin 1991. *Language*, 69(4):828–832. [49]

Cowan, H. K. J. (1959). A note on statistical methods in comparative linguistics. *Lingua*, 8:223–46. [17]

del náhuatl, N. D. (2012). http://whp.uoregon.edu/dictionaries/nahuatl/index.lasso. Accessed 2/3/12. [-]

Dictionary, S. O. (2012). shoshonidictionary.com. Accessed 3/4/12. [50]

Dixon, R. M. W. (1997). *The Rise and Fall of Languages*. Cambridge: Cambridge University Press. [15, 24]

Dolgopolsky, A. B. (1986). A probabilistic hypothesis concerning the oldest relationships among the language families in northern Eurasia. In Shevoroshkin, V. V. and Markey, T. L., editors, *Typology, Relationship and Time: A collection of papers on language change and relationship by Soviet linguists*. Ann Arbor: Karoma. [i, 2, 16, 31, 33, 34, 35, 41, 43, 44, 47, 57, 63, 74, 85, 100]

Durie, M. and Ross, M., editors (1996). *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. Oxford: Oxford University Press. [6]

Dyen, I., Kruskal, J. B., and Black, P. (1992). An Indoeuropean classification: a lexicostatistical experiment. *Transactions of the American Philosophical Society*, 82:1–132. Data available at http://ldc.upenn.edu, accessed 3/4/12. [49, 50]

Ehret, C. (1995). *Reconstructing Proto-Afroasiatic (Proto-Afrasian). Vowels, Tones, Consonants and Vocabulary*. Berkeley; Los Angeles; London: University of California Press. [37]

Embleton, S. (1986). *Statistics in Historical Linguistics*. Bochum: Brockmeyer. [16, 22, 24, 26, 49, 65, 90]

Embleton, S. (2000). Lexicostatistics/glottochronology: from Swadesh to Sankoff to Starostin to future horizons. In Renfrew, C., McMahon, A., and Trask, L., editors, *Time Depth in Historical Linguistics*. Cambridge: The McDonald Institute for Archaeological Research. [20]

Emeneau, M. B. (1961). *Kolami: A Dravidian Language*. Annamalainagar: Annamalai University, 2nd edition. [50]

Fabricus, J. P. (1972). *Tamil and English Dictionary*. Tranquebar: Evangelical Lutheran Mission Publishing House, 4th revised edition. (online version: http://dsal.uchicago.edu/dictionaries/fabricius/, accessed 10/2/12). [50]

Fernández, Z. E. (2009). Yaqui vocabulary. In Haspelmath, M. and Tadmor, U., editors, *World Loanword Database*. Munich: Max Planck Digital Library, http://wold.livingsources.org/vocabulary/32. Accessed 3/2/12. [50]

Fortson, B. W. (2004). *Indo-European Language and Culture*. Malden, Mass.: Blackwell. [6, 62]

Fox, A. (1995). *Linguistic Reconstruction: An Introduction to Theory and Method*. Oxford: Oxford University Press. [1, 4, 9, 10, 20, 24, 25]

Gamkrelidze, T. V. and Ivanov, V. V. (1995). *Indo-European and the Indo-Europeans*. Berlin; New York: Mouton de Gruyter. [6]

Georg, S. (2004). Review of *Etymological Dictionary of the Altaic Languages*. *Diachronica*, 21(2):445–450. [49]

Greenberg, J. (1963). *The Languages of Africa*. Bloomington: Indiana University Press. [1]

Greenberg, J. (1987). *Language in the Americas*. Stanford: Stanford University Press. [37]

Greenberg, J. (2000). *Indo-European and its Closest Relatives*. Stanford: Stanford University Press. [-]

Greenhill, S. J., Blust, R., and Gray, R. D. (2008). The Austronesian basic vocabulary database: From bioinformatics to lexomics. *Evolutionary Bioinformatics Online*, 4(27):271–283. [15, 47, 50, 52, 55, 56, 58, 88]

Gudschinksy, S. (1956). The ABCs of lexicostatistics (glottochronology). *Word*, 12:175–210. [79]

Gwynn, J. P. L. (1991). *A Telugu-English Dictionary*. Delhi; New York: Oxford University Press. [50]

Haspelmath, M. and Tadmor, U. (2009). World Loanword Database. Munich: Max Planck Digital Library. http://wold.livingsources.org/, accessed 29/5/12. [15, 28, 34, 45, 47, 76, 78, 85, 86, 87, 88, 89, 93]

