

ISSN 2752-7417

**the Undergraduate
Linguistics Association of Britain**

2012 proceedings

**EDITED BY
CAITLIN WILSON**

Undergraduate Linguistics Association of Britain

ULAB II Proceedings

University of Edinburgh and University of West England, Bristol

14-15 April 2012

Editor: Caitlin Wilson

Copyeditors: Beatrix Livesey-Stephens, Andrew Tobin, Hui Zhu

Cover design: Lydia Wiernik

Copyright © Undergraduate Linguistics Association of Britain 2022

ISSN 2752-7417

Contents

ULAB National Committee 2012	4
ULAB Local Committee 2012	4
Introduction to the ULAB II Proceedings <i>Lydia Wiernik and Caitlin Wilson, University of Edinburgh</i>	5
Old Frisian Gemination-Degemination through Strata <i>Jurij Božič, University of Ljubljana</i>	7
On Tocharian Exceptionality to the Centum-Satem Isogloss <i>Richard Littauer, University of Malta</i>	18
Reanalysis and Regularisation: The Development of Morphological Structure <i>Gwilym Lockwood, School of Oriental and African Studies, University of London</i>	26
‘English corrupted by Singaporeans’: a Study of Speaker Perceptions of Colloquial Singapore English in 2012 <i>Nicholas Marsh, University of Oxford</i>	33
The Grapheme <V> in Classical Latin: Did it Represent Two Phonemes or Three? <i>Devon Smith, University of Cambridge</i>	49
A Psycholinguistic Experimental Approach to Affect and Vertical Position <i>Thomas Williams, University of the West of England, Bristol</i>	59

National 'Core' Committee 2012-2013

David Arnold <i>University of Edinburgh</i>	Founder, Chair
Richard Littauer <i>University of Malta</i>	Founder
Gina Bruckner <i>University of Edinburgh</i>	Secretary
Claire Swaffield <i>University of West England, Bristol</i>	Treasurer

Local Committee 2012-2013 (UWE, Bristol)

Tom Williams	Chair
Jessica Roberts	Secretary
Claire Swaffield	Treasurer
Maryam Bham, Amy Haines, David Etheridge	Communications Officers
Joshua Creaven	Web Administrator
Christian Ohman	Timetabling Officer
Catherine Haynes	Venue Officer
Alex Hurst	Social Secretary
Tanya Woll	Transport Officer
Gina Ko	Accommodation Officer
Antonia Kelly, Tina Hung	Committee Members

Introduction to the Proceedings of ULAB II

Lydia Wiernik and Caitlin Wilson, University of Edinburgh

As ULAB's 2021-2022 Archivist, I endeavoured to unearth what we were missing: a record of our past. There was very little by way of a structured archive prior to 2019, though our annual conference began in 2011. With help from [Richard Littauer](#), one of ULAB's founders, I was able to recover a majority of this archive, a section of which contained the unfinished 2012 Proceedings.

Richard asked if we'd be willing to finish what he'd started and complete the Proceedings, despite the elapsed decade. We agreed. Everything we needed was there: a list of committee members, authors' information and their respective papers. Though "all" we needed to do was put it together, this was no small feat; I leave it to Caitlin to describe her and the copyeditors' process.

Completing the 2012 Proceedings was not only an homage to our history, but a promise for our future. We utilised the Proceedings as a handover; I transitioned from the role of Archivist to JoULAB's editor and Caitlin from Local Chair to Archivist and JoULAB Associate Copyeditor. As I compiled [2021's Proceedings](#), I was able to offer technical advice and help oversee the process. Caitlin and the new cohort of JoULAB could familiarise themselves with the routine of editing and our Formatting and Styling Guide. It brought us together, too; publishing the 2012 Proceedings was a phenomenal way to get to know one another while contributing something meaningful to the ULAB story.

The Proceedings are the product of a massive, brilliant synergy by Caitlin and the copyediting team, aided by Richard's original work. We had such fun throughout this process, and I hope you enjoy it as much as we did.

If you have your own materials from past ULAB conferences and would like to share them with us, don't hesitate to reach out to Caitlin at ulablinguistics@gmail.com.

- Lydia

The purpose of this project, as Lydia has said, was to put ULAB history down in writing and make it available to everyone. We've come a long way from the small organisation that was created in 2011. Since the 2012 conference, we've had 9 ULAB conferences hosted in universities across the UK.

While a lot of things have changed in the decade between my term as Local Chair and the 2012 conference, at its core ULAB is, and always will be a place for passionate undergraduate linguists to come together and celebrate our research. The founders of ULAB themselves, Richard Littauer and David Arnold, got to experience this first-hand when they attended ULAB 2022. It was a huge honour to see them in the audience and to talk to them about our shared love for linguistics and ULAB.

To say that starting my term as Archivist with this project was like being thrown into the deep end is no understatement. Compiling and copyediting papers that were written and due to be published over a decade ago, all while learning the ropes of archiving and copyediting has been an amazing but challenging experience. The copyediting team and I came across all sorts of challenges when preparing these proceedings, but we had a great time along the way. If you'd like to read more about what we thought of this journey, you can read our blog post on the [ULAB website](#).

I'd like to thank Beatrix Livesey-Stephens, Andrew Tobin, and Hui Zhu for copyediting these Proceedings with me, Lydia Wiernik for all the advice they've given me, Richard Littauer for making this project happen, and to the authors of the papers you are about to read for trusting me with their words. I hope you enjoy reading back on research written over a decade ago, and that it brings you back to what I'm sure was an amazing time at the 2012 conference.

Lastly, I'd like to thank all current and past ULAB committee members for making this organisation what it is today. Please enjoy this little slice of ULAB history, the 2012 Proceedings of ULAB.

- Caitlin

Papers Accepted to the ULAB II Proceedings

The following papers have been accepted to the Proceedings of ULAB II, 2012.
Final acceptance was given by the National Chair and Founders of the 2012-2013 committee.

Old Frisian Gemination-Degemination through Strata

Jurij Božič

University of Ljubljana, Slovenia

Abstract. The present paper studies the Old Frisian phenomenon of gemination and degemination under the framework of Stratal Optimality Theory. Old Frisian geminates such as [ken:e]-‘kin’ dat.sg. were diachronically triggered by [j]-suffixation which triggered gemination in all West Germanic dialects. The synchronic state of Old Frisian geminates is examined whereby it is deduced that they must be specified in the Input. A process of Old Frisian word-final degemination, however, is assumed and analysed as a newly established Postlexical phenomenon which only affects word-final geminates, generating [ken]-‘kin’ nom.sg. alongside [ken:e]. By applying the theory of stratal sound change (Bermúdez-Otero, 2007), it is posited, that the ranking responsible for Postlexical degemination eventually percolates to the stem level, where it degeminates stem-final geminates (e.g., [ken:-] > [ken-]), which accounts for the unified lack of geminates at the end of the Old Frisian and the beginning Middle Frisian period.

Keywords: Old Frisian; gemination; degemination; Optimality Theory

1 Introduction

Consonant gemination is a distinctive feature of old West Germanic languages, which partly distinguishes them from North Germanic and quite drastically from East Germanic. It occurred systematically whenever the glide [j] immediately followed a consonant. This process had a profound influence on the creation of specific features in verbal and nominal paradigms (Murray and Venneman, 1983). Geminated consonants are, therefore, rampant in these paradigms, which somewhat complicates a purely linear analysis, especially since their trigger can no longer be seen on the surface, cf. Old English *cynn* ‘kin’ and the Old Frisian cognate *ken*. To further complicate matters, Old Frisian seems to have lost some geminates, as shown above, while early Old English retains them.

While general West Germanic and Old English, specifically, have received thorough descriptive treatment as well as attention in generative phonology, Old Frisian has received much scarcer treatments. The present paper, therefore, concentrates on synchronic and diachronic aspects of Old Frisian gemination in *ja*-stems as well as the later degemination processes that affected this paradigm.

The analysis is performed under the wider generative framework of Optimality Theory (cf. Prince and Smolensky, 1993; McCarthy, 2002), and, more specifically, under Stratal Optimality Theory (cf. Kiparsky, 2000a; Bermúdez-Otero, 1999a), which is tailor made for dealing with situations where the surface state is not directly deducible from its context, i.e., it exhibits opacity, and it also serves as a means of explaining phonological change in a straightforward manner. The paper attempts to show why a stratal approach is well suited for accounting for Old Frisian geminates in a generative fashion in terms of synchrony and diachrony.

Section 2 presents a brief introduction to classical Stratal Optimality Theory. While section 3 deals with the early Germanic phenomenon of gemination; section 4 presents the Old Frisian state of affairs. Section 5 discusses the theoretical notions presupposed for Input specification, while 6 and 7 sections present the process of generating Old Frisian geminates (including their degemination) in a synchronic and diachronic context. Section 8 concludes the paper.

2 Strata in Optimality Theory

Optimality Theory (OT) is an Input-Output theory of constraint interaction (Prince and Smolensky, 1993; McCarthy, 2002). Stratal OT (Kiparsky 2000a, 2000b, 2003a, 2003b; Bermúdez-Otero 1999a), on the other hand, incorporates the notions developed by Lexical Phonology (Kiparsky 1982, 1985) into classic OT, combining stratified derivation with the computational aspects of constraint interaction. Unlike the classic (parallel) variety, it proposes three specific strata, or levels, on which structure is generated:

(1) INPUT → STEM LEVEL → WORD LEVEL → POSTLEXICAL LEVEL → OUTPUT

It is vital to note that the output of each level constitutes the input for the higher level and, according to Kiparsky (2003a, p.110), constraint rankings may be different on ' $n+1$ ' as compared to ' n '. Each level is subject to the specific demands of its phonology: stem phenomena are generated on the stem level, word processes on the word level (these make up the 'lexical phonology') and postlexical processes on the postlexical level. The candidates on each level, therefore, pass through GEN and EVAL.

Such stratification allows us to elegantly explain opaque states: Kiparsky (2000a) notes that processes on the higher $n+1$ strata cause those on the lower levels to appear opaque. Thus, a word-level process may well obscure the operation of a stem-level process, causing opacity. Given two strata, n and $n+1$, the constraint ranking determines that *process x* takes place in environment α on stratum n , but on stratum $n+1$ the different ranking generates a candidate affected by *process x*, but without α . In a rule-based approach, a new rule, *rule y* would be said to delete x 's environment α . This is one example of opacity – a phenomenon where the surface state is not deducible from its surface context.

The presence of a postlexical stratum in language is, in turn, 'adduced from boundary interaction', and processes that govern it have to be fully 'transparent' (Koontz-Garboden, 2001, p.14), as there is no further level to make them opaque, but *they* are the ultimate culprit for opacity, or in other words, they obscure the processes that operate on the lower strata.

A noteworthy remark is also the assumption that diachronic processes, for instance, phonological innovations triggered by phonetic factors start out on the postlexical level, but eventually 'percolate' downwards (Kiparsky, 1995, In press; Bermúdez-Otero, 1999a, 1999b; 2007; Bermúdez-Otero and Trousdale, in press), (possibly) causing opacity on the way, until they reach the Lexicon. This prediction will be of great importance for explaining the process of degemination in a unified fashion in Old Frisian.

3 West Germanic Geminatio

A well attested phenomenon in West Germanic languages is consonant gemination. It was triggered after light-stem nouns, whenever $^{**}[j]^1$ was followed any consonant except $^{**}[r]$, before which it was systematically retained, while it was often deleted everywhere else (though we may reconstruct it for a hypothetical Proto-West-Germanic language) (Hogg, 1992, p.71; Ringe, 2009, p.130).

This phenomenon is also accountable for the gemination in Germanic weak verbs, *ja*-stem and *jō*-stem declensions, resulting in specific situations across the West Germanic attestations. Of primary

¹ For the sake of clarity, reconstructed forms are marked with $\langle^{**}\rangle$ and ungrammatical forms with $\langle^{*}\rangle$.

concern for this paper is the *ja*-stem declension, particularly in Old English and Old Frisian, which host a reflex of the following Proto-Germanic situation, according to Hogg and Fulk (2011, p.20):

Table 1: *Proto-Germanic ja-stem.*

	SG.	PL.
<i>Nom.</i>	**kunjam	**kunjō
<i>Acc.</i>	**kunjam	**kunjō
<i>Gen.</i>	**kunjas	**kunjōō
<i>Dat.</i>	**kunjāai	** kunjomiz

While traditional reconstructionist literature is abundant on West Germanic gemination, some recent generative insight has also been offered on the subject, particularly in connection with syllabification, cf. Kiparsky (1998), Ham (1998), Bermúdez-Otero (1999a, 1999b), Crist (2001). Particularly insightful is Bermúdez-Otero's work; though confined to general Western Germanic, the discussion would be of use to Old English treatments, but since the latter is only employed as a means of establishing contrast with Old Frisian, this discussion is omitted here.

4 Old Frisian Situation

Some old Western Germanic languages have received very thorough investigation. One such language is Old English. Since Old English is one of the most widely studied old Germanic languages and since it is attested in an enormous medieval manuscript corpus, it is not surprising that it has received thorough descriptive treatments, cf. Campbell (1959), Hogg (1992), Hogg and Fulk (2011), etc. Though, the latter, particularly Hogg (1992), verge on generative grammar. Also, a fair deal of generative attention as such has been paid to Old English, cf. Kiparsky and O'Neil (1976), Hogg (1976, 1979, 2000), Bermúdez-Otero and Hogg (2003), Fulk (2010) and Bermúdez-Otero (2005).

Old Frisian, however, suffers from rather poor and relatively late² attestations, compared to, for instance, Old English. It has also received scarcer treatments than Old English: for instance, Heuser (1903), Boutkan (1996), Bremmer (2009), and Versloot (2004; 2008).

In Old Frisian, geminate consonants may only be found in the stem when followed by a vowel inflection (cf. Bremmer, 2009, p.61). Consider an example from the *ja*-stem declension – in <ken> [ken] 'kin' word-final geminates are missing:

Table 2: *Old Frisian ja-stems.*

	SG.	PL.
<i>Nom.</i>	ken	ken
<i>Acc.</i>	ken	ken
<i>Gen.</i>	kennes	kenna
<i>Dat.</i>	kenne	kennum

According to Heuser (1903, p.24, p.33), the only remnant of *ja*-stems in Old Frisian is the [e]-ending in the nominative and accusative singular of the noun *here* 'army'. However, geminate consonants are also a prominent remnant of the *ja*-declension. While in Old English, for instance, gemination is fully predictable, this cannot be said for Old Frisian. In Old English, the trigger of gemination, the [j] suffix

² Aside from a few runic inscriptions, the first Old Frisian attestation is from 1200 (Bremmer, 2009, p.6), which is really contemporaneous with Middle English in England, but since Old Frisian retains more original Germanic features than Middle English, it is contrasted with Old English as a typical Old Western Germanic representative.

was still present in the Input, as it surfaced whenever [r] preceded it: cf. Old English <herges> ['herjes]-GEN.SG. 'army' vs. <cynnes> [kyn:es]-GEN.SG. and <fremman> ['frem:an]-INF. 'do, make' vs. <nerian> ['nerjan]-INF. 'save'. [r] blocked gemination and caused [j] to surface, while any other consonant was geminated, rendering gemination fully predictable (Kiparsky and O'Neil, 1976; Hogg, 1976).

This is not, however, the case in Old Frisian: /j/ does not surface in Old Frisian derivations with [r], cf. <heres> ['heres]-GEN.SG. vs. <ken> ['ken:es]-GEN.SG. and <nera> ['neran]-INF. 'save' (<*narjan> vs. ['setta]-INF. 'set' (<*sattjan> (Bremmer, 2009, pp. 61–63, p.79). Gemination is completely unpredictable in Old Frisian, as /j/, the trigger of gemination, never surfaces before [r]. This means that geminate consonants have to be stored underlyingly in Old Frisian. This also implies that, while gemination was a synchronic phenomenon in Old English, it was only a historical left-over in Old Frisian.

As may be observed in table 2, geminates fail to occur in word-final position in Old Frisian. There have been attempts to argue for the existence of word-final geminates in Western Old Frisian on the basis of vowels that preceded geminates, as they sometimes underwent lengthening. However, this seems to have been nothing more than a 'tendency', according to Versloot (2008, p.91), as only very isolated (and geographically limited) examples are attested. However, according to the comparative evidence, the Ingvaconic languages, where Old Frisian belongs together with Old English and Old Saxon, did at some point contain geminate consonants in the nominative and accusative singular of *ja*-stems (Krogh, 1996, pp. 298–290). This phenomenon and the given manuscripts of the time, lead to the assumption that Old Frisian was undergoing, or had undergone, degemination in the 15th century, completely losing geminate consonants (Versloot, 2008, pp. 86–91).

Despite all this, very little has been deduced about the status of word-final geminates in Old Frisian. As noted above, there is some evidence, though very limited, that they existed. Boutkan (1996, p.41) noted that 'geminates were simplified orthographically [...] in absolute final position', <ken> vs. <kenne> being a prime example. The reason why word-final geminates should still be posited for Old Frisian, according to Boutkan (ibid., p.40), is the phrase *all a iechta* 'all in confession', where <ll> [l:] is preserved. However, late Old Frisian scribes still adhered to the Latin spelling practice, which was initially in use until the early 15th century, demanding that all geminates be spelt out (i.e., <CC> and not <C> was used) (Versloot, 2008, pp. 86–91). This means that we should expect <CC#> for word-final geminates and <VCCV> for intervocalic geminates, unless word-final geminates were *not* spelt out for some unknown reason.

The present paper is going to argue that (in most cases) there are no surface word-final geminates in Old Frisian because they have undergone degemination. The loss of geminate consonants in word-final position is a common sound change and precisely the same is assumed for Old English ([kyn:] > [kyn]) (Kurath, 1956). As we shall see in §6 and §7, this assumption will be borne out, even explaining a seeming exception to such as *all a iechta*. Accordingly, we can speak of two major periods of degemination in Old Frisian: the early (i) word-final degemination, which produced paradigms as in table 2, and (ii) the late intervocalic degemination, bridging the period between Old and Middle Frisian, after which contrastive geminates disappear from the language.

5 Specifying the INPUT

Before we may proceed to the stage of formalisation of the advances made above within Stratal OT, an incredibly simple but easily misleading aspect of INPUT specification must be discussed. In the standard moraic approach, Hayes (1989) postulates that coda consonants are not underlyingly moraic, but are licensed a mora during the process of generation (ibid., p.258), as demonstrated below:



Figure 1: *Weight-by-Position Parameter.*

This is referred to as the Weight by Position parameter, which is turned on for most Germanic languages. With geminates, however, the situation is different, as they are already dominated by a mora underlyingly: when syllabification takes place, in case of intervocalic geminates, the geminate undergoes ‘flopping’ – its ‘consonantal melody is “flopped” onto a following vowel initial syllable’ (Hayes, 1989, pp. 267–258):

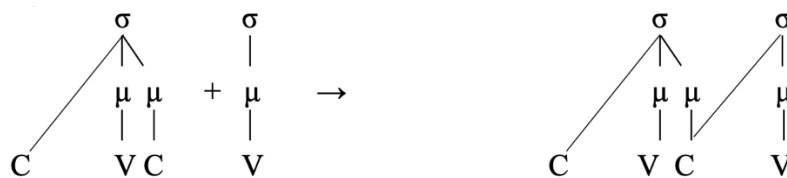


Figure 2: *Consonant flopping.*

As a side note it should also be emphasised that, as stipulated by Moraic Phonology (Hayes, 1989), word final geminates have the representation as shown in (6) – the geminate is a consonant underlyingly dominated by a mora:

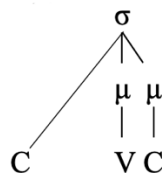


Figure 3: *Word-final geminate.*

But since Moraic Phonology has some difficulty with distinguishing simple word final geminates from degeminated consonants, the present paper presupposes that word final geminates are demorified and then attached immediately at the syllable maximal projection level, since word-final singletons in languages with word-final geminates are predicted to be light (cf. Topintzi, 2008; Ham, 2001):

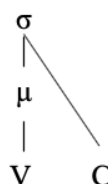


Figure 4: *Word-final singleton.*

6 Old Frisian Geminate in *ja*-stems

Old Frisian geminates are assumed to be underlyingly present here, as [j], the historical trigger of gemination, cannot be recovered from any derivational process where we still find it in Old English (cf. Old English *herjes* vs. Old Frisian *heres*), rendering geminate consonants completely unpredictable. Furthermore, Old Frisian *ja*-stem phonology is stratified, which presumes that affixation takes place at word level, while word final degemination at the postlexical level. This must be the case since: (i) word-final degemination is transparent itself, but in turn it obscures the distribution of geminate consonants in the paradigm, as predicted by Stratal OT (see §2); (ii) there is no phonetic data that we could study, but this should be the case, as the postlexical level is where phonological innovations start out in the grammar (for more elaboration see §7). This is reinforced by Boutkan's (1996, p.41) assumption that geminates are present in Old Frisian since they occur in the phrase *all a iechta*, which must be syllabified as *al.la.iech.ta*, implying that the otherwise word-final geminate in *all* is flopped to the following vowel postlexically and thus need not be degeminated. This indicates that word-final geminates were still present on the word level, which was the input for the postlexical level (in this case *all*): when *all* was followed by a vowel, postlexical flopping must have taken place, but if it was followed by nothing, it must have been degeminated, hence the [ken] vs. [ken:e] distinction (though the geminate in [ken:e] was already flopped on the word level, due to [e] suffixation).

The postlexical degemination, therefore, deletes the stem-environment for consonant flopping, thus obscuring the distribution of geminate consonants in the paradigm:

(2) Input:	/ken:/ or /ken:/ + /e/	
	[ken:]	[stem level]
	[ken] / [ken:e]	[word level]
	[ken] / [ken:e]	[postlexical level]

The flopping applies to the suffixed cases, or in case of postlexical phrases to word-final geminates followed by a vowel, while degemination to those with a zero-ending followed by nothing. This means that word-final degemination is a transparent phenomenon which, from an *n+1* level, obscures the operations of processes on an *n* level, from the level of the paradigm, in this case, the distribution of geminate consonants in the stem.

Before proceeding with Stratal OT formalisation, it is necessary to present the Markedness and Faithfulness constraints that govern the generation process of Old Frisian *ja*-stems. As expected, Faithfulness constraints will be concerned with preserving moraic structure, while Markedness constraints will inhibit geminates and marked syllabic structure:

(3) NOCMORA: 'The head of a mora must be a vowel' (Broselow et al., 1997, p.65)

NOCMORA constraint forbids geminate consonants of any kind since they are by definition headed by a mora.

(4) *GEM#: 'Geminates are disallowed in word final-position' (Davis, 2003, p.28)

*GEM# is the Markedness constraint mentioned in the first paragraph of this chapter; it will play an important role in the degemination of word final geminates on the postlexical level.

(5) IDENT^μ: ‘Let α be a segment in the input. Let β be a correspondent of α in the output. Let α be linked to n morae.

IDENT^μ = (a) \wedge (b)

(a) β is linked to n morae;

(b) β is positionally μ -licensed.’ (Bermúdez-Otero, 1999a, p.49)

IDENT^μ is a ‘macro constraint’ consisting of two ‘micro constraints’ (Crowhurst and Hewitt, 1997), which are the result of a ‘local conjunction’ (Smolensky, 1993) (in this case they are (a) and (b)). Bermúdez-Otero (1999a, p.50) defines Smolensky’s ‘local conjunction’ in the following manner: ‘a candidate c violates the micro constraint (a) \wedge (b), if, and only if c violates both micro constraint (a) and micro constraint (b)’. This implies that a segment present in the OUTPUT must be either (i) attached to the same number of moras as in its INPUT, (ii) or it is attached to no moras in the INPUT because it is licensed by Weight by Position. This constraint will be easily violated in our case since we are dealing with underlying geminates, which implies that the micro constraint (b) will have little effect on the analysis.

The ranking of the constraints suggested above, can account for the generation of Old Frisian *ja*-stem nouns. The following tableau illustrates the *stem level*:

INPUT – STEM LEVEL		Candidates	IDENT ^μ	*GEM#	NOCMORA
/ken ^μ /	☞	‘ken ^μ ’		*	*
		ken	*!		

Figure 5: *Stem level*.

The constraint IDENT^μ dominates *GEM# and NOCMORA on the stem level (IDENT^μ >> *GEM#, NOCMORA). The latter two are not in a dominance relationship, i.e. their ranking with respect to each other does not matter in light of the material presented here. While IDENT^μ prohibits the degeminated form, *GEM# and NOCMORA prohibit geminated consonants (in fact, the candidates that incur a violation of *GEM# represent a subset of those that incur a violation of NOCMORA). Let us now turn to the word level:

INPUT – WORD LEVEL		Candidates	IDENT ^μ	*GEM#	NOCMORA
NOM.SG. [ken ^μ]	☞	‘ken ^μ ’		*	*
		‘ken’	*!		
GEN.SG. [ken ^μ +es]	☞	‘ken ^μ .nes’			*

		ke.nes	*!		
DAT.SG. [ken ^μ +e]	↗	‘ken ^μ .ne			*
		‘ke.ne	*!		

Figure 6: Word level.

The general ranking of the word level *ja*-stem constraints would be IDENT^μ >> *GEM#, NOCMORA, which makes it identical to the stem level ranking. *GEM# is vacuously satisfied in the genitive and dative, but its ranking is absolutely crucial to generate word-level word-final geminates, as in the nominative.

To cater for the lack of geminates finally in Old Frisian, we have to accommodate a re-ranking of the markedness constraint *Gem# to dominance status on the postlexical level:

INPUT – POSTLEXICAL LEVEL		Candidates	*GEM#	IDENT ^μ	NOCMORA
NOM.SG. [ken ^μ]		‘ken ^μ	*!		*
	↗	‘ken		*	
GEN.SG. [ken ^μ .nes]	↗	‘ken ^μ .nes			*
		ke.nes		*!	
DAT.SG. [ken ^μ .ne]	↗	‘ken ^μ .ne			*
		‘ke.ne		*!	

Figure 7: Postlexical level.

The postlexical-level ranking in *ja*-stems, which licenses the degemination (i.e., demora-fication) of *kenn* to *ken*, is the following: *GEM# >> IDENT^μ >> NOCMORA. Again, note that the *GEM# >> IDENT^μ ranking is crucial for licensing word-final degemination, while the fact that these constraints dominate NOCMORA is crucial for maintaining the intervocalic geminates in the genitive and dative cases. As mentioned before, only a subset of geminate structures is banned from the grammar. This is precisely what predicts the phrase *all a iechta* (cf. §4), its syllabification being *al.la.iech.ta*, which means that our ranking still permits this phrasal configuration, as postlexically, it reveals no word-final geminates.

This postlexical re-ranking can, therefore, account for the word final degemination in Old Frisian up to the systematic degemination, which occurred in the 15th century. Needless to say, what applies to NOM.SG. also applies to ACC.SG., and also NOM. and ACC.PL.; the same goes for the other cases in plural, the ranking of which is isomorphic with the singular.

7 Why Strata?

As is obvious from the tableaux above, positing strata may at first seem redundant since we could simply posit *GEM# >> ... on the word level or in parallel OT as the dominating constraint, which the candidates of the dative and genitive cases would vacuously satisfy. However, with a stratified approach, we can, at the same time, straightforwardly account for the diachronic state of geminates in Old and Middle Frisian.

As noted in §3, stratification offers us the ability to single out the ‘innovatory’ features in phonology, which, after the stabilisation of a phonetic process, are manifested as a re-ranking of markedness constraints on the postlexical level, which may, eventually, trigger re-ranking on the lower levels, as well, due to a restoration of ‘ranking uniformity throughout the grammar’ (Bermúdez-Otero, 1999a, p.99; Kiparsky, 1995). We may therefore speak of ‘domain narrowing’: a postlexical innovation may percolate to the Word level, eventually to the stem level and finally enter the Lexicon (Bermúdez-Otero, 2007; Bermúdez-Otero and Trousdale, in press):



Figure 8: *Sound change pathway.*

Thus, all we have to posit is *GEM# as the dominating markedness constraint on: (i) THE WORD LEVEL to account for a unified lack of geminates word-finally; (ii) THE STEM LEVEL to account for regular degemination of the ‘intervocalic’ segments in the 15th century (though, they are degeminated as word-final on the stem level, or rather ‘stem final’). The ensuing step (iii) is Lexicalisation whereupon degeminated forms enter the Lexicon and *ken-* already comes specified in the INPUT, giving rise to the late Old Frisian and Middle Frisian state of affairs concerning geminate consonants.

This explains why positing postlexical degemination seems to be the right solution: only after a phonetic rule has undergone stabilisation and has ushered in a change of grammar (i.e., in postlexical phonology) does it begin to percolate to the lower strata and start exhibiting ‘morphosyntactic sensitivity’ (Bermúdez-Otero, 2007) – in our case, this refers to the later stem-level degemination. This is precisely why the initial degemination must be treated as postlexical – it was a new process that exhibited no morphosyntactic sensitivity and applied across the board (i.e., anywhere finally on the postlexical level, but later it specifically applies to uninflected stems).

Such a solution also shows that Stratal OT can account for all degemination processes in Old Frisian by simply promoting a single markedness constraint, viz. *GEM#, to dominance status, which is permitted if the three strata are posited, as with this stem forms can be separated from the fully inflected values. Such an approach should also work for other instances of gemination-degemination in Old Frisian other than *ja*-stems, i.e. *jō*-stems and weak class 1 verbs, all of which are based on diachronic /j/-suffixation.

8 Conclusion

The Old Frisian examples of gemination and degemination have shown that they are accountable under the framework of Stratal OT in a straightforward fashion. The paradigm of *ja*-stems is analysed for its status of geminate consonants, the degemination of which in word-final position triggers paradigm opacity. Though this is not an issue for parallel OT, it is argued that by positing INPUT specified geminates with word level suffixation and postlexical degemination, we can account for the status of this innovatory phenomenon and consider it as the culprit for further developments on the lower strata, or, in other words, we license both phases of degemination by re-ranking only one markedness constraint.

References

- Bremmer, R. H. (2009). *An Introduction to Old Frisian: History, Grammar, Reader, Glossary*. Amsterdam: John Benjamin's Publishing Company.
- Bermúdez-Otero, R. (1999a). *Constraint interaction in language change: quantity in English and Germanic*. PhD Thesis. Manchester: University of Manchester.
- Bermúdez-Otero, R. (1999b). Opacity: Evidence from West Germanic Geminatation. In *Handout: Spring Meeting of the Linguistics Association of Great Britain*. Retrieved from Bermudez Otero: <<http://www.bermudez-otero.com/lagb.pdf>>.
- Bermúdez-Otero, R. (2003). The acquisition of phonological opacity. In J. Spenader, A. Eriksson & Ö. Dahl (Eds.), *Variation within Optimality Theory: Proceedings of the Stockholm Workshop on 'Variation within Optimality Theory'* (pp. 25–36). Stockholm: Department of Linguistics, Stockholm University.
- Bermúdez-Otero, R. (2005). *A-stems in West Saxon: Synchrony* [Master's Thesis]. University of Newcastle-upon-Tyne.
- Bermúdez-Otero, R. (2007). Diachronic phonology. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 497–517). Cambridge: Cambridge University Press.
- Bermudez-Otero R. & Hogg R. M. (2003). The actuation problem in Optimality Theory: phonologization, rule inversion, and rule loss. In D. Eric Holt (Ed.), *Optimality Theory and language change* (pp. 91–119). Dordrecht: Kluwer.
- Bermúdez-Otero, R. and Trousdale, G. (in press). Cycles and Continua: On Unidirectionality and Gradualness in Language Change. In T. Nevalainen and E. C. Traugott (Eds.), *Handbook on the History of English: rethinking and extending approaches and methods*. Oxford: Oxford University Press.
- Boutkan, D. (1996). *A concise grammar of the Old Frisian dialect of the First Riustring Manuscript*. Odense: Odense University Press.
- Broselow, E., Chen, S. and Huffman M. (1997). Syllable Weight: convergence of phonology and phonetics. *Phonology*, 14(1), 47–82.
- Campbell, A. (1959). *Old English Grammar*. Oxford: Clarendon Press.
- Crist, S. J. (2001). *Conspiracy in Historical Phonology* [Doctoral dissertation]. University of Pennsylvania.
- Crowhurst, M. & Hewitt, M. (1997). Boolean Operations and Constraint Interactions in Optimality Theory [Master's thesis]. University of Rutgers.
- Davis, S. (2003). The Controversy over Geminates and Syllable Weight. In C. Féry and R. van de Vijver (Eds.), *The Syllable in Optimality Theory* (pp. 77–99). Cambridge: Cambridge University Press.
- Ham, W. H. (1998). A new approach to an old problem: Geminatation and constraint reranking in West Germanic. *The Journal of Comparative Germanic Linguistics*, 1, 225–262.
- Ham, W. H. (2001). *Phonetic and phonological aspects of geminate timing*. Routledge: New York.
- Hayes, B. (1989). Compensatory Lengthening in Moraic Phonology. *Linguistic Inquiry*, 20(2), 253–306.
- Heuser, W. (1903). *Altfriesisches Lesebuch: Mit Grammatik und Glossar*. Heidelberg: Carl Winter's Universitätsbuchhandlung.
- Hogg, R. M. (1976). The Status of Rule Reordering. *Journal of Linguistics*, 12 (1), 103–123.
- Hogg, R. M. (1979). Analogy and Phonology. *Journal of Linguistics*, 15(1), 55–85.
- Hogg, R. M. (2000). On the (non-)existence of High Vowel Deletion. In A. Lahiri (Ed.), *Analogy, levelling, markedness: principles of change in phonology and morphology* (pp. 353–76). Berlin: Mouton de Gruyter.
- Hogg, R. M. (1992). *A Grammar of Old English: Volume 1: Phonology*. Chichester: Wiley-Blackwell.
- Hogg, R. M. & Fulk, R. D. (2011). *A Grammar of Old English: Volume II: Morphology*. Chichester: Wiley-Blackwell.

- Fulk, R. D. (2010). The Roles of Phonology and Analogy in Old English High Vowel Deletion. *Transactions of the Philological Society*, 108(2), 126–144.
- Kiparsky, P. & O’Neil, W. (1976). The Phonology of Old English Inflections. *Linguistic Inquiry*, 7 (4), 527–557.
- Kiparsky, P. (1982). From cyclic phonology to Lexical Phonology. In H. van der Hulst and N. Smith (Eds.), *The structure of phonological representations, part 2* (pp. 131–76). Dordrecht: Foris Publications.
- Kiparsky, P. (1985). Some consequences of Lexical Phonology. In C. Ewen and E. Kaisse (eds.) *Phonology Yearbook 2* (pp. 85–138). Cambridge University Press.
- Kiparsky, P. (1995). The Phonological Basis of Sound Change. In J. A. Goldsmith (ed.), *The Handbook of Phonological Theory* (pp. 640–670). Cambridge, MA: Blackwell Publishers.
- Kiparsky, P. (1998). Sievers’ Law as Prosodic Optimization. In J. Jasanoff, H. C. Melchert, and L. Olivier (Eds.), *Mir Curad: Studies in Honor of Calvert Watkins*. Innsbruck: Innsbruck Beiträge zur Sprachwissenschaft.
- Kiparsky, P. (In press). *Paradigm Effects and Opacity*. Stanford: CSLI monograph.
- Kiparsky, P. (2000a). Opacity and Cyclicity. *The Linguistic Review*, 17, 351–367.
- Kiparsky, P. (2000b). Analogy as Optimization: “exceptions” to Sievers’ Law. In Aditi Lahiri (ed.), *Analogy, levelling, markedness: Principles of Change in Phonology and Morphology*. Trends in Linguistics. Studies and Monographs 127 (pp. 15–47). Berlin: Mouton de Gruyter.
- Kiparsky, P. (2003a). Finnish Noun Inflection. In D. Nelson and S. Manninen (Eds.), *Generative Approaches to Finnic and Saami Linguistics*. Stanford: CSLI Publications.
- Krogh, S. (1996). *Die Stellung des Altsächsischen im Rahmen der germanischen Sprachen*. Vandenhoeck & Ruprecht: Göttingen.
- Koontz-Garboden, Andrew. (2001). *Tiberian Hebrew spirantization and related phenomena in stratal OT* [Master’s thesis]. Stanford University.
- Kurath, H. (1956). The loss of long consonants and the rise of voiced fricatives in Middle English. *Language* 32, 435–45.
- McCarthy, J. J. (2002). *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University Press.
- Murray, R. W. & Vennemann, T. (1983). Sound Change and Syllable Structure in Germanic Phonology. *Language*, 59(3), 514–528.
- Prince, A. & Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar* [Master’s thesis]. Rutgers University and University of Colorado.
- Ringe, D. (2009). *From Proto-Indo-European to Proto-Germanic: A Linguistic History of English: Volume I*. New York: Oxford University Press.
- Smolensky, P. (1993). *Harmony, markedness, and phonological activity*. Retrieved from *Rutgers Optimality Archive*: <<http://roa.rutgers.edu/article/view/88/>>.
- Topintzi, N. (2008). On the existence of moraic onset geminates. *Natural Language and Linguistic Theory*, 26 (1), 147–184.
- Versloot, A. P. (2004). Why Old Frisian is Still Quite Old. *Folia Linguistica Historica*, XXV/1–2, 257–302.
- Versloot, A. P. (2008). *Mechanisms of Language Change: Vowel Reduction in 15th Century West Frisian*. University of Groningen: LOT.