Hedges, B. A. (1981). On positive correlation between means and standard deviations of claims ratios: commentary. *The Journal of Risk and Insurance*, 48(4):649–652. [83]

Heggarty, P. (2010). Beyond lexicostatistics: How to get more out of 'word list' comparisons. *Diachronica*, 27(2):301–24. [9, 20, 22, 24]

Herrara, F. (2004). *Nahuatl-English/English-Nahuatl (Aztec)*. New York: Hippocrene Books. [51]

Hock, H. H. and Joseph, B. D. (2009). *Language History, Language Change, and Language Relationship*. Berlin; New York: Mouton de Gruyter, 2nd revised edition edition. [25]

Hoijer, H. (1956). Lexicostatistics: A critique. *Language*, 32:49–60. [17, 26]

Holman, E. W., Wichmann, S., Brown, C. H., Velupillai, V., Müller, A., and Bakker, D. (2008). Explorations in automated language classification. *Folia Linguistica*, 42:331–54. [25, 28, 40]

Hopi Dictionary Project (1998). *Hopi Dictionary/Hopìikwa Lavàytutuveni: A Hopi Dictionary of the Third Mesa Dialect with an English-Hopi Finder List and a Sketch of Hopi Grammar*. Tuscon, AZ: University of Arizona Press. [50]

Hopper, P. J. (1973). Glottalized and murmured occlusives in Indo-European. *Glossa*, 7(2):141–166. [6]

Jakobson, R. (1960). Why 'mama' and 'papa'? In Kaplan, B. and Wapner, S., editors, *Perspectives in Psychological Theory*. New York: International Universities Press. [8]

Krishnamurti, B. (2003). *The Dravidian Languages*. Cambridge: Cambridge University Press. [52, 60, 62]

Kroeber, A. L. and Grace, G. W., editors (1960). *The Sparkman Grammar of Luiseño*, volume 16 of *University of California publications in linguistics*. Berkeley: University of California Press. [50]

Kruskal, J. B., Dyen, I., and Black, P. (1973). Some results from the vocabulary method of reconstructing language trees. In Dyen, I., editor, *Lexicostatistics in Genetic Linguistics*. The Hague; Paris: Mouton. [21, 22, 49, 65, 90]

Laet, J. d. (1643). *Notae ad Dissertationem Hugonis Grotii*. Amsterdam; Paris: apud viduam Gvilielmi Pele. [24]

Lees, R. (1953). The basics of glottochronology. *Language*, 29:113–27. [21, 23, 79]

Letchmajee, L. (1902). *An introduction to the grammar of the Kui or Kandh language*. Calcutta: Bengal Secretariat Press. [50]

Lohr, M. (1999). *Methods for the Genetic Classification of Languages*. PhD thesis, University of Cambridge. [i, 2, 5, 9, 16, 20, 22, 25, 26, 31, 36, 37, 38, 39, 41, 43, 44, 47, 49, 51, 52, 63, 74, 79, 84, 85, 86, 87]

Luce, G. H. (1981). *A Comparative Word List of Old Burmese, Chinese and Tibetan*. London: School of Oriental and African Studies. [37]

Mahadevan, S. (2011). *English-Tamil, Tamil-English Dictionary*. London: ibsBOOKS (UK). [50]

Mallory, J. P. and Adams, D. (2006). *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World*. Oxford: Oxford University Press. [13, 52, 57, 58, 59]

Matisoff, J. (1990). On megalocomparison. *Language*, 66:106–20. [1]

McLaughlin, J. (1992). A counter-intuitive solution in Central Numic phonology. *International Journal of American Linguistics*, 58(2):158–81. [52]

McMahon, A. and de McMahon, M. A. (1959). *Cora y español*. México: Instituto Lingüístico de Verano en cooperación con la Dirección General de Asuntos Indígenas de la Secretaría de Educación Pública. [50]

McMahon, A., Heggarty, P., McMahon, R., and Slaska, N. (2005). Swadesh sublists and the benefits of borrowing: an Andean case study. *Transactions of the Philological Society*, 103(2):147–70. [25]

McMahon, A. and McMahon, R. (1995). Linguistics, genetics and archaeology: Internal and external evidence in the Amerind controversy. *Transactions of the Philological Society*, 93(2):125–225. [1]

McMahon, A. and McMahon, R. (2005). *Language Classification by Numbers*. Oxford: Oxford University Press. [1, 4, 9, 16, 20, 62]