Acknowledgments

I would like to thank Janez Orešnik, Tatjana Marvin and Peter Jurgec for discussing my work with me.

On Tocharian Exceptionality to the Centum-Satem Isogloss

Richard Littauer

University of Malta

Abstract. I argue that there may have been pressures inhibiting phonological change in Tocharian, particularly in relation to the centum/satem Indo-European split. Firstly, if the overall complexity of language strata may change over time (Sampson, 2009), syntactic, morphological, and lexical changes may have inhibited a change in the phonemic inventory of the language. Secondly, I examine the historical evidence for other languages in historical Central Asia in order to show that contact languages may have also horizontally influenced the centum nature of Tocharian. Thirdly, I posit that the merge of the PIE dorsal consonants may have been influenced by the decline in the general use, domain, and range of Tocharian, which would be in line with recent research correlating small phonemic inventories with small language communities (Hay & Bauer, 2007). Tocharian therefore may not have been susceptible to a weak dialect wave, or to a possible early branching of IE. This paper takes a new perspective on the history and development of Tocharian, and in turn on the nature of the ‘textbook’ isogloss example.

Keywords: Tocharian; centum; satem; contact; isogloss

1 Introduction

The *centum* and *satem* division of the Indo-European (IE) family is a near perfect isogloss. In Proto-Indo-European (PIE), there were three sets of phonologically distinct velar consonants: labiovelars, velars, and palato-velars, each with voiced and voiceless and aspirated variants. In modern IE languages, the treatment of the palatal stops differs. There is a split, where western IE branches, such as Italic, Celtic and Germanic, feature a merge of the palato-velars and the velars (centum languages), and eastern IE branches, such as Indo-Iranian and Balto-Slavic, represent the palatal stops as sibilants, and feature a merge of the labiovelars and plain velars.

‘The word for ‘hundred’ is the shibboleth: it appears as Latin *centum*, Greek *hekatón*, English *hundred* (with $h < x < k$), Old Irish *cēt*, but as Sanskrit *śata*, Avestan *satəm*, Lithuanian *szimtas*, and Russian *sto*.’ (Pulleyblank, 2002, p. 456)

There are two exceptions: Tocharian, from the 6th–9th centuries in the Xinjiang Tarim Basin in western China, and Hittite, an older language from central Anatolia. Tocharian and Hittite are two centum languages located within the *satem* language area. Pulleyblank (2002, p.456) notes the cognate of ‘centum’ and ‘satem’ in Tocharian: ‘The Tocharian word is *kānt* in the Eastern A dialect and *kante*, *kānte* in the western B dialect’ (here, ‘*kānte*’ is the archaic Tocharian B. form). Two theories are normally used to justify this anomaly: either there was a late split along geographical lines after Tocharian and Hittite would have been influenced, or a growing dialect wave from Mesopotamia or above the Black Sea did not reach the Tarim Basin, or Europe, due to the long distances and geography involved. In the second theory, Hittite would not have been affected, as it was already dormant or extinct by the time the wave would have begun (Lockwood, 1969, p. 31). It should be noted that while Hittite is a centum language, Luvian, a sister language discovered recently, kept the three sets of velars distinct, which suggests that Proto-Anatolian might have, as well (cf. Melchert, 1987; Kloekhorst, 2008).

Regardless of whether change occurred in the centum languages due to either a late split or a dialect wave, ‘it is now clear that the centum languages share nothing other than a failure to participate

in the palatalisation of the palatal series, and as such they cannot be held to be a subgroup of PIE' (Clackson, 2007). The remaining question, then, is not how to classify these languages together, but to understand the nature of this pronunciation change. Taking this view, the centum languages are the conservative remnants of a satem change. If this is the case, why is it that Tocharian did not also palatalise with the other satem languages nearby?

Following the Neogrammarian hypothesis for language change, it is taken for granted that sound changes occur regularly and completely. However, for the Tocharian language this assumption may not be easily defensible, due to its peculiar history. Tocharian, as we know it, is extant today only in a few thousand texts, which are mostly liturgical in nature. There are two Tocharian dialects, suitably named Tocharian A and B. As well as being phonologically different from modern, surrounding Indo-Iranian IE languages, both of the dialects show a radical reorganisation of the PIE noun morphological inflections, among other grammatical anomalies.

Historically, although the Tocharian manuscripts are dated from the 6th to the 8th centuries (and perhaps as early as the 4th century (Peyrot, 2008, p. 205)), the Tocharian culture is referenced much earlier, notably by Ptolemy, Strabo, and Apollinorus (Sinor, 1963, p. 151). Various migration theories suggest the Tocharians were emigrants from central or west Eurasia, although archaeological evidence for the migration is currently non-existent (Pulleyblank, 2006, p. XII.416). Another theory suggests that the Tocharians migrated from the Afanasievo culture in Siberia, which has been suggested to have spoken an early IE language, before settling in the Tarim Basin. (Pulleyblank, p. 2006, p. XIII.7). However, without archaeological evidence, all that can be definitively known is that there were two separate communities, with A being more conservative linguistically. It is unclear, even, exactly what communities outside of the two recorded in the manuscripts spoke Tocharian, as the Yüeh-chih and Wu-sun are thought to be speakers of Tocharian languages (Mallory & Mair, 2000, p. 278).

As the origins of Tocharian are unclear, and as the extant history of the language is relatively short, any theoretical work on the nature of the centum classification must depend on reconstructed evidence from the manuscripts, or must depend upon knowledge of factors influencing language change in other, similar situations. There is enough information on the Tocharian speakers to know that the communities were small; their use of the language was in decline during the written periods, before finally being extinguished by neighbouring Turkic languages, and that the Tocharian language had already undergone significant changes morphologically and syntactically, but not phonologically (in the centum features, although the phonological system in other respects did change), before being written down in the Tarim Basin.

With this in mind, I explore here three separate ways of viewing language change in Tocharian which have not been explored sufficiently in the current literature, to the best of my knowledge. Firstly, I take a holistic view of language complexity, where each level only changes in relation to the language as a whole, and where the overall complexity of a language may change over time (Sampson, 2009). If complexity is viewed as a changing variable, it follows that a language may alter its phonological structure minimally while drastically changing its grammatical structure (or vice versa). Examples of this sort of shift can be found in several contact-influenced languages in Kupwar, India, and in the Bantu language Ma'a. Tocharian may be another example, as the morphology and syntax diverged from PIE.

Secondly, the neighbouring languages which might have influenced Tocharian, about which there is some information in the historical record, do not feature a merge of the palato-velars and velars. Given this, a possible approach to the centum problem would be that the phonology of Tocharian may have resisted palatalisation, unlike its related languages, in order to maintain distinctiveness in manner of speaking from neighbouring (both related and unrelated) languages, which would go against the split theory for centum PIE languages, which holds that there was only one origin for the phonological change. If there were influences that would bring about two changes, the dating of the split theory would become irrelevant, when considering this single variable out of all others which might be used to date

language change. Both of these suggested hypotheses hinge upon viewing language or language strata as holistic, which goes against both linguistically-informed intuition and the majority of the research into Tocharian, but is not unprecedented.

Thirdly, the size of the language community may have had a direct influence on the phonemic inventory of the language. The amount of archaeological evidence and Tocharian texts indicate that the communities were small, and that the language was in decline. Both size and changes in use have elsewhere been noted to influence inventory size, and this influence may have been enough to cause an independent merger, separate from the western IE languages. If this is the case, the split theory is again irrelevant, and the wave theory can be discarded as a major influence as it would not be able to completely influence the language, even if the Tocharian language was within the reach of the isogloss wave. As the age of the culture, and inferentially, the language, is often used along with its status as a centum language to posit that it split from PIE relatively early, perhaps in the same timescale as Hittite, any argument which could negate the need for an early split is worth considering as a viable alternative. By showing that these are valid possibilities, it is hoped that future work will be less hesitant in trying different theoretical starting points when it comes to historical linguistics.

2 Variation in Language Complexity

Stating that language complexity is variable, either between languages or within a language, is not without controversy. Indeed, much of the controversy comes from differing definitions of complexity, which is far beyond the scope of this paper. Here, complexity can be understood as a general term indicating the amount of variables in a system and the diversity and scope of interactions between them — as this is a loose definition, all suggestions based on complexity here are to be taken with a grain of salt, and are to be understood as possible ways of viewing the research instead of conclusions based on the predicative power of complexity as a linguistic variable. While anecdotally it is clear to second-language learners that some languages are more difficult to learn or more complex than others, linguists traditionally have viewed languages as equally complex. As Sampson et al. (2009) note in the aptly-named volume, *Linguistic Complexity as an Evolving Variable*, many linguists hold that:

‘...impressionistically it would seem that the total grammatical complexity of any language, counting both morphology and syntax, is about the same as that of any other. This is not surprising, since all languages have about equally complex jobs to do, and what is not done morphologically has to be done syntactically.’ (Hockett, 1958, p. 180–1, as quoted in Sampson et al., 2009, p. 2)

This notion is recently being contested, often by linguists such as Gil (2009), working on syntax in Riau Indonesia, or Everett (1986), working on Pirahã — in both cases, the languages are clearly different from a western, Indo-Europeanist perspective, and it isn't immediately clear that they are equally complex.

Another controversial statement is that there are tradeoffs – word order in Latin, which is relatively free, is much less constrained than word order in English. However, the morphology is much more complex in Latin than in English. While the level of complexity may differ itself between languages (and between speakers), for Tocharian the important point here is that there are differing levels of complexity between strata within a language. Hidden within both of these points is a third, implied and empirically testable assumption: language complexity for an entire language can change over time, as the strata change themselves. For instance, English is generally taken to be not as complex morphologically as Old English or French, although it is substantially based on both languages. As the

strata can move and change, the complexity of the language changes. While there is no roof or standard for complexity, there remains a tradeoff among the other strata to some extent. This can be seen clearly in languages like Pirahã, which have a heavier functional load on prosody than on phonemes, which enables the language to be whistled or hummed understandably. Here, the tonal system is more complex on a lexical level, while the phonemic inventory is inversely simple (Pirahã has one of the smallest phonemic inventories known, with around 13 phonemes (Everett, 1986)). Given the complex nature of language change, it is of course also possible for a trade-off to not happen under certain circumstances.

In the case of language areas, variance on different linguistic strata — e.g., morphology, phonology, syntax, or in the lexicon — is clearly apparent when there has been extensive language contact. For instance, in Kupwar, India, Urdu and Marathi, both Indo-European, coexist with Kannada, a Dravidian language. There has been both minimal lexical changes from the original stocks for each of the three languages and minimal phonological borrowings across languages (although Urdu and Marathi are understandably closer by nature of being from the same family.) However, syntactically, ‘the sentences can be seen as exact calques of each other’ (Myers-Scotton, 2002, p. 176). This is at odds with the normal process of borrowing, where the lexicon is borrowed first. The case of Kupwar illuminates the process of trade-offs — complexity changes to fit the surrounding environment on one strata, but may not on others.

In the case of Tocharian, strata tradeoffs which occur under the influence of other languages can be used to shed light on the centum issue. Tocharian underwent changes syntactically, morphologically, and lexically from Indo-European.

‘In Tocharian we find a total reorganisation of the case-endings system. In PIE each case had its own ending, with completely different forms for the singular and the plural. Tocharian, however, had a nominative, an ‘oblique’ case (congruent, in part, with the accusative) and a genitive case, and furthermore a number of ‘endings’ which were the same for both singular and plural.’ (Beekes, 1995, p. 92)

The other case endings were also changed, although minimally (and certainly not to the same extent as other Indic languages later, which changed into split-ergative or ergative systems). Given that there was morphosyntactic change but only minimal phonological borrowings, it is relatively straightforward to propose that changes on the syntactic, morphological, and lexical levels go along with conservative phonological changes. All such proposals must be taken lightly, however, as there were definitely some phonological changes in the development of Tocharian A and B, and Proto-Tocharian (see Ringe 1996), and this process is still not well understood.

Nevertheless, such a shift on different strata may have occurred not merely from horizontal influence in the sense of borrowing, but also may reflect the change of the language over time within the local linguistic environment. In such a case, it is possible that either the satem wave or the split may not have affected Tocharian, as other influences were at work which led to the conservation of the phonological centum realisation.

3 Language Contact in the Tarim Basin

The amount of horizontal influence on the language is not clear, due to the lack of documentation and records from the period. What is clear is that the Tocharian language was influenced by non-IE languages in the area. By looking at the phonological inventories of these languages, it is possible that part of the nature of the centum merge into [k] in Tocharian can be understood better. To do so, the origins of Tocharian must be analysed. Tocharian culture is referenced much earlier than the 6th century

by Ptolemy, Strabo, and Apollinorus, although the Tocharians to which they refer may have been different from the speakers of Tocharian A or B (Sinor, 1963, p. 151). There are two migration theories: either the Tocharians migrated from central or west Eurasia, although archaeological evidence for such a feat is non-existent, or they migrated from the Afanasievo culture of Siberia before settling in the Tarim Basin. In either event, the list of contact languages would be similar. There have been previously suggested sources for borrowings, namely ‘nomadic pastoralists speaking Ural-Altaic languages’ (Sinor, 1990). But ‘the major event that led to the Turkicisation of Xinjiang was the collapse of the nomadic steppe empire of the Turkic-speaking Uighurs’ (Pulleyblank, 2002, p. 45). Although the Turkic domination of the area occurred at the end of the Tocharian era, it can be assumed that those languages were in the area beforehand.

As well, there would have been more Indic, Dravidian, and especially Sino-Tibetan languages in the area, as Tocharian continued on ‘into the period of the Tang Dynasty (AD 618-907) when the Chinese first regained the Tarim basin and then lost it in the face of Tibetan and, subsequently, Turkish incursions’ (Mallory & Mair, 2000, p. 272). In any event, there were a plethora of languages in the area. Six centuries previously, Mithridates VI of Pontus reputedly knew all twenty-two languages in his Anatolian empire. The languages which we knew were contemporary with Tocharian and geographically close is itself a long list: Bactrian Iranian, Khotanese Saka, Old Persian, Han, Tang, Shang, and Chou Old and Middle Chinese variants, Sogdian, Greek, Uighur and Kyrgyz Turkish, Ossetic, Avestan, Tai, Prakrit, Tibetan, Kuchean, Burushaski, Scythian, and Cimmerian (Pulleyblank, 2002).

Listing languages in the area is not immediately useful. This is done to show the linguistic complexity of the area; in such a diverse linguistic environment, it is possible that the Tocharians may have affected their language in order to maintain distinction from others, as a linguistic realisation of a cultural movement. This is not unprecedented: the Ma’a language has one idiosyncratic phoneme, a voiceless lateral fricative, which speakers use ‘to emphasise the differentness of their other language [...] they sometimes introduce it into Bantu words.’ (Thomason, 2001) In this vein, the velar merging might then have been used as a way of keeping distinct from other sounds in the region. As well, allophonic variants might have been minimalised to retain maximum distinction. And ‘if we remember that in the Han period Chinese dental affricates are used to represent foreign palatals’ (Pulleyblank, 2002, p. IX.19), while the neighbouring western PIE languages would have also used an affricate [s], then the centum pronunciation could have been used intentionally as a linguistically distinctive phoneme. Any intentional use of distinctive phonemes would depend on a perceived threat from differing languages and cultures — something we have already established would not be surprising in the area, given the amount of neighbouring languages and successive invasions.

This argument must be taken with a grain of salt — the centum/satem difference was only a small part of the IE phonological inventory and complex history, and it is very improbable that speakers of any IE language would consciously use it as a marker of linguistic identity. That having been said, it is possible that the centum pronunciation may have been latently culturally influenced precisely to keep Tocharian separate from neighbouring languages, which include the neighbouring satem languages.

4 Demographics and Phonological Merging

Along with strata tradeoff influence and distinction maintenance, there is a third, to my knowledge unproposed, influence on the centum nature of Tocharian: that the demographic changes within the Tocharian culture may have had an influence on any change in the phonological inventory. It is clear from the archaeological record that the Tocharian population was never exceedingly large. As such, it may have been influenced by the ‘founder effect,’ where the linguistic diversity of a subset of a language

community which migrates is logically less than the diversity and variation in the larger, home community (cf. Atkinson, 2011). ‘It is a commonly noticeable phenomenon of dialect geography that a language tends to show greater variety closer home in its essential linguistic features than in its country of migration’ (Sinor, 1990, p. 154). This would hold for either of the two previously mentioned migration theories. Going with this, research correlates phonemic inventory size with the size of the language community (Hay & Bauer, 2007). Given this, it is possible that a shrinking in the Tocharian population — ultimately terminal, as Tocharian was moribund already in the 9th century — caused a decrease in the phonemic diversity of the language. In this case, the frication of the satem languages may have had minimal effect on the language, as it would involve the assimilation of more phonemes.

Tocharian may have already been showing signs of decaying use in the preserved documents. Tocharian A is found almost exclusively in liturgical documents, and is noticeably different from Tocharian B. While this is not enough to show that the language was moribund — consider that finding a Latin text in an Italian church does not indicate that Italian is dying — the death of the language after the 9th century, along with other features of the language, may point to a decrease of use. Conversely, the fact that the language was written down may show vitality — in such a case, it can be assumed that this argument may not be without detractors. However, the marked difference between A and B shows a loss of forms, as ‘although both Tocharian languages have seen fairly brusque losses of their earlier endings, Tocharian A reveals far greater loss of syllables and endings’ (Mallory & Mair, 2000, p. 272). It is worth noting that ‘attrition, the overall simplification and reduction of a language's linguistic structures, without concomitant complication elsewhere in the system, occurs only as a prelude to language death’ (Thomason, 2003, p. 704). The difference between the two dialects, as stated before, can be seen in the main shibboleth for the centum isogloss: ‘The Tocharian word is *kānt* in the eastern A dialect and *kante*, *kānte* in the western B dialect’ (Pulleyblank, 2002, p. 456). Dialect A was found in the area of modern day Ürümqi, and is only attested in Buddhist, Manichean, and other religious and political documents, written in the borrowed Sanskrit script Brāhmī, while B was more widespread, and also had secular corpora, such as trade lists. Interestingly, ‘Tocharian B, the ‘living’ language, was on the whole the more conservative of the two languages’ (Mallory & Mair, 2000, p. 277). This conservation may be due to proactive conservation of the language by the speakers, or due to the migratory history of the language and the founder effect.

The central point is that Tocharian was a language family with an idiosyncratic history: it was the result of a migration to a new linguistic environment (with different contact languages than many of the other IE languages), with two languages which lasted for only a short time, and possibly showed signs of being moribund when they were written down. As the speaking community shrank, it can be assumed the language degraded further. In such a case, the influence of a dialect wave or of a split may have been minimal, and the merge of the velars may have occurred as a side effect of the shrinking community. It may not be a matter of geographical separation which caused the centum anomaly, but rather that the language was resistant to phonemic change already, and never succumbed to the external influence that affected the satem languages. However, it should again be noted that there are counter arguments to be considered. For instance, Tocharian B must have been living and well for several centuries, from our extant documents; the velar merge may have been part of wider phonological processes involving merging and palatalisation earlier in the history of Tocharian (cf. Ringe, 1996, pp. 39f, 102ff; and the Tocharians may have had closer influence from western IE languages before migrating to the Tarim basin (Hackstein, 2005).

5 Conclusion

Without more substantial data on Tocharian and the surrounding languages which may have influenced it, triangulated with archaeological and historical information, it is difficult to posit any hypotheses which explain fully the anomaly of the centum classification. Here, I have argued for three alternative ways of viewing the problem: a tradeoff in linguistic complexity that meant that the phonology was minimally influenced while the lexicon and syntax were, the influence of neighbouring languages and the possibility of retaining an older inventory as a means of cultural distinction, and the possibility that demographic and geographic changes in the language community influenced the susceptibility of the language to influence. In each case, the power of a dialect wave or of a clear split as ways of understanding the centum nature of Tocharian is lessened in the face of individual properties of Tocharian. It is hoped that these ways of viewing the data are enlightening, and that they influence future researchers working on Tocharian by encouraging them to think of alternative viewpoints for language contact and language change situations before deciding on a single, unified, and overly explanatory theory.

References

- Atkinson, Q. (2011). Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. *Science*, 332(6027), 346-349.
- Beekes, R. (1995). *Comparative Indo-European Linguistics*. Amsterdam: John Benjamins Publishing Company.
- Clackson, J. (2007). *Indo-European Linguistics: An Introduction*. Cambridge: Cambridge University Press.
- Everett, D. (1986). Pirahã. In D. Derbyshire & G. Pullum (Eds.), *Handbook of Amazonian Languages* (pp. 211, 326). Berlin: Mouton de Gruyter.
- Gil, D. (2009). How Much Grammar Does It Take To Sail A Boat? In G. Sampson, D. Gil & P. Trudgill (Eds.), *Language Complexity as an Evolving Variable* (pp. 19-34). Oxford: Oxford University Press.
- Hackstein, O. (2005). Archaismus oder historischer Sprachkontakt. Zur Frage westindogermanisch-tocharischer Konvergenzen. In: Gerhard Meiser, Olav Hackstein (ed.), *Sprachkontakt und Sprachwandel* (pp. 169-184). Wiesbaden: Reichert.
- Hay, J., & Bauer, L. (2007). Phoneme inventory size and population size. *Language*, 83(2), 388–400.
- Kloekhorst, Alwin. 2008. *Etymological Dictionary of the Hittite Inherited Lexicon*. Leiden Indo-European Etymological Dictionary Series 5. Leiden–Boston: Brill.
- Lockwood, W. (1969). *Indo-European Philology*. London: Hutchinson of London.
- Mallory, J., & Mair, V. (2000). *The Tarim Mummies*. London: Thames and Hudson.
- Melchert, H. C. (1987). PIE velars in Luvian. In: C. Watkins (Ed.), *Studies in Memory of Warren Cowgill*, (pp. 182- 204). Berlin/New York: de Gruyter.
- Myers-Scotton, C. (2002). *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University Press.
- Peyrot, M. (2008). Variation and Change in Tocharian B. *Leiden Studies in Indo-European*, 15. Amsterdam–New York: Rodopi.
- Pulleyblank, E. (2002). *Central Asia and Non-Chinese Peoples of Ancient China*. Hampshire, UK: Ashgate.
- Ringe, D.A. Jr. (1996). *On the Chronology of Sound Changes in Tocharian. Volume 1: From Proto-Indo-European to Proto-Tocharian*. American Oriental Series 80. New Haven: American Oriental Society.

- Sampson, G., Gil, D., & Trudgill, P. (Eds.) (2009). *Language Complexity as an Evolving Variable*. Oxford, UK: Oxford University Press.
- Sinor, D. (1963). *Introduction a l'étude de l'Asie centrale*. Wiesbaden: Harrasowitz.
- Sinor, D. (Ed.) (1990). *The Cambridge History of Early Inner Asia*. Cambridge: SMC Publishing Inc.
- Thomason, S. (2001). *Language Contact*. Edinburgh: Edinburgh University Press.
- Thomason, S. (2003). Contact as a Source of Language Change. In B. Joseph & R. Janda (Eds.), *The Handbook of Historical Linguistics*, 23 (pp. 687–712). Oxford, UK: Blackwell Publishers.

Acknowledgements

I would like to thank Jim Hurford for his helpful comments on my original paper, David Arnold for his repeated proofing and support, and Neil Ashton and several anonymous reviewers for their very helpful comments.

Reanalysis and Regularisation: The Development of Morphological Structure

Gwilym Lockwood

School of Oriental and African Studies, University of London

Abstract. This paper explores the changing nature of English morphology by examining the tendency of English verbs to “regularise” - that is, the tendency of irregular verbs to be gradually reanalysed into regular verbs according to English morphological structures. The regular English past tense is the suffix “-ed”, although historically there was a far more complex system of rules for forming the past tense, such as word-internal vowel change. However, this irregular system is gradually being replaced by the regular past tense. In order to show this, I will give a Government Phonology account of analytic and non-analytic morphology and how it interacts with phonology. I will then show that when a word is morphologically reanalysed, it almost always shifts from non-analytic (or irregular) to analytic (or regular) morphology. Furthermore, new verbs, such as loanwords and verbs derived from nouns or onomatopoeia, are also analysed with analytic morphology. Finally, I will explore the idea that the reanalysis of a word’s morphology is inherently linked to the frequency of its use. By referring to previous studies on evolutionary linguistics, I will show that more commonly used words have “stickier” morphology, which is reluctant to shift from non-analytic to analytic, while less commonly used words tend to shift from non-analytic to analytic much more quickly.

Keywords: morphology; reanalysis; regularisation

1 Introduction

This paper explores the changing nature of English morphology by examining the tendency of English verbs to “regularise” - that is, the tendency of irregular verbs to be gradually reanalysed into regular verbs according to English morphological structures. The regular English past tense is the suffix *-ed*, although historically there was a far more complicated system of rules for forming the past tense, such as word-internal vowel change. However, this irregular system is gradually being replaced by the regular past tense.

In order to show this, I will give an account of analytic and non-analytic morphology and how it interacts with phonology. I will then show that when a word is morphologically reanalysed, it almost always shifts from non-analytic (or irregular) to analytic (or regular) morphology. Finally, I will explore the idea that the reanalysis of a word’s morphology is inherently linked to the frequency of its use. By referring to previous studies on evolutionary linguistics, I will show that more commonly used words have “stickier” morphology, which is reluctant to shift from non-analytic to analytic, while less commonly used words tend to shift from non-analytic to analytic much more quickly.

2 Morphology: An Overview

Within Government Phonology, Kaye (1995, p. 302) proposed the two functions of *concat*, meaning to concatenate two phonological strings into one, and ϕ , meaning to apply phonology to string X. The application of these two functions is what distinguishes analytic and non-analytic morphology.

There are two types of analytic morphology. The most relevant type of morphologically complex structure to this discussion is found in compounds of a free morpheme and a bound morpheme, such as with the regular English past tense where the verb stem is free (and also constitutes the present tense)

and the past tense morpheme *-ed* is bound. This structure can be written as $[[A]B]$. This bracketed notation shows the presence of phonological domains; in this case, there are two, namely $[A]$ and the concatenation of $[A]$ and B into $[[A]B]$. In essence, phonological processes are applied to the domains individually (the ϕ function). Firstly, $[A]$ undergoes phonological processing. Then, the post-phonological $[A]$ is concatenated with B to make $[A]B$, which then itself undergoes phonological processing as a domain. This can be expressed as:

$$\phi(\text{concat}(\phi(A), B))$$

or, “apply phonological processes to A and concatenate the result with B to form a string; then apply phonological processes to that resulting string”.

I will take Kaye’s example (1995, p. 304) of the word *seep*, with the regular past tense *seeped*, to illustrate how this works.

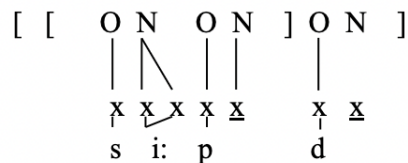


Figure 1: *seeped*.

Here, a domain-final empty nucleus is found in the domain $[A]$, *seep*. If the word *seeped* were one domain and only one domain, we would expect to see either a phonetic realisation of an unlicensed empty nucleus, for example $*[si:pəd]$. However, with analytic morphology, phonological processes such as the licensing of empty nuclei in a domain-final position apply at the domain level and are irreversible. Therefore, when a suffix is added and the ϕ function is applied to the entire domain, the phonological processes which have already occurred are not affected; however, as morpheme B , the past tense *-ed*, does not constitute a phonological domain, the assimilation of $/d/$ to $[t]$ to match the unvoiced final consonant of $[si:p]$ still applies.

The other type of analytic morphology is found in compound nouns such as *blackboard* and can be written as $[[A][B]]$, where $[A]$ and $[B]$ are individual domains, as is the concatenation of the two. This can be expressed as:

$$\phi(\text{concat}(\phi(A), \phi(B)))$$

or, “apply phonological processes to $[A]$ and to $[B]$; then concatenate the results to form a string; then apply phonological processes to that string”.

In both types of analytic morphology, the morphological structure of the word is visible to the phonology; we can see with $[si:pt]$ that there is an empty nucleus. $[pt]$ could form a legitimate rhyme-onset cluster, but the fact that there is a preceding long vowel, $[i:]$, means that putting the p into a post-rhymal nuclear complement would violate the condition of binarity within the Rhyme. Therefore, we can tell that $[si:pt]$ is morphologically complex and distinguish it from a morphologically simplex word. This is how the morphology is visible to the phonology.

Non-analytic morphology, meanwhile, is where the morphology is invisible to the phonology; that is to say, the phonology treats the concatenation of morphemes as one domain without processing any sub-domain within the word. Therefore, a non-analytic morphological structure can be written $[A$

B], where phonology only applies to the concatenation of the strings. As phonology does not apply to any individual domains, a non-analytic, morphologically complex word is not distinguishable from a morphologically simplex word; it is not phonologically parsable into separate domains, and only a knowledge of the meaning of the word will indicate its structure. To clarify; a native English speaker will automatically know that [si:pt] is the past tense of the verb [si:p], even if they have never encountered this verb, because the long vowel [i:] indicates that [pt] is not a legitimate rhyme-onset cluster. However, the same speaker would not know that [kept] is the past tense of [ki:p], or even that it is a verb in the past tense, because the short vowel [e] makes it indistinguishable from morphologically simplex words like *adept*. As it is not phonologically parsable, only a knowledge of the meaning of [kept] will give any information as to its morphological structure.

Similarly, there are irregular verbs where the past tense is formed only by changing the vowel, such as *speak* [spi:k] and *spoke* [spəʊk]. These forms are also indistinguishable from simplex structures - indeed, *spoke* [spəʊk] itself can be a simplex morphological structure when it means the long thin metal rod used to reinforce a bicycle wheel. Meanwhile, other irregular verbs where the past tense is entirely different from the present, such as *go* [gəʊ] and *went* [went], give no phonological clue to their structural identity.

This led Kaye (1995, p. 311) to propose that irregular morphology is always non-analytic. With this principle in mind, let us now examine the tendency of English irregular verbs to regularise; that is, to be reanalysed from non-analytic morphological structures to analytic; that is, for their morphology to become visible to phonology.

3 Subject to Change: Morphological Reanalysis

While English does have a set of irregular verbs, these are historical relics from English's Anglo-Saxon past – all Modern English irregulars are monosyllabic Anglo-Saxon words with Anglo-Saxon sound patterns (Pinker, 2000). There is some consistency to the changes made to derive irregular past tenses, such as *ring* becoming *rang* and *sing* becoming *sang*, which gives these rules some degree of predictability; however, it only applies to very specific cases. *String*, for example (such as in the phrase “to string somebody along”), follows the regular past tense rule and becomes *stringed*, while *bring*, another irregular, follows an entirely different irregular sound pattern and becomes *brought*.

Modern English tolerates these fossilised exceptions to the formation of the past tense, but it has by and large moved on - all new verbs which enter the English language are regular verbs. This can be seen with the verb *google*, taken from the name of the internet search engine and meaning “to look up information online”. This is conjugated regularly: *google* - *googled* - *was googled*. Similarly, verbs which derive from onomatopoeia also follow regular past tense rules, even when the words in question could easily be mistaken for monosyllabic irregular Anglo-Saxon words. The same applies for verbs which come from nouns, such as *to grandstand*, meaning “to show off, to impress an audience”. Even though the original noun, *grandstand*, is formed from the adjective *grand* and the irregular verb *stand*, the past tense is *grandstanded*, not **grandstood*.

As Modern English shows a universal preference for regularity in new additions and developments, so Modern English shows a tendency to gradually align its irregular fossils by regularising its Anglo-Saxon irregular verbs. Old English, spoken around 800 AD, had 177 irregular verbs, which was approximately 25% of all verbs in the language (Lieberman & Michel, 2007, p. 713). By the time of Middle English, which was spoken around 1200 AD, this figure had already decreased to 145; and in Modern English, only 98 verbs – less than 3% – remain irregular.

One example of an irregular verb which has regularised is *cleave*; its past tense used to be derived by creating a legitimate cluster and shortening the vowel, resulting in *cleft*; in Modern English, it follows the regular morphology to result in *cleaved*.

Cleft is a non-analytic morphological structure, indistinguishable from simplex words like *left* (as in, the opposite of right) and *deft*. As Kaye observes (1995, p. 311), [vd] is not a possible true cluster in English; even when it is preceded by a short vowel, it is always parsed as past tense, which rules out the interpretation of complex morphology by a violation of binarity with a long vowel and post-rhymal nuclear complement, as seen above in [pi:pt]. Accordingly, Old English found the closest possible true cluster to [vd], which was its devoiced counterpart, [ft]. The word [kleft] is undeniably non-analytic, as all possible indicators of its morphological complexity have been phonologically hidden. English phonology treated *cleft* as [A B], and processed *cleave* and *cleft* as two separate lexical entries.

This is no longer the case. Modern English has reanalysed the past tense of *cleave* and regularised it into *cleaved*, by applying the regular “-ed” suffix. This makes it analytic, and its function expressed as:

$$\varphi(\text{concat}(\varphi(A), B))$$

where [A] represents the verb root *cleave* and [B] represents the *-ed* regular past tense suffix. Like [si:pt] above, [kli:vd] is phonologically parsable:



Figure 2: *cleaved*.

Again, a domain-final empty nucleus is found in the domain [A], *cleave*. If the word *cleaved* were one domain and only one domain, we would expect to see either a phonetic realisation of an unlicensed empty nucleus, for example *[kli:vɔd]. However, with analytic morphology, phonological processes such as the licensing of empty nuclei in a domain-final position apply at the domain level and are irreversible. Therefore, when a suffix is added and the φ function is applied to the entire domain, the phonological processes which have already occurred are not affected; however, as morpheme [B], the past tense *-ed*, does not constitute a phonological domain, the assimilation process would still apply, although there is no change, since /d/ is itself voiced and therefore already matches [v], the voiced final consonant of [kli:v].

The same process happened with the verb *help*. In Old English, the past tense was derived simply by changing the vowel, resulting in *holp*, but in Modern English, it has regularised and become *helped*. [holp] is non-analytic; it is unparsable as morphologically complex, and so its [A B] structure is indistinguishable from a morphologically simplex structure. Its modern counterpart, [helpt], like [si:pt], is analytic; the domain-final empty nucleus in [help] is licensed and is not phonetically realised, which shows that the phonological processes apply at the domain level and are irreversible. It is therefore phonologically parsable as having the structure [[A]B].

There are some exceptions to this trend where regular verbs have become irregular, such as *sneak*. Some speakers have reanalysed the past tense to *snuck*, rather than the regular *sneaked*. Pinker (2000) believes that this is due to accidental confusion with (or possibly deliberate and creative misuse of) the pattern of the irregular verbs *stick* - *stuck* and *strike* - *struck*. A similar process has happened in some

dialects of American English with the verb *squeeze*; some speakers have reanalysed its past tense from the regular *squeezed* to the irregular *squoze*, probably arising from a confusion with the irregular verb *freeze* - *froze*. However, these new irregular forms have not replaced the regular forms, but seem to be used alongside as dialectal or even idiolectal variations; in contrast, regularised Old English irregulars (such as *holp*) are now defunct and do not occur in Modern English.

While *cleave* has been regularised, another verb which follows the same sound pattern, *leave*, remains irregular. It is not coincidental that *leave*, still irregular, is more commonly used in English than *cleave*, the now-regularised verb. The next section will explore the link between regularisation of verbs and the frequency of their usage.

4 The Relationship between Common and Regular

Two recent studies have shown that the more common a word is, the less likely it is to change. Evolutionary biologist Mark Pagel has traced the pathway of 200 meanings from the original Indo-European to 87 different modern languages. His study has shown that very frequently used words have barely changed - the word *two* is a cognate in all 87 languages studied - while less frequently used words change to the point where they are unrelated and/or unintelligible as cognates (Pagel, 2007, p. 717).