McMahon, A. and McMahon, R. (2006). Why linguists don't do dates: evidence from Indo-European and Australian languages. In Forster, P. and Renfrew, C., editors, *Phylogenetic Methods and the Prehistory of Languages*. Cambridge: McDonald Institute for Archaeological Research. [20]

Merlon, F. (1982). *Mangaray*. Amsterdam: North-Holland Publishing Company. [34]

Miller, W. R. (1983). Uto-Aztecan languages. In Sturtevant, W. C., editor, *Handbook of North American Indians*, volume 10: Southwest. Washington: Smithsonian Institution. [52]

Miller, W. R. (1986). Numic languages. In Sturtevant, W. C., editor, *Handbook of North American Indians*, volume 11: Great Basin. Washington: Smithsonian Institution. [52]

Mithun, M. (1999). *The Languages of Native North America*. Cambridge: Cambridge University Press. [52]

Nettle, D. (1999). Towards a future history of macrofamily research. In Renfrew, C. and Nettle, D., editors, *Nostratic: Examining a linguistic macrofamily*. Cambridge: McDonald Institute for Archaeological Research. [2]

Nichols, J. (1996). The comparative method as heuristic. In Durie, M. and Ross, M., editors, *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. Oxford: Oxford University Press. [5]

O'Neil, W. A. (1964). Problems in the lexicostatistic time depth of Modern Icelandic and Modern Faroese. *General Linguistics*, 6(1):27–37. [23]

Oswalt, R. L. (1970). The detection of remote linguistic relationship. *Computer Studies in the Humanities and Verbal Behaviour*, 3:117–29. [25, 36, 38]

Oswalt, R. L. (1991). A method for assessing distant linguistic relationship. In Lamb, S. M. and Mitchell, E. D., editors, *Sprung from some common source: Investigations into the prehistory of languages*. Stanford: Stanford University Press. [25, 36]

Pagel, M., Atkinson, Q. D., and Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, 449(7163):717–720. [14, 24, 53, 64]

Press, M. L. (1979). *Chemehuevi: a grammar and lexicon*. Berkeley; London: University of California Press. [50]

Rao, H. R. (1967). *Kannada Made Easy*. Bombay: D. B. Taraporevala sons & co. [50]

Reece, L. (1970). *Grammar of the Wailibri Language of Central Australia*. Sydney: University of Sydney Press. [34]

Ross, M. and Durie, M. (1996). Introduction. In Durie, M. and Ross, M., editors, *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. Oxford: Oxford University Press. [6]

Ruhlen, M. (1991). The amerind phylum and the prehistory of the new world. In Lamb, S. M. and Mitchell, E. D., editors, *Sprung from some common source: Investigations into the prehistory of languages*. Stanford: Stanford University Press. [1]

Ruhlen, M. (1994). *On the Origin of Languages: Studies in Linguistic Taxonomy*. Stanford: Stanford University Press. [-]

Saltarelli, M. (1988). *Basque*. London: Croom Helm. [7]

Sapir, E. (1931). Southern Paiute dictionary. *Proceedings of the American Academy of Arts and Sciences*, 65(3):537, 539–730. [50]

Shaul, D. L. (1994). A sketch of the structure of Oob No'ok (Mountain Pima). *Anthropological Linguistics*, 36(3):277–365. [50]

Shipley, W. F. (1963). *Maidu Texts and Dictionary*. Berkeley; Los Angeles: University of California Press. [34]

Slaska, N. (2005). Lexicostatistics away from the armchair: Handling people, props and problems. *Transactions of the Philological Society*, 103(2):221–42. [53]

Smith, E. W. (1964 [1907]). *A Handbook of the Ila Language: commonly called the Seshukulumbwe, spoken in north-western Rhodesia, south-central Africa, comprising grammar, exercises, specimens of Ila tales, and vocabularies*. Ridgewood, New Jersey; Gregg Press Incorporated., revised edition. [34]

Sridhar, S. N. (1990). *Kannada*. London: Routledge. [50]

Starostin, G. (2009). Review of "Language Classification: History and Method" by Lyle Campbell and William J. Poser. *Journal of Language Relationship*, 2:158–174. [16]

Starostin, S. (1991). *Altajskaja problema proischoshdenie japonskovo jazyka [The Altaic problem and the genesis of the Japanese language]*. Moscow: Nauka. [28, 40]