Mathematicians Lieberman and Michel have explored this further by looking strictly at the rate of change of English irregular past tenses. As noted earlier, irregular verbs account for less than 3% of all verbs in Modern English; however, the 10 most commonly used English verbs are all irregular. Lieberman and Michel listed all 177 irregular verbs from Old English and sorted them into sets, or “bins”, by the frequency at which the verbs occurred in the CELEX corpus. They found that the first two bins, 10^{-1} to 1 (i.e. a frequency of above 1 word in every 10) and 10^{-2} to 10^{-1} (i.e. a frequency of between 1 word in every 100 and 1 word in every 10), contained only irregular verbs. On the other hand, of the 12 Old English irregular verbs in the lowest frequency bin, 10^{-6} to 10^{-5} , all but one had regularised. I have reproduced below in Figure 1 the 177 irregular verbs, sorted by frequency, which makes the relationship between frequency and regularisation evident (Lieberman & Michel, 2007, p. 714).

This confirms the tendency shown by Pagel for less common words to change more quickly, but more interestingly still is the direct mathematical relationship that Lieberman and Michel have established between frequency and regularisation. They found that irregular verbs regularise at a rate that is inversely proportional to the square root of their frequency; in other words, if a verb is used 100 times less frequently, it will regularise 10 times more quickly. This explains why English’s most common verbs are irregular, and are likely to remain irregular, while less common English irregular verbs have regularised over time.

Table 1 | The 177 irregular verbs studied

Frequency	Verbs	Regularization (%)	Half-life (yr)
10 ⁻¹ –1	be, have	0	38,800
10 ⁻² –10 ⁻¹	come, do, find, get, give, go, know, say, see, take, think	0	14,400
10 ⁻³ –10 ⁻²	begin, break, bring, buy, choose, draw, drink, drive, eat, fall, fight, forget, grow, hang, <i>help</i> , hold, leave, let, lie, lose, <i>reach</i> , rise, run, seek, set, shake, sit, sleep, speak, stand, teach, throw, understand, <i>walk</i> , win, <i>work</i> , write	10	5,400
10 ⁻⁴ –10 ⁻³	arise, <i>bake</i> , bear, beat, bind, bite, blow, <i>bow</i> , burn, burst, <i>carve</i> , <i>chew</i> , <i>climb</i> , cling, creep, <i>dare</i> , dig, <i>drag</i> , flee, <i>float</i> , flow, fly, <i>fold</i> , freeze, grind, leap, lend, <i>lock</i> , melt, reckon, ride, rush, shape, shine, shoot, shrink, sigh, sing, sink, slide, slip, smoke, spin, spring, starve, steal, <i>step</i> , stretch, strike, stroke, suck, swallow, swear, sweep, swim, swing, tear, wake, wash, weave, weep, <i>weigh</i> , wind, yell, yield	43	2,000
10 ⁻⁵ –10 ⁻⁴	bark, bellow, bid, blend, braid, brew, cleave, cringe, crow, dive, drip, fare, fret, glide, gnaw, grip, heave, knead, low, milk, mourn, mow, prescribe, redden, reek, row, scrape, seethe, shear, shed, shove, slay, slit, smite, sow, span, spurn, sting, stink, strew, stride, swell, tread, uproot, wade, warp, wax, wield, wring, writhe	72	700
10 ⁻⁶ –10 ⁻⁵	bide, chide, delve, flay, hew, rue, shrive, slink, snip, spew, sup, wreak	91	300

177 Old English irregular verbs were compiled for this study. These are arranged according to frequency bin, and in alphabetical order within each bin. Also shown is the percentage of verbs in each bin that have regularized. The half-life is shown in years. Verbs that have regularized are indicated in red. As we move down the list, an increasingly large fraction of the verbs are red; the frequency-dependent regularization of irregular verbs becomes immediately apparent.

Figure 3: *The 177 irregular verbs studied.* (Lieberman & Michel, 2007, p. 714).

5 Summary

Analytic morphology is when a complex morphological structure is phonologically parsable as complex; phonological processes apply at domain-level. Non-analytic morphology is when a complex morphological structure is indistinguishable from a simplex structure, as phonological processes only apply at the entire word domain-level without detecting any morphological domains. Regular English past tense morphology is phonologically parsable, and hence is analytic; irregular English past tense morphology is not phonologically parsable and constitutes a separate lexical item from its present tense counterpart, and hence is non-analytic. New verbs, verbs derived from nouns and verbs derived from onomatopoeia all take regular, or analytic, past tense morphology, which shows Modern English's preference for analytic structures in past tense verbs. Studies on language change have shown that the number of irregular verbs in Modern English has decreased compared with Old English; some irregular verbs have regularised, which is to say that their morphological structures have been reanalysed from non-analytic to analytic. Lieberman and Michel's study has shown a mathematical link between the rate of this reanalysis and the frequency with which a verb is used. Therefore, there is an evident and predictable tendency for the morphological structure of English verbs to be reanalysed from non-analytic to analytic over time, depending on their frequency. Further research into this area could provide new insights into usage-based models.

6 References

- Hare, M & Elman, J. (1995). Learning and morphological change. *Cognition*, 56, 61-98.
- Kaye, J. (1990). 'Coda' licensing. *Phonology Yearbook*, 7(1), 301-330.
- Kaye, J. (1995). Derivations and interfaces. In J. Durand & F. Katamba (Eds.), *Frontiers of Phonology* (pp. 289-332). London & New York: Longman
- Kiparsky, P. (1982). Lexical morphology and phonology. In I. S. Yang (Ed.), *Linguistics in the morning calm: selected papers from SICOL-198* (pp. 3-91). Seoul: Hanshin

- Lieberman, E & Michel, J. B. (2007). Quantifying the evolutionary dynamics of language. *Nature*, 449, 713-716.
- Pagel, M. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, 449, 717-720.
- Pinker, S. (2000). The irregular verbs. *Landfall*, Autumn, 83-85.
- Used, The. (2004). All that I've got [song]. On *In Love and Death*. Los Angeles: Foxy Studios.

'English corrupted by Singaporeans'³: a Study of Speaker Perceptions of Colloquial Singapore English in 2012

Nicholas Marsh

University of Oxford

Abstract. The goal of this investigation is to clarify the sociolinguistic status of Colloquial Singapore English (CSE), otherwise known as Singlish, in light of its stigmatisation by the Singapore government. Using original data personally collected by surveying English-speaking Singaporeans in 2012, it will firstly consider whether Singlish exists in a lectal continuum or in a diglossic relationship with Standard Singapore English. It will then build on these findings in order to compare popular attitudes towards CSE within the speech community, with the official, anti-Singlish state line propagated in recent years.

Keywords: sociolinguistics; Singaporean English; language policy

1 Introduction

'Do not speak Singlish! If you do, you are the loser. Only foreign academics like to write about it. You have to live with it.' (former Prime Minister Lee Kuan Yew, 1994)⁴

1.1 Aims of this paper

This paper will attempt to establish whether Singaporeans share their government's anti-Singlish stance by answering the following question: how can the Singapore government's language policy during the past twenty years be reconciled with the sociolinguistic reality of the language situation in Singapore in 2012? In order to do this, it was necessary for the investigation to take two major lines of enquiry. One line is linguistic; the government line, making conscious or unconscious use of somewhat out-dated sociolinguistic models, heavily implies that Singlish exists in a strict diglossic relationship with SSE (Gupta, 1989), effectively acting as the so-called 'Low' variety (Ferguson 1959). The study seeks to establish whether this is in fact the case, or whether, instead, Singlish ought to be assessed using an 'indexical approach' (Leimgruber, 2012) as a more adaptable variety in a linguistic cline or lectal continuum, loosely based on the idea of a post-creole continuum (DeCamp, 1971), whose syntax and vocabulary may differ more or less from the standard depending on the social context of the variety's use (Pakir, 1991; Bao & Hong, 2006). The other line will examine social attitudes towards Singlish and its usage, aiming to clarify issues over the language's prestige and cultural currency in light of what some have claimed to be a divided opinion among the population (Wee, 2011, pp. 7-8). Therefore, in building a clearer linguistic definition and concurrently gaining an insight into Singaporean views on Singlish's social status, my investigation will be able to effectively tally speaker reality with government rhetoric.

³ This now famous description of Singlish was coined by former Prime Minister Goh Chok Tong in a speech given at the National Day Rally in 1999 (quoted in *The Straits Times*, 29 August 1999).

⁴ From a speech given to students at National University of Singapore (quoted in *The Sunday Times Singapore*, 31 July 1994)

1.2 Historical and linguistic context

Since the establishment of Singapore as a British trading post by Sir Stamford Raffles in 1819, the role and usage of the English language on the south-east Asian island has grown significantly in prevalence and stature, and this growth shows little sign of abating. As of 2010, almost 80% of Singaporeans claimed to be literate in English, whilst 32% spoke English as their first language at home⁵. Since the city-state's independence in 1965, English has enjoyed official language status, along with Chinese (Mandarin), Malay, and Tamil. Furthermore, it has always been the most important language of instruction in all levels of education and, in 1987, it officially became the sole medium of instruction in schools (Tickoo, 1996, p. 439).

However, the sociolinguistic reality is not so simple. Working within the 'Three Circle Model of World English' (Kachru, 1985a), Singapore is placed quite firmly in the Outer Circle of 'new' World Englishes (Crystal, 1997; Schneider, 2007; Jenkins, 2009), along with, for example, South Africa, India, and the Philippines. Due to English's co-existence with other prominent languages in Singapore, many English speakers are L2 speakers, that is to say they have learned English as their second language (ESL) in addition to their mother tongue. But over a century of British rule meant that every aspect of the state's current infrastructure – including its schools, universities, media, judiciary, and military – began life, and has generally remained, an English language institution. English's official status and obvious bureaucratic importance led to the future educated elite growing up as L1 English speakers (Platt & Weber, 1983), yet the overwhelming migration of Straits Chinese, mainland Chinese, Indians, Malays, and other non-English speakers ensured that the English spoken in Singapore has remained for a very significant period a non-native, outer circle variety of the new World Englishes. Until the arrival of the British, the island had a population of fewer than 50 permanent inhabitants and Singapore has since been a multiracial and multilingual territory, now officially composed of three ethnic groups: Chinese (74%), Malays (13%), Indians (9%), as well as a miscellaneous 'others' category (3%). The fact is that, with no real 'indigenous' population and, arguably, no real indigenous language, English has accordingly become the *de facto* national language of the republic (Tickoo, 1996, p. 431).

The growth of English on the island largely came, however, post-independence, and as a result of an economically motivated and forcefully implemented series of language planning measures carried out by the People's Action Party (PAP)⁶ from the 1970s onwards (Spolsky, 2004, pp. 170-1)⁷. Nevertheless, there is substantial variation in English proficiency across the population, both in range and depth (Pakir, 1991, p. 176). This makes terms such as 'native speaker' and 'native variety' problematic as the development of English in the modern Singaporean state (as well as its perception by various parties) has always been greatly subject to the influence of such a mixed socio-cultural context.

1.3 Language policy and language planning post-independence

⁵ All demographic information in this paper comes from Singapore Department of Statistics, Census of Population 2010 Statistical Release 1: Demographic Characteristics, Education, Language and Religion. <http://www.singstat.gov.sg/pubn/popn/C2010sr1/cop2010sr1.pdf> [accessed 28/12/11].

⁶ Since assuming power in 1965, Singapore has been uninterruptedly governed by the People's Action Party (PAP). Although officially a parliamentary democratic republic with an opposition Workers' Party, Singapore has been variously described as a 'dominant one-party system' (Chan, Quah & Seah, 1985, p. 146), a 'semi-authoritarian regime' and 'the only affluent nation in the world that is not a genuine liberal democracy' (Francesch-Huidobro, 2008, p. 4). For this reason, the terms 'government' and 'state' are used interchangeably.

⁷ For a more detailed overview of the unilateral implementation of Singapore's language policy, see Tickoo (1996).

It is worth bearing in mind that in Singapore the term ‘mother tongue’ does not have the same definition as it does in linguistic discourse; according to the state, one’s Mother Tongue (MT) is the heritage language assigned to a citizen by the government based on the ethnic group in which they are categorised, rather than the first language a speaker learns or speaks most proficiently (Lim, 2010, p. 30). It thus follows that there are only three Mother Tongues in Singapore (Mandarin Chinese, Malay, and Tamil); English is not considered, by the state at least, any Singaporean’s Mother Tongue. Since the 1970s, the Ministry of Education has proscribed that every Singaporean is required to learn English as well as their own MT: a form of language planning known as English-based bilingualism (EBB) (Tickoo, 1996, p. 438).

The reason for this curious situation, which Lim, Pakir and Wee (2010, p. 6) call ‘Singapore’s aggressively bilingual language policy’, is down to the young state’s twin key goals: economic development and the preservation of cultural heritage. English is viewed in Singapore as an important ‘neutral’ language for the purposes of interracial communication, yes, but more importantly it is promoted as an essential tool for innovation and competitiveness in science, technology, trade and education. Given Singapore’s complete lack of natural resources, English is thus viewed as one of its primary intellectual resources. This is in contrast to the arguably economically stifling post-colonial monolingual language policies – described by Chew (1999, p. 40) as indulgence in ‘linguistic nationalism’ – of south-east Asian neighbours such as Malaysia and, more so, Indonesia. The then Deputy Prime Minister Lee Hsien Loong’s views from 2001 are as explicit as they are emblematic of the English language policy in Singapore:

‘It is important, to get Singaporeans to speak good, standard English... We need to plug into the global economy, developing ourselves into an attractive hub for foreign MNCs and talent, and linking up with the rest of the world through trade and investment.’ (DPM Lee Hsien Loong, 2001)

Yet in spite of all this, the state still views English as an essentially *Western* language. Adopting it as a first language or fostering any sense of cultural identification with it would compromise Singapore’s so-called ‘Asianness’, erode citizens’ sense of heritage and unsettle the cultural conception of the nation. Pakir, Lim and Wee’s recent study (2010, p. 6) asserts that, ‘English can of course serve as a lingua franca, but Singaporeans are not generally expected to develop a sense of identification with the language’. This study will test the limits of rationality when it comes to the state’s policy of linguistic pragmatism. And the best way to properly critique its rationality is to assess the most culturally-charged variety of English in use in the city-state, that being Colloquial Singapore English (CSE). It is hoped that analysing the stance of government and speakers respectively, such a balanced critique will be attained.

1.4 The growth of Colloquial Singapore English (CSE)

Perhaps the biggest insurgency against the state ideology outlined above is the rapid development of Colloquial Singapore English. Following the five stages of development outlined by Schneider (2007) in his Dynamic Model of postcolonial Englishes, Singapore English is currently in the fourth stage known as ‘endonormative stabilisation’. During this period, after indigenous residents have begun to take some degree of ownership of English and more innovative language arises, the language begins to stabilise and perhaps comes to signify a mode of identity within the community. This is the point where English in Singapore becomes Singaporean English. Additionally, in the fourth stage, different varieties of the language begin to emerge within the same speech community. This has no doubt occurred in

Singapore since the 1970s (Low & Brown, 2005, p. 30), with the emergence of an informal variety academically known as Colloquial Singapore English (more commonly known as Singlish). This paper focuses solely on CSE, and it is important to make the vital distinction between this informal variety and Standard Singapore English (SSE). Standard Singapore English is virtually indistinguishable – on a lexical and grammatical basis at least – from Standard British or American Englishes, aside from the occasional semantic discrepancy in vocabulary (Deterding, 2007; Leimgruber, 2011). SSE’s only true divergence from Inner Circle Englishes lies in its phonology (Platt, 1975; Forbes, 1993; Deterding, 2007). Singlish, on the other hand, differs heavily and frequently from the standard in its grammar and its lexis⁸. It is known to be spoken by Singaporeans of all ages and is often seen as a reflection of Singapore’s multiracial and multilingual composition (Deterding, 2007, p. 4). Establishing Singlish’s relationship with SSE is a point of considerable debate. Previous studies have assigned it the place of the ‘low’ variety in a diglossic relationship with Standard Singapore English, or as the basilect in a more fluid lectal continuum (Platt, 1975; Gupta, 1989; Leimgruber, 2012). Clarifying this linguistic judgement is vital to assessing the veracity of the state’s more social judgements.

1.5 Government attitudes to Singlish

The Singapore government is, and always has been, opposed to CSE. The ideology underpinning practically every facet of the PAP policy is one of ‘pragmatic viability’ (Chew, 1999, p. 40) or the ‘ideology of “pragmatism”’ (Chua, 1995, p. 68). While this is an understandable line of policy for a young state in a volatile political region and in a precarious economic position, there are limits to its judiciousness. This is because, in reality, ‘pragmatism’ translates into a sole aim: economic prosperity. As Chua Beng Huat points out:

‘[C]ultural elements have been made to serve as the handmaidens to the economic development effort... The primacy of English was rationalised entirely on the basis of its utility for science, technology and commerce, i.e. it is essential to economic development.’ (Chua Beng Huat, 1995, p. 65)

Given what was to follow five years later – with the launch of the Speak Good English Movement in 2000 (see below) – Chua’s comments seem remarkably prescient. As earlier mentioned, the Singapore government believes there is no such thing as a Singaporean native English speaker, in spite of the fact that 32% of citizens speak it as their first language at home. This denial of linguistic and cultural ownership of English to Singaporeans has seen the development of a prescriptive belief in what Stroud and Wee (2012, p. 12) call ‘the ideology of a standard’, and from this the government encouraged an influential ‘Native Speaker/Non-Native Speaker’ dichotomy (Alsagoff et al, 2007). Essentially, Singaporeans must speak (Standard) English, but not own it because ‘Inner Circle speakers of English continue to be regarded as the true owners of English’ (Bokhorst-Heng et al, 2010, p. 138). Furthermore, ‘speakers of the local variety of English, especially the colloquial variety of Singlish, are portrayed as uneducated, uncouth and unworldly’ (Bokhorst-Heng et al, 2010, p. 138).

This notion was borne out most fully in official policy through the birth of the Speak Good English Movement (SGEM) – one of the state’s many national campaigns for social amelioration – on 29 April 2000, and it certainly sparked the sharpest conflict between the ideology of (economic)

⁸ It is not this paper’s intention to provide a full lexical and grammatical description of Singlish. Broadly, Singlish is marked by its adoption of Chinese and Malay syntactical patterns (and its own, Chinese-influenced mode of tense inflection), as well as words and expressions from the various ethnic languages spoken in Singapore, mainly southern Chinese dialects, and Malay. For much more detailed overviews, see Lim (2004) and Deterding (2007).

pragmatism and theories of cultural and linguistic rights. Although ostensibly a public amelioration programme targeting exonormative standards for Singaporeans, it is in reality an aggressive anti-Singlish movement. The SGEM promotes 'good' English and scorns 'bad' English, and emblematises core government attitudes towards CSE. It views Singlish as:

- (1) A hindrance to economic development (for the nation) and mobility (for the individual):

'Let us be gracious hosts to our guests, greet and serve them using clear and grammatical English, and leave them a lasting impression of Singapore, and make them want to come back to Singapore again. If instead we speak in a dialect which only some Singaporeans can understand, then we are handicapping ourselves, and cutting ourselves off from the rest of the world.' (PM Lee Hsien Loong on Singapore's hosting of the annual meetings of the boards of governors of the International Monetary Fund and World Bank, 2005)⁹

'The benefits of good English are much more than you think. It may mean impressing an interview panel and getting a better job with better pay. It may mean convincing a customer who brings you more business. Whatever you are doing, speaking English brings a significant advantage.' (PM Lee Hsien Loong, 2005)

- (2) An indication of, and contributor to, lack of intelligence:

'Poor English reflects badly on us and makes us seem less intelligent or less competent.' (PM Goh Chok Tong, 2000)

- (3) Not a marker of Singaporean identity:

'We are no less Singaporean by speaking good English.' (PM Lee Hsien Loong, 2005)

'It is wrong to think that we are only Singaporean if we speak Singlish. We want to strengthen our common Singaporean identity, but let us do so in other ways, and not by using Singlish.' (DPM Lee Hsien Loong, 2001)

A combination of its vehemently anti-Singlish policies and 'the state's inability to free itself of the native speaker ideology' (Lim, Pakir & Wee, 2010, p. 10) ensures that the SGEM attempts to ignore (or even combat) Scheider's Dynamic Model, as it clings to exonormative standards in its conception of what constitutes Singapore English. According to the government, 'speaking good English has national merit and is connected to notions of good citizenship' (Bokhorst-Heng et al, 2010, p. 218). This paper will test this assumption and the notions of loyalty or duty which accompany it.

2 Methodology

2.1 The survey¹⁰

⁹ All comments from government officials, unless stated otherwise, quoted in this paper can be found on the official Speak Good English Movement website (www.goodenglish.org.sg) under the Media Coverage ('Official Speeches') section. [Accessed 19/01/12]

¹⁰ The survey received ethical approval from Oxford Central University Research Ethics Committee on 31 January 2012.

To answer these questions, a survey was conducted by means of an online questionnaire accessible between 8 February 2012 and 20 February 2012. In creating the questionnaire, I wanted to ask:

- How familiar is Singlish to Singaporeans?
- How is Singlish used and in what social contexts?
- What are Singaporeans' attitudes to Singlish, especially in light of government and so-called elitist views on the variety's usage and impact?

My survey takes the form of 29 multiple-choice questions, each with a view to directly or indirectly answer the above questions. In order to ensure that my survey was suitable for a Singaporean audience, it was checked by 4 native Singaporeans for linguistic inaccuracies (as well as general cultural competency), and my thanks go to them for their time, comments and suggestions.

2.2. The participants

A total of 484 Singaporeans responded to the survey, of which 263 completed it fully. In the interests of consistency this study will only consider results from the 263 fully completed questionnaires. This is because it is difficult to rely on the sincerity of answers from only partially completed surveys, even if it is at the expense of a larger quantity of data. For the purposes of this study it would have been most desirable to obtain a fully representative demographic sample, particularly given the sociolinguistic nature of the enquiry. However, practical considerations were such that this was not entirely possible, meaning that my investigation makes use of a largely exploratory sample which nevertheless did still cover all the official ethnicities, languages, and levels of education.

Speakers' socioeconomic background forms an important part of my investigation into Singlish, yet it was neither plausible nor desirable to ascertain confidential information from participants, such as income, type of residence or occupation. Asking for income can often act as a deterrent for potential participants and would have also led to misleading data from well-educated but technically unemployed current students. The sheer variety of respondents' professional occupations is too difficult to easily assimilate into categories, and therefore to use as accurate socioeconomic markers. As such, both were overlooked in favour of asking for respondents' highest level of education attained, which is used in this paper as the chief socioeconomic index¹¹. Only one participant indicated that primary education was their highest level attained; this was not deemed a sufficiently representative sample, and their responses were therefore discarded.

3 Results and discussion

3.1 Linguistic findings

¹¹ The formal educational qualifications available to Singaporeans are, in an approximate ascending order:

- Primary ('Pri')
- Secondary ('Sec')
- Junior College ('JC')
- Institute of Technical Education certificate ('ITE')
- Polytechnic diploma ('Poly')
- University degree ('Uni')

In (brackets) is the abbreviation used when labelling results figure keys presented in this paper. This is not a linear progression; typically, gaining an ITE certificate or Polytechnic diploma is viewed as an alternative pathway to attending Junior College (roughly equivalent to a UK 'Sixth Form') and university.

3.1.1 Frequency of usage

Familiarity with Singlish was ascertained by question 17 ('Would you ever say you use 'Singlish'?'). 96% of respondents answered 'yes', while 4% said they were not familiar, indicating that Singlish usage is nearly universal in Singapore. This suggests it is inaccurate to confine Singlish speakers to a particular socio-economic bracket, as implied by other H/L diglossic models (see Fishman, 1967). This is confirmed by the lack of any discernible correlation between a Singaporean's level of education and their propensity to use any amount of Singlish, confirmed by the responses to question 18.

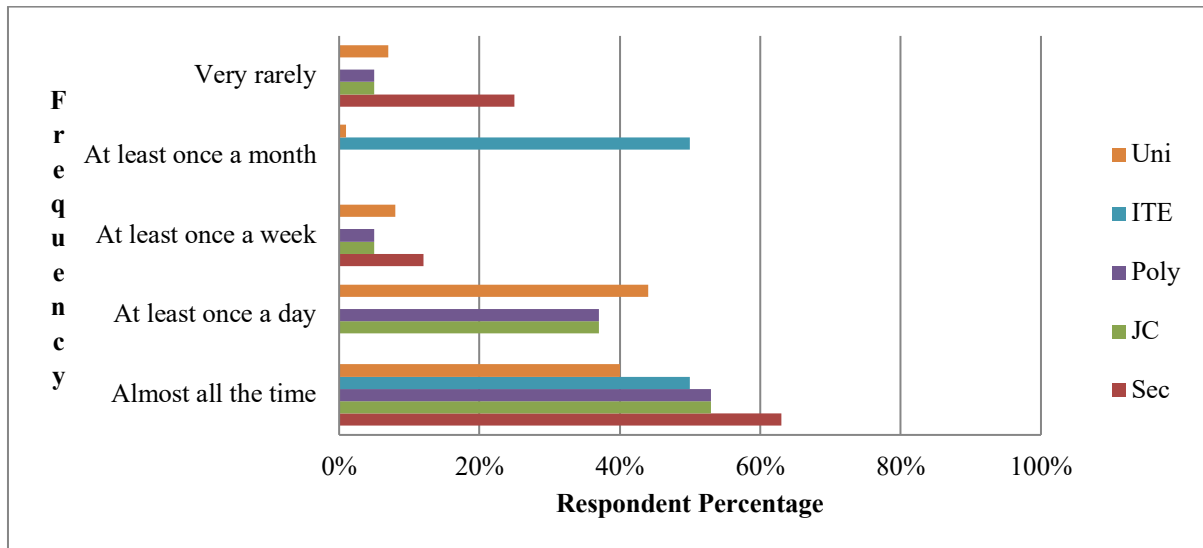


Figure 1: Results for question 18, 'how often do you use Singlish?', cross-tabulated by level of education.

However, as figure 1 shows, a speaker's level of education was indeed relevant in the case of those speakers who use Singlish almost all the time; this does seem to confirm the government link between a citizen exclusively using Singlish and socio-economic opportunity (Stroud & Wee, 2012, pp. 13-4). However, a good 40% of university graduates in the study claimed they used Singlish at least once a day, which suggests that it is not CSE itself which affects educational success and social status, but rather extreme frequency of usage. Judging from these results, then, we see that the state's attempts to wholly dismantle Singlish and its usage (whether large or small), known as the policy of 'creative destruction' (Rubdy, 2001), are neither necessary nor advisable.

These results, then, refute the link between the use of Standard English and economic prosperity so frequently repeated by government officials. Furthermore, it is not possible to ascribe the usage of CSE to race, as indicated by the low rate of variation in Singlish usage in table 1.

Table 1: Results for question 17, 'Would you say you ever use Singlish?', cross-tabulated by ethnic background

Race	Chinese	Malay	Indian	Eurasian	Other
Yes (%)	96	100	89	100	93
No (%)	4	0	11	0	7

Having established that CSE is a language common to almost all Singaporeans, question 18 tells us how 82% of participants said they spoke Singlish at least once a day, with 38% of these speakers using it ‘almost all the time’. It is clear then that government efforts to reduce Singlish usage, channelled from 2000 through the Speak Good English Movement, have largely failed.

When asked about the nature of their language use in informal conversation (question 19), only a minority of respondents said they would use Singlish for the entirety of a conversation (12%), while the rest (88%) use Singlish only for the occasional sentence or word. The government conception of Singlish holds that the so-called L and H variety do not mix in the speech of Singaporeans, but are rather used discretely for separate purposes (along the lines of, for example, Hungarian and German, evidenced by Susan Gal’s 1979 study). The state thus assumes that Singlish is a fully extensive social language (Stroud & Wee, 2012, p. 13) used as a medium for conducting entire exchanges. This, in turn, is buttressed by the assumption that Singaporeans communicate in CSE because they are unable to master the standard, leading to the state’s negative conception of Singlish as a marker of poor education, and thus as a national problem. As we can see, this is not the case for CSE, as speakers overwhelmingly use only snippets of the so-called ‘L’ variety in their everyday speech. It is due to the responses to question 19, then, that I have opted for the verb ‘use’ rather than ‘speak’ in relation to Singlish. So the results yielded from question 19 suggest that the socioeconomically-induced diglossia (in the traditional sense) so feared by the government does not in fact exist. Further confirmation is provided by a cross-tabulation of results by level of education in figure 2:

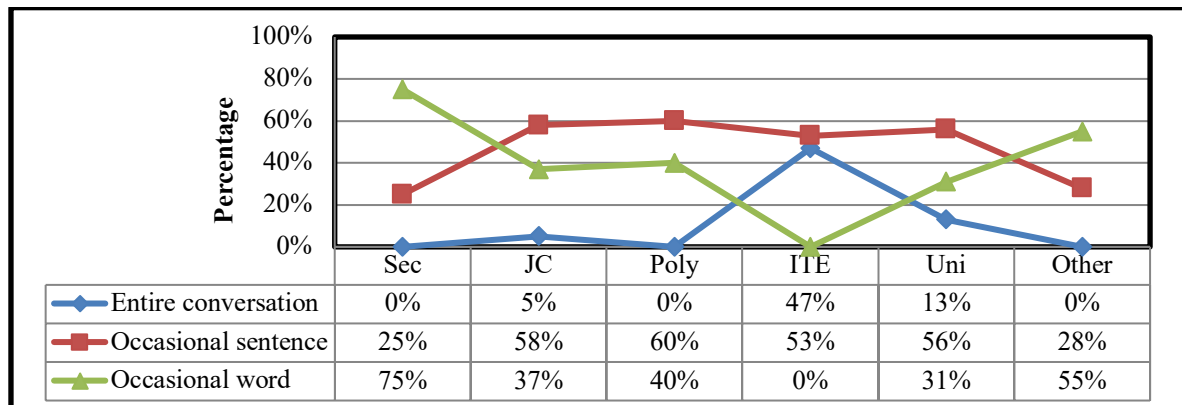


Figure 2: Responses to question 19, ‘When you use Singlish do you use it for...?’, cross-tabulated by education.

The government’s diglossic model is not supported by the fact that more speakers with a university degree use Singlish more than those with a polytechnic diploma, those with Junior College qualifications, or those who went no further than secondary school (the majority of whom, in fact, only use the occasional word). If, as indicated above by the figure above, speakers who are less educated are using Singlish only for the occasional word or sentence, the government must more seriously consider the possibility of Singlish being used as ‘a play language...perhaps one that speakers occasionally and deliberately cross into’ (Rampton, 1995, quoted in Stroud & Wee, 2012, p. 13).

3.1.2 Proficiency

Interestingly, however, when participants were asked to judge fellow speakers’ linguistic habits, the results supported the government stance much more strongly. The effectiveness of government campaigns and elements of social snobbery could well have had parts to play. Central to the

government's anti-Singlish campaign is the fear that the very presence of CSE will result in its adoption as a first language by citizens, to the great detriment of Standard English. The worst case scenario, from the government's point of view, is that Singaporeans will grow up with Singlish as their first language. And indeed, 83% of respondents thought some Singaporeans are more capable of using Standard English than others, which suggests agreement with the government's preoccupation with falling standards and broken English (see PM Goh's comments from 2000 above). Furthermore, when asked whether they thought some Singaporeans are only ever capable of using Singlish (question 25), participants agreed with the government's class-based assumptions over proficiency, and by definition over the existence of a form of diglossia in Singapore:

Table 2: *Results for question 25, 'Aside from their mother tongue, do you think some Singaporeans are only ever capable of using Singlish?'*

Yes (%)	57
No (%)	32
Not sure (%)	12

Question 28 asked whether participants used the discourse particle 'lah' (e.g. 'Thailand is very beautiful, lah') or the syntactic reduction can (e.g. 'Do you accept \$100 notes?' 'Can') even when they do not mean to do so. The fact that 73% admitted they did only strengthens the findings from question 25 above. If speakers are using markers (or, in this case, stereotypes) of Singlish unintentionally, some (i.e., the government) would argue that Singlish is damaging the proficiency, or at least the acceptable native speaker norms, of the standard variety: 'English corrupted by Singaporeans'. In fact, when we look at those who responded to this question who did not go to university, 85% acknowledged that they unintentionally used 'lah' or 'can', which is 12% above the sample average. The official link between being a socioeconomic 'loser' and using what SGEM called in 2011 'English as it is Broken' seems, then, to be confirmed. This is another explanation, however. The idea behind this link is that the proliferation of Singlish harms Singaporeans' proficiency in English, leading to this curious diglossia comprised of SE speakers and non-SE speakers. However, I would argue that the results to question 28 in fact dispel the government's fear of a CSE/SE diglossia because use of an L or an H variety implies conscious choice on the part of the speaker to speak that variety, and therefore refusal of the other on the basis of class and education. It is more plausible that the inadvertent use of 'lah' and 'can', rather than being an indication of lack of proficiency in Standard English, is a sign of potential endonormative stabilisation of Standard Singapore English. This cannot be deemed detrimental to intelligibility or socioeconomic prospects. The fact that 71% of university graduates said they involuntarily used 'lah' and 'can' (lower than the sample average, but negligibly so) is strong evidence that government and sometimes popular agitation over Singlish's effect on proficiency and economic prosperity are exaggerated, to say the least.

Returning to question 25, where the majority of participants agreed with the government that some speakers can only speak Singlish, it is interesting to note that, on closer inspection, it was in fact only those who spoke English and Chinese as their first language at home that did indeed agree as a majority.

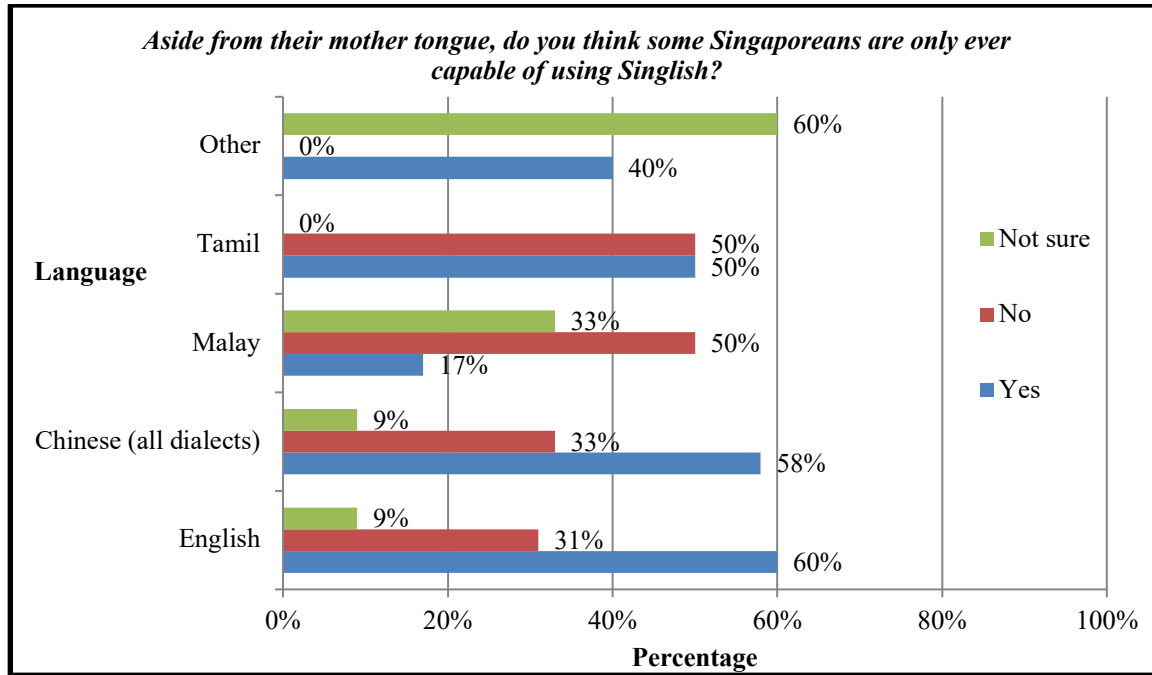


Figure 3: Results for question 25, ‘Aside from their mother tongue, do you think some Singaporeans are only ever capable of using Singlish?’, cross-tabulated by language spoken most frequently at home.

Participants were certainly more inclined to agree with the government line when asked to judge the speaking habits of others, rather than themselves. There may well be an element of socioeconomic snobbery in this; English and Chinese L1 speakers are not only numerically more prevalent in Singapore, they are also more advantaged educationally and financially. Furthermore, Malay and Tamil speakers are required to speak in a language other than their MT much more than their anglo- and sinophone fellow citizens due to the dominance of English and Chinese on the island. It follows, then, that non-English and non-Chinese speakers would be judged unfavourably in terms of English proficiency. Yet the fact that Malay speakers (17%) and Tamil speakers (50%) themselves would not agree so emphatically with Chinese (58%) and English speakers (60%), suggests more about elitist government viewpoints being readily absorbed by certain citizens than it does about a sociolinguistic reality.