Starostin, S. (2000). Comparative-historical linguistics and lexicostatistics. In Renfrew, C., McMahon, A., and Trask, L., editors, *Time Depth in Historical Linguistics*. Cambridge: The McDonald Institute for Archaeological Research. [9, 21, 25, 40, 79]

Starostin, S., Dybo, A., Mudrak, O., Gruntov, I., and Glumov, V. (2003). *Etymological dictionary of the Altaic languages*. Leiden; Boston: Brill. 3 vols. [49]

Stubbs, B. (2011). *A Comparative Uto-Aztecan Vocabulary*. Shumway Family History Services: Flower Mound, TX; Rocky Mountain Books and Productions: Blanding, UT. [49, 52]

Swadesh, M. (1950). Salish internal relationships. *International Journal of American Linguistics*, 16(4):157–167. [9, 16, 19, 24, 79]

Swadesh, M. (1952). Lexico-statistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society*, 96(4):453–63. [19, 28, 76, 96]

Swadesh, M. (1955). Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics*, 21(2):121–137. [19, 21]

Swadesh, M. (1972). *The Origin and Diversification of Language*. London: Routledge and Kegan Paul. [1, 9, 11, 17, 24]

Tadmor, U. (2009). Indonesian vocabulary. In Haspelmath, Martin & Tadmor, U., editor, *World Loanword Database*. Munich: Max Planck Digital Library, http://wold.livingsources.org/vocabulary/27. Accessed 8/6/12. [58]

Tadmor, U., Haspelmath, M., and Taylor, B. (2010). Borrowability and the notion of basic vocabulary. *Diachronica*, 27(2):226–46. [i, 2, 11, 15, 16, 17, 18, 24, 28, 29, 30, 31, 40, 41, 42, 43, 44, 66, 69, 70, 71, 73, 75, 76, 77, 78, 79, 80, 82, 83, 84, 85, 86, 87, 90, 91, 92]

Tamil-English-Tamil Dictionary (2012). tamildictionary.org. Accessed 10/2/12. [50]

Taylor, C. (1985). *Nkore-Kiga*. London: Croom Helm. [34]

Telugu-English-Telugu Dictionary (2012). telugudictionary.org. Accessed 16/2/12. [50]

Thord-Gray, I. (1955). *Tarahumara-English, English-Tarahumara Dictionary and an Introduction to Tarahumara Grammar*. Coral Gables, FL. : University of Miami Press. [50]

Tovar, A., Bouda, K., Lafon, R., Michelena, L., Swadesh, M., and Vycichl, W. (1961). El método léxico-estadístico y su aplicación a las relaciones del vascuence. *Boletin de la Real Sociedad Vascongada de los Amigos del País*, 17:249–281. [1]

Trask, R. (2000). *The Dictionary of Historical and Comparative Linguistics*. Edinburgh: Edinburgh University Press. [9, 11, 83]

Trask, R. L. (2007 [1996]). *Historical Linguistics*. London: Arnold. [1, 7]

Tryon, D. T., editor (1995). *Comparative Austronesian Dictionary*, volume 1-5. Berlin; New York: Walter de Gruyter & Co. [50, 55]

Uldall, H. J. and Shipley, W. (1966). *Nisean Texts and Dictionary*. Berkeley; Los Angeles: University of California Press. [34]

Voegelin, C. F. (1958). A working dictionary of Tübatülabal. *International Journal of American Linguistics*, 24(3):221–8. [50]

Vovin, A. (2005). The end of the Altaic controversy. Review of Starostin et al. 2003. *Central Asiatic Journal*, 49(1):71–132. [49]

Wang, F. and Wang, W. S.-Y. (2004). Basic words and language evolution. *Language and Linguistics*, 5(3):643–662. [16, 25, 74]

Yerington Paiute Tribe (1987). *Paiute-English English-Paiute dictionary*. Anchorage, AK: Bilingual Education Series. [50]

Zilva Wickremasinghe, M. d. and Menon, T. N. (2005). *Malayalam self-taught by the natural method with phonetic pronunciation (Thimm's system)*. New Delhi: Asian Educational Services, 2nd revised edition. [50]

Zorc, R. D. (1995). A glossary of Austronesian reconstructions. In Tryon, D. T., editor, *Comparative Austronesian Dictionary: An Introduction to Austronesian Studies*. Berlin; New York: Mouton de Gruyter. Part 1, fasc. 2. [37]