3.1.2 Context of usage

Questions 11-16 served to clarify the CSE’s relationship with the standard with a degree of finality. Building on Platt’s (1975) theory of Singlish as ‘creoloid’ existing as the basilect in a speech continuum, the questions put the participants in 2 different scenarios and offered them 3 ways (loosely, an acrolectal form, a mesolectal form and a basilectal form) of saying the same thing. Participants were then asked to indicate which of the phrases they would use when speaking to different people in various situations of formality. The average results taken from the two scenarios are shown below:

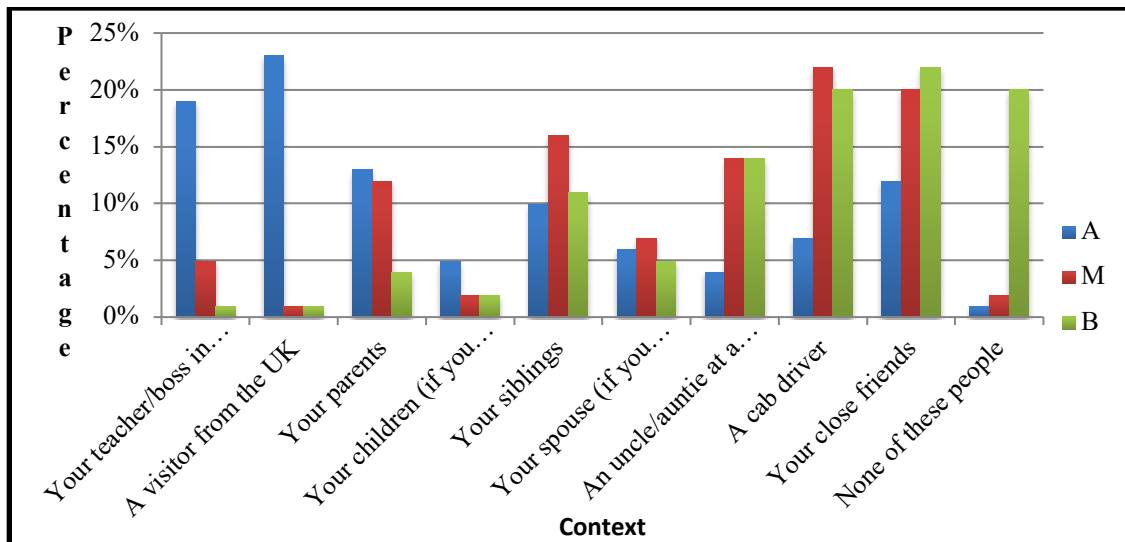


Figure 4: Results for questions 11-16, asking in which social situations participants would use a hypothetical acrolect, mesolect and basilect of Singapore English.

The social settings are positioned in order of approximate formality along the x-axis of the chart and as can be seen, the use of the Standard Singapore English acrolect drops as the situation becomes less formal and conversely the use of the basilect noticeably increases as the context becomes less formal. As expected, use of the mesolect remains fairly consistent, especially in situations of medium formality. These results offer strong evidence in favour of the notion of CSE existing in a lectal continuum of Singapore English. However, Platt's (1975, p. 369) model suggests that while there is a speech continuum in Singapore English, not all of it is available to all speakers due to economic factors. See figure 5, taken from Platt (1975).

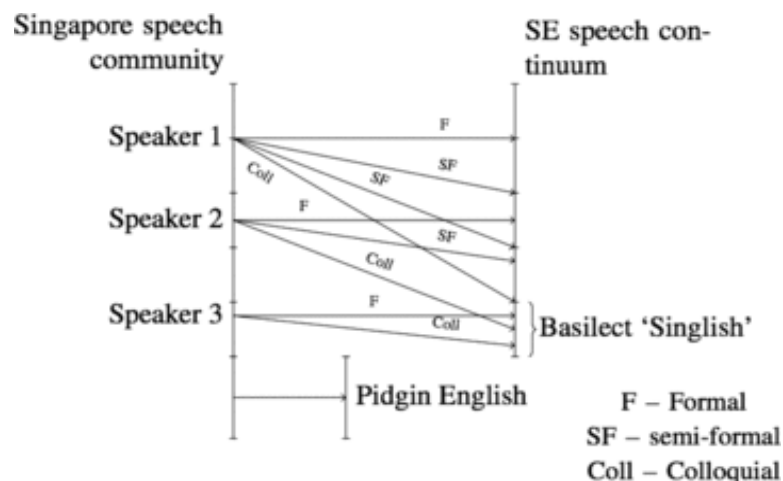


Figure 5: 'Relation between socio-economic factors and the usage of sub-varieties of [SgE] available to a speaker'.

In this model, speaker 1 belongs to the highest socioeconomic class and speaker 3 to the lowest. Yet this model, whilst denying the ever more implausible diglossic conception of Singapore English, implies discrete categories and that some Singaporeans cannot speak the acrolect, which supports the government view. An alternative and more recent model, proposed by Leimgruber (2012), endorses an

‘indexical approach’ to describing Singapore English, whereby each grammatical, lexical and phonological variable of Singlish is indexically subject to the speaker, the listener and the social context of the situation. The main consequence of this approach is the deconstruction of a definable acrolect, mesolect and basilect for Singapore English, as the emphasis is shifted to individual variables. This supports the findings from question 19 where many Singaporeans claimed they only used the occasional Singlish word or sentence, and also supports the possibility for Singaporeans to ‘code-switch’ between varieties, even at an intra-sentence level. The new model and my own findings suggest that government fears that a generation of Singaporean children ‘would grow up with Singlish as their first language’ (DPM Hsien Loong, 2001) are unfounded. In this case, the term ‘using’ Singlish, I would argue, seems a more appropriate one than ‘speaking’ Singlish, which implies exclusive usage of one variety and preoccupies the government. Adopting this model means, again, the government view does not tally completely with the sociolinguistic reality.

3.2 Social findings

3.2.1 *Personal attributes*

As mentioned earlier, the ‘ideology of the standard’ behind the government’s language policy moves quite seamlessly from a judgement on Singlish to a judgement on Singlish speakers, often pejoratively implying certain moral or personal traits. In fact, Paul Bruthiaux even goes as far as saying that ‘the main concern of SGEM is not intelligibility but respectability’ (2010, p. 96). Prominent among these personal traits associated with Singlish is the notion of a lack of intelligence and incompetence from its speakers as suggested by PM Goh in 2000 (referenced above). Working under the principle that ‘most language behaviours are in fact socially diagnostic’ (Coupland & Giles, 1991, p. 32), questions 20-22 made use of the ‘direct approach’ (Garrett, 2010, p. 37) in surveying attitudes to micro-sociolinguistic features in order to ascertain broader macro-sociolinguistic trends (Wardhaugh, 1992, p. 12). Participants were asked to imagine a person using Singlish compared to one using Standard English and evaluate how they would appear in terms of intelligence, friendliness and formality.

Table 3: *Results to question 20, ‘Compared to a person using Standard English, a person using Singlish appears...’*

More intelligent (%)	1
Less intelligent (%)	48
No effect (%)	51

There is no doubt that the respondents felt that using Singlish does not increase one’s semblance of intelligence, which is understandable given the informality of the variety (see results to question 21). Yet the fact that over half stated that the use of CSE was not an indicator of intelligence at all is quite remarkable, especially given the precedent in similar sociolinguistic studies of personal attributes associated with regional, socio-economically inflected varieties of English, which often reveal heavy stigmatisation (see, for example, Coupland and Bishop’s (2007) matched-guise test investigating perceptions of regional British accents). Singaporeans appear to not be in concordance with PM Goh’s view, then. Participants displayed a similarly ambivalent attitude when it came to friendliness (question 22), with 58% stating that a Singaporean using Singlish would appear friendlier than one using Standard English and 40% stating it would not make a difference to how friendly they appeared. Merely 2%,

however, believed using Singlish would make the person seem less friendly. It was only when assessing the formality of a person using CSE that respondents gave decisive judgement:

Table 4: *Results to question 20, ‘compared to a person using Standard English, a person using Singlish appears...’*

More formal (%)	2
Less formal (%)	94
No effect (%)	4

The results reveal both the cultural and (economically) pragmatic values of Singaporeans and the government. Figure 6 is a good illustration of this clash of values:

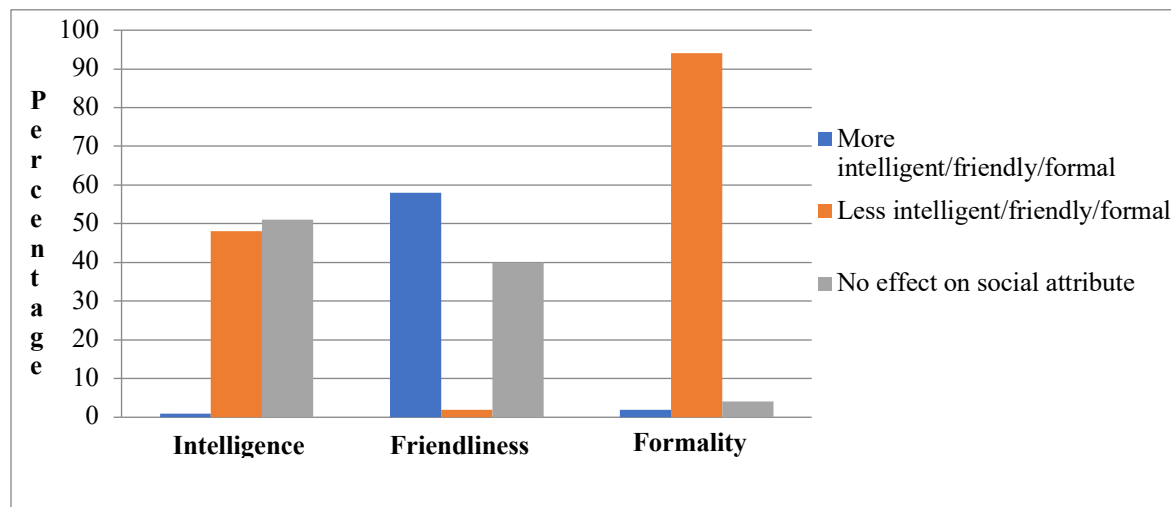


Figure 6: *Summary of responses to questions 20-22, asking for the social attributes (intelligence, formality and friendliness) of a person using Singlish compared to using Standard English.*

In the eyes of the government the positive comparatives here would be ‘more intelligent’ and ‘more formal’, as they are the best aligned with its political philosophy and economic aims. And the low percentages shown in figure 6 show that participants share the view that CSE is certainly not a suitable variety to achieve these aims compared to SE, especially in relation to ‘formality’ which is arguably the quality most rooted in pragmatism and the world of commercial transaction. It is telling, though, that, to borrow Chua’s phrase, the only ‘cultural element’, ‘friendliness’, is indeed best served by Singlish. The warm response to the link between Singlish and being perceived as friendly – coupled with participants’ reluctance to condemn speakers’ intelligence on the basis of Singlish usage (in stark contrast to state rhetoric) – highlights once again the government’s foregrounding of industrial pragmatism over more dispensable notions of cultural and social well-being. It thus displays an attitude at quite significant odds with contemporary social attitudes, and consequently, sociolinguistic reality.

3.2.2 Direct attitudes towards government policy

Question 26 provided the most direct response to speaker perceptions of Singlish in light of government policy. SGEM and other initiatives, as shown by the results in table 5, have clearly failed to work.

Table 5: Results to question 26, ‘How far do you agree with Prime Minister Goh Chok Tong’s claim that, “if young people speak Singlish when they can speak good English, they are doing a disservice to Singapore”’.

Strongly agree (%)	6
Agree (%)	13
Neither agree nor disagree (%)	19
Disagree (%)	39
Strongly disagree (%)	23

Indeed, neither accusations of socioeconomic elitism levelled at policy makers, nor claims from government officials such as Col. David Wong, former Chairman of the Speak Good English Movement in 2003, that ‘this push for Singaporeans to speak good English is a people-driven movement and not a top-down campaign’, are vindicated by a breakdown of results by education.

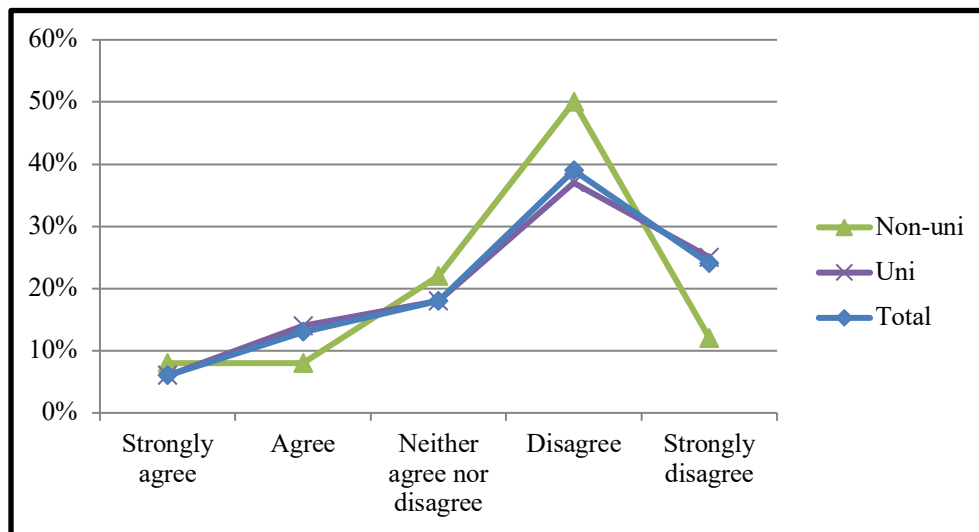


Figure 7: Results to question 26 cross-tabulated by level of education.

Non-degree holders are more likely to disagree with Goh more than university graduates, yet those with a degree are more likely to strongly disagree. Therefore regardless of class or education, these results show Singaporeans united in opposition against the anti-Singlish movement.

4 Conclusion

The study has shown that Singaporeans, contrary to government ideology, do feel ownership of Singlish, and generally do not support efforts to reduce or destroy its usage. The results also revealed a significant ideological gap between speakers and policy makers in their conception of Singlish as it relates to SE or SSE, in both linguistic and social terms. It is not a mere isolated ‘low variety’ nor is it the harbinger of moral and financial decline. When Lui Tuk Yew, former Minister for Education, insisted in 2006 that Singaporeans needed to improve Standard English language skills in order to build ‘emotional literacy to create the social glue that holds us together as a people and as a nation’, he badly missed the point. He said that ‘the language of the heart, of feelings and emotions is the key to

connecting with our parents, children, colleagues, our neighbours and especially, our loved ones'. Unfortunately for him, this language is Singlish.

The limits of practicality of Singapore's language policy have been directly and indirectly exposed by a sample of a nation whose polity is marked by a 'mixture of coercion and consent' (Francesch-Huidobro, 2008, p. 4). Indeed, Singapore's history of censorship, tight grip on the press and documented restrictions on freedom of speech make the fact that Singaporeans continue to use Singlish and view it, as Hoon (2003, p. 52) puts it, as 'a form of English... that most Singaporeans count without hesitation as part of their identity', a testament to its cultural significance and to the government's severely skewed vision of the sociolinguistic reality of English in Singapore. In 2005, Lee Hsien Loong admitted that losing some of Singapore's 'values, ancient cultural heritages and a sense of identity' in order to solely speak standard English 'was a painful, but rational trade off', given the commercial benefits. He asked, 'Do we want to start developing a new language all to ourselves?' The answer, as this study shows, is yes.

References

- Alsagoff, L., Bokhorst-Heng, W. D., McKay, S. L., & Rubdy R. (2007). English Language Ownership among Singaporean Malays: Going beyond the NS/NNS dichotomy. *World Englishes*, 26(4), 424-45.
- Alsagoff, L. (2010). English in Singapore: Culture, Capital and Identity in Linguistic variation. *World Englishes*. 29(3), 336-48.
- Bao, Z., & Hong H. Q. (2006). Diglossia and register Variation in Singapore English. *World Englishes*. 25(1), 105-14.
- Bishop, H., & Coupland, N. (2007). Ideologised values for British Accents'. *Journal of Sociolinguistics*, 11(1), 74-93.
- Bokhorst-Heng W. D., Rubdy R., McKay S. L., & Alsagoff L. (2010). Whose English? Language Ownership in Singapore's English Language Debates. In L. Lim, A. Pakir & L. Wee (Eds.), *English in Singapore: Modernity and Management* (pp. 133-58). Hong Kong: Hong Kong University Press.
- Brown, A., & Low E. L. (2005). *English in Singapore: An Introduction*. Singapore: McGraw-Hill.
- Bruthiaux, P. (2010). The Speak Good English Movement: A Web User's Perspective. In L. Lim, A. Pakir & L. Wee (Eds.), *English in Singapore: Modernity and Management* (pp. 91-108). Hong Kong: Hong Kong University Press.
- Chan H. C., Quah, J. S.T., & Seah, C. M. (1985). *Government and Politics of Singapore*. Singapore: Oxford University Press.
- Chew, P. G. L. (1999). Linguistic Imperialism, Globalism and the English Language. *AILA Review*, 13, 37-47.
- Chua, B. H. (1995). *Communitarian Ideology and Democracy in Singapore*. London: Routledge.
- Crystal, D. (1997). *English as a Global Language*. Cambridge: Cambridge University Press.
- DeCamp, D. (1971). Towards a Generative Analysis of a Post-Creole Speech Continuum. In D. Hymes (Ed), *Pidginisation and Creolisation of Languages* (pp. 349-71). Cambridge: Cambridge University Press.
- Deterding, D. (2007). *Singapore English*. Edinburgh: Edinburgh University Press.
- Ferguson, C. A. (1959). Diglossia. *Word*, 15, 325-340.
- Fishman, J. A. (1967). Bilingualism with and without Diglossia; Diglossia with and without Bilingualism. *Journal of Social Issues*, 23(2), 29-38.
- Forbes, D. (1993). Singlish. *English Today*, 9(2), 18-21.
- Francesch-Huidobro, M. (2008). *Government, Politics and the Environment: a Singapore Study*. Singapore: Institute of Southeast Asian Studies.
- Gal, S. (1979). *Language Shift: Social Determinants of Language Change in Bilingual Austria*. New York: Academic Press.

- Garrett, P. (2010). *Attitudes to Language*. Cambridge: Cambridge University Press.
- Giles, H., & Coupland, N. (1991). *Language: Contexts and Consequences*. Milton Keynes: Open University Press.
- Gupta, A. F. (1989). Singapore Colloquial English and Standard English. *Singapore Journal of Education*, 10(2), 33-9.
- Hoon, C. H. (2003). 'You see me no up': Is Singlish a problem? *Language Problems and Language Planning*, 27(1), 45-62.
- Jenkins, J. (2003). *World Englishes*. London: Routledge.
- Kachru, B. B. (1985a). Standards, Codification and Sociolinguistic Realism: English Language in the Outer Circle'. In R. Quirk & H. G. Widdowson (Eds.), *English in the World: Teaching and Learning the Language and Literatures* (pp. 11-30). Cambridge: Cambridge University Press.
- Leimgruber, J. R. E. (2011). Singapore English. *Language and Linguistics Compass*, 5(1), 47-62.
- Leimgruber, J. R. E. (2012). Singapore English: an indexical approach. *World Englishes*, 31(1), 1-14.
- Lim, L., Pakir, A., & Wee, L. (2010). *English in Singapore: Modernity and Management*. Hong Kong: Hong Kong University Press.
- Pakir, A. (1991). The range and depth of English-knowing bilinguals in Singapore. *World Englishes*, 10(2), 167-79.
- Platt, J. T. (1975). The Singapore English Speech Continuum and its Basilect 'Singlish' as a 'creoloid'. *Anthropological Linguistics*, 17(7), 363-74.
- Platt, J. T. & Weber, H. (1983). *Singapore and Malaysia*. Amsterdam, PA: Benjamins.
- Rubdy, R. (2001). Creative destruction: Singapore's Speak Good English Movement. *World Englishes*, 20(3), 341-55.
- Schneider, E. (2007). *Post-colonial English: Varieties around the World*. Cambridge: Cambridge University Press.
- Spolsky, B. (2004). *Language policy*. Cambridge: Cambridge University Press.
- Stroud, C., & Wee, L. (2012). *Style, Identity and Literacy: English in Singapore*. Bristol: Multilingual Matters.
- Tickoo, M. L. (1996). Fifty years of English in Singapore: All Gains, (a) Few Losses? In J.A. Fishman, A.W. Conrad & A. Rubal-Lopez (Eds.), *Post-Imperial English: Status Change in Former British and American Colonies* (pp. 1940-1990). New York: Mouton de Gruyter.
- Wardhaugh, R. (2010). *An Introduction to Sociolinguistics*. Oxford: Blackwell. (Original work published 1992).
- Wee, L. (2011). *Language Without Rights*. Oxford: Oxford University Press.

About the Author

Nicholas Marsh, University of Oxford.
nickmarsh1878@gmail.com

The Grapheme <V> in Classical Latin: Did it Represent Two Phonemes or Three?

Devon Smith

University of Cambridge

Abstract. In Classical Latin, the grapheme <V> could be used to represent [u:], [u], and [w]. There are clear minimal pairs showing that [u:] and [u] were separate phonemes /u:/ and /u/. However, there has been much debate about whether [u] contrasted with [w]. In one sense, it seems clear that [w] was a consonant phoneme that contrasted with other consonants: *pestis* /pestis/ ~ *festis* /festis/ ~ *uestis* /westis/ while [u] was a vowel phoneme that contrasted with other vowels: *malum* ~ *malam*. However, the two sounds are phonetically similar enough to qualify as allophones in complementary distribution – [w] occurring in syllable peripheries, and [u] occurring in syllable nuclei. I will focus on the complication that arises when we observe [w] and [u] in C_V or V_C environments. There are four elements that I will bring into the discussion of the phonemic status of these sounds to conclude that [w] was an allophone of /u/. First, I will discuss conditioning elements found in C_V and V_C environments. Interesting patterns emerge where u follows /l/, /r/, /s/, and /n/. Secondly, I will discuss the evidence provided by poetry. In many cases, poetry seems to contradict the normal rules for the distribution of [u] and [w]. Thirdly, I will discuss the Claudian inversed digamma, which was used to represent [w], and what the inscriptional evidence can tell us about speakers' perceptions of sounds. Finally, I will briefly discuss the development of [w] into the Romance languages. This discussion hopes to provide a more nuanced understanding of the dissimilarities between popular Latin and Classical Latin and demonstrates the application of sociolinguistics to historical linguistics.

Keywords: historical linguistics; sociolinguistics; Latin

1 Introduction

The Latin phoneme /u/¹² has been said by some to contrast with a separate phoneme /w/ (Watbled (2005) and McCullagh (2011)), while others have stated that /u/ has an allophone [w] which has no phonemic status in Latin (Ballester (1996b.), Touratier (2005b)). It should be noted that [w] passes the substitution test for determining the phonemic status of a sound: *pestis* ~ *festis* ~ *uestis*. On the surface, it seems as though the phonemic status of the sounds represented by the grapheme <V> should be clear: there is one version, a consonant, which contrasts with other consonants (here, /p/ and /f/), and another version, a vowel, which contrasts with other vowels (*malum* ~ *malam*). However, from another angle, they are two variants of the same sound – one occurring at syllable peripheries, and the other occurring as a nucleus. Below I shall present the arguments given by both sides and attempt to reconcile the evidence they provide.

2 Allophonic [w]

¹² Though this paper discusses the grapheme <V>, the symbol I will be using throughout to represent this grapheme is u. There are two strong reasons for this: First, if I were to use <V> throughout, I would only be able to discuss the grapheme itself and any commentary on its pronunciation would have to be carried out in square brackets – requiring me to comment on the precise nature of any other sounds, however controversial, which happen to occur in the same word. Second, the use of u allows me to comment on whether a specific token of u was consonantal or syllabic (through the use of and respectively); it would be inappropriate to apply these phonetic symbols to a grapheme. As an aside to this last point, it should be noted that stating the pronunciation of u in this way is not controversial – these pronunciations have been well established by meter, contemporary commentary, and basic phonological principles.

As stated above, it is reasonable to perceive *u* as [w] in syllable peripheries and [u] in syllable nuclei. If this were the case everywhere, these sounds share enough phonological features to be considered allophones of the same phoneme on the basis of their complementary distribution.¹³ What is interesting is what happens to *u* when it borders both a consonant and a vowel; here we find some tokens where it appears that *u* shows contrast between its consonantal and vocalic realisations.

The discussion of the phonological status of the sound *u* in Latin has recently been found in Xavier Ballester's 'Fonemática de /u/ en latín. El argumento de los pares mínimos' (1996b.). He argues that [w] is simply an allophone of /u/ "...hasta época imperial al menos..." (137). Though he does not expand on his reason for assigning the advent of [w] as a phoneme until after the imperial period, we can imagine that this may be based on the development of the sound in the Romance languages (see section 6 below). For now, let us focus on his argument that [w] is an allophone of /u/ and not a phoneme in its own right, which he bases on the lack of an acceptable minimal pair.

Ballester is of the view that 'minimal pairs' like *salūi* : *salūi* are not acceptable because they do not show contrasts in the same context. In the first word, [w] is part of the lexical morpheme and occurs before a grammatical morpheme /i/, whereas the [u] it is being contrasted with is a morpheme in its own right – a grammatical morpheme added on to a lexical morpheme *sal* and followed a grammatical morpheme /i/: *salū+i* : *sal+ū+i* (137f). The sounds [w] and [u], he argues, are in complementary distribution on a morphophonological level. That is, with reference to the example above, the phoneme /u/ is realised as [u] when it is part of the morpheme [PERFECT TENSE], but not when it is part of the lexical morpheme (i.e. the root of the word).¹⁴

2.1 Behaviour of /u/ Following /l/ and /r/

Where there is a morpheme boundary following *lu* or *ru*, *u* is realised as [w] (Godel, 95). The clusters *lu* and *ru* are not permitted onsets or codas and must be analysed [l.w] and [r.w]. Examples are as follows:

- (1) (a) *silūa*: sil.ū+a
- (b) *alūo*: al.ū+o
- (c) *parūus*: par.ū+us
- (d) *arūa*: ar.ū+a

Conversely, the following do not work:

- (2) (a) *colūmbus*: co.lum.b+us – *u* follows /l/ in the lexical morpheme, but *u* must be the syllable nucleus since /m/ is not a permitted nucleus in Classical Latin and [lw] is not a permitted syllable onset.
- (b) *hellūatus*: hel.lū.+a.t+us – the analysis *hell.ū+a.t+us is not permitted because it violates the Onset First Principle of phonology.
- (c) *elūentur*: e.+lū.+e+n.tur – *e+l.ū+e+n.tur is not a permitted analysis because though /l/ can be a syllable coda, the syllabification of the root *lūo* (lū.+o) is maintained; *u* must be syllabic in this word for, as stated above, [lw] is not a permitted syllable onset, nor is the other possibility (*l.ū+o) acceptable

¹³ On the need for phonetic similarity between two sounds in complementary distribution to justify them as allophonic realisations of the same phoneme, see Clark and Fletcher (2007), Carr (1993), Hock (1991), Jones (1950), etc.

¹⁴ He bases his distribution on Robert Godel's 'Les semi-voyelles en latin', discussed below.

- because /l/ is not an acceptable nucleus in Classical Latin.
(d) *rŭo*: rŭ.+o – for the same reasons as given for *lŭo*.

Thus *uolŭit* ‘it rotates’, which contains a *u* in its stem, is realised [wol.w+it], while *uolŭit* ‘he wished’, where the *u* is morphological, is realised [wo.l+u.+it]. We can state this rule as follows:

- (3) /u/ → [w] / /l, r/ . _ +

2.2 Behaviour of /u/ Following /s/ and /n/

Another complication is found when *u* forms a syllable onset cluster with <s>. Here it is realised as [w] unless a morpheme boundary immediately follows: thus *suadeo* (*sŭa.d+e.+o*) versus *sŭile* (*sŭ.+il.e*), *sŭa* (*sŭ.+a*) (Godel, 94). Of course, these clusters are subject to rule I.b.i, stating that it must be syllabic if it happens to fall between consonants.

It seems odd that [sw] clusters should be exempt from the general rule in Classical Latin that [w] does not occur anywhere except as a syllable onset. Watbled dismisses the unusualness of this cluster in Classical Latin phonology as phonetic property of [s] i.e. that “Le sifflante [s] a traditionnellement, dans de nombreuses langues un statut special do point de vue syllabique: elle peut en quelque sorte être adjointe à différents types d’attaques consonantiques, en ne respectant pas l’échelle de sonorité (2005, 44)” citing G. N. Clements’ research on “The role of the sonority cycle in core syllabification.” While this could very well be the case, it seems unsatisfactory and does not attempt to account for the poetic variation we find (more on this below). However, the suggestion by Devine and Stephens (1977) that there was a consonant /s^w/ also meets with some problems.

First and foremost, if there is a phoneme /s^w/, it is in very limited distribution – Devine and Stephens describe its distribution as follows:

‘The graphemic cluster *su* represents [sw] rather than [su] when it is followed by a homosyllabic *ā* or *ē*: this only happens in post junctural position, and, with a couple of rare exceptions, is confined to the words *suadeo*, *suavis* and etymologically or popularly related *s(u)avium*, and *suesco*...’ (Devine & Stephens, 1977, p. 78f)

The phoneme /s^w/, they maintain, is a common enough phoneme cross-linguistically, and would fit in nicely with the labialised consonants /k^w/ and /g^w/, which they argue are present in Latin (as opposed to the clusters [kw] and [gw], which will be discussed in 4.1 below). Poetry seems to suggest that *su*, where pronounced [sw], was monophonemic, but the evidence is not solid; even though there is a tendency to treat [sw] monophonemically, there is almost equally a tendency not to do so (Devine & Stephens, 1977, 79-80).¹⁵ With this ambiguous evidence, they conclude as follows:

‘...syllabification and system congruity...point to /sw/, by-forms in *su-* suggest /sw/. In Indo-European, of course *sw was biphonemic and supported by similar clusters: the question is whether sw came to be monophonemic in upper class Latin with the elimination of non-syllable-initial w; after this development it was supported not by sister clusters, as in Indo-European, but by the sister phonemes qu, gu.’ (Devine & Stephens, 1977)

With this evidence, it is difficult to conclude one way or the other. For now, it will be easiest to assume

¹⁵ Devine and Stevens give a reasonably comprehensive account of variations in the scanning of *su*.

a special case for /s/ and [w]. Even if further research does show that this pattern is suggestive of a phoneme /s^w/, we are not harmed by stating that we can predict a [w] in the circumstances given above. The problem of poetry will be discussed in 4.1 below.

We see a similar pattern where *u* follows /n/. The example we will focus on is *tenuis*. Though it seems the usual pronunciation of this term in Classical Latin was [tenūis],¹⁶ it does have a rather high frequency of scanning as [ten̥is] in poetry. This might suggest that we should take this consonant into consideration in our list of rules for the Classical Latin distribution of [w]. We do note that in *tenuis* a morpheme boundary intervenes between the *u* and the *i*, suggesting a rule like those given for /l/ and /r/ above. Godel hints that this might be a good solution:

‘La synizèse, si fréquente, dans *tenuis* (et surtout *tenuia*, *tenuior*) a chance de représenter la prononciation normale : c’est le seul adjectif où -nu- (-nw-) précédé d’une brève appartienne au radical : on a *tenuior* comme *suauior*, *saeuior*, *antiquior*, tandis que dans *ingenuus*, *continuus*, au contraire, il y a une limite de monème entre *n* et *u*, comme dans les parfaits *genui*, *tenui* ; et là il n’y a jamais de synizèse. On peut donc ajouter aux exemples donnés ci dessus : *tenu|is*, *tenu|ior* (-w-) et *ingen|uus*, *gen|ui* (-u-) et dès lors, *genua*, disyllabe (Virg., En. V 432) indiquerait une analyse *genu|a*.’ (Godel, 1953, p. 95)

However, if we posit a rule such as that, we will find ourselves faced with the opposite problem when we find *tenūis* scanning as trisyllabic. We will return to this discussion in section 3.1 below.

2.3 Complementary Distribution of [w] and [u]

Based on these discussions, the distribution of the realisations of *u* as suggested through comparison of our sources is as follows:

- (4) *u* →
- (a) [w]
 - (i) / V _ V (*auis*)
 - (ii) / /l, r/ . _ + (*silua*)
 - (iii) / . + /s/ _ V (*suauis*)
 - (iv) (?) / /n/ . _ + (*tenuis*)
 - (b) [u]
 - (i) / . + C₀ _ C₀ . + (*unus*, *cornu*)
 - (ii) / /l, r/ + _ . (*merui*)
 - (iii) / . + /s/ _ + (*sua*)

These rules show that the presence of one or the other phones is completely predictable and therefore that they are allophones of the same phoneme.

3 An Objection to this Analysis

Jean-Phillipe Watbled (2005) believes that the analysis of [w] as an allophone of /u/ is problematic once the labiovelars enter the scene. The labiovelar [k^w] presents no difficulty as it is clearly recognised

¹⁶ Devine & Stephens state that “Stop + w is an inadmissible combination (except when divided by compound boundary) in the classical language...” (1977, 60)

as a single complex phoneme. In meter, it scans as a monomoraic segment, so it cannot be /kw/. Furthermore, it would be typologically odd if Latin were to permit [kw-] as a syllable onset, but not *[tw-], *[dw-]¹⁷, *[pw-], or *[bw-]. The phoneme /k^w/ regularly appears in a variety of environments. However, Watbled claims that the segment [g^w] only ever appears following /n/ and before a vowel (Watbled, 2005, 45).¹⁸ This means that the distribution of [g^w], if it is phonemic, is incredibly limited. However, I do not see why a limited distribution of a sound should make it implausible as a phoneme. Devine and Stephens (p. 90) agree, citing the distribution of // in English.

Watbled establishes three possible means of treating *gu* in Latin (p. 45-49). His first is to treat *gu* as a labiovelar phoneme /g^w/ with the minimal pairs /ling^wam/ (*linguam*) and /link^wam/ (*linquam*). The problem for Watbled is that /g^w/ can only appear after /n/. However, surely this is more satisfactory than his other suggestions: that [gw] and [w] are in complementary distribution or that *gu* represents two phonemes, /g/ and /w/. The first of these two assessments requires that /w/ contrasts with /k^w/. Even if we accept that /k^w/ and /w/ are phonetically similar enough to contrast with each other (both are [+labial], [+velar], and [+consonantal]), it seems odd to prefer a contrast between [-continuant] and [+continuant] in a phonological system that does not show this contrast elsewhere, when a perfectly typical contrast of [+voice], [-voice] presents itself as a viable alternative; though this does not make an allophone [gw] implausible (see also Devine and Stephens (p.94)). As to the second suggestion, *i.e.*, that *gu* represents two phonemes, we are presented with the problem that [w] can appear after this stop, but no others (Watbled accepts /k^w/ as phonemic). I am inclined to agree with Touratier (2005b) and Devine and Stephens (1977, 94) and state that there was a phoneme /g^w/ albeit in a very limited distribution. More importantly, however, I do not believe that the status of the labiovelars need concern the phonemic status of *u* and it is outside the range of this paper to discuss the labiovelars further here.

4 The Problem of Perception

4.1 Poetry

Before we are too satisfied with the conclusion that [w] is an allophone of a phoneme /u/, we must account for the alternation between trisyllabic and disyllabic variants of *silua*, *genua*, *tenuis*, and *suadeo* and *suesco* found in poetry. It seems that these must have been a conscious decision on the part of the poet: he might normally use the version of *silua* that was disyllabic, but when necessary, he could choose to use the version that was trisyllabic. He knew that there was a way to alter the pronunciation of the word and he knew which segment of the word altered the pronunciation. In short, he was aware of a contrast between these two sounds.

Let us begin with *genua* and *tenuis*¹⁹. Above, we discussed the possible creation of an optional rule in which a morpheme boundary between *u* and the endings allowed these words to scan trisyllabically. We could say that the [u] and [w] are in free variation in the position after *u*. This works perfectly well for *tenuis*, especially if we believe that the disyllabic variant is taken from allegro or popular speech (see Weiss (2009)), and perhaps licensed by analogy with the treatment of /u/ following the /l/ and /r/. Devine and Stephens (1977) believe many anomalous poetic forms were taken from popular or archaic speech where it was quite acceptable for a [w] to follow a consonant.

The case for *genua*²⁰ is only slightly more complex. If we believe that there was a rule as stated in 2.1 and 2.3 above that licensed free variation instead of an absolute conditioning environment, we

¹⁷ Though this was acceptable in Early Latin (*duenos*, *duellum*).

¹⁸ “Ce complexe [gw] ne se rencontre qu’entre une nasale et une voyelle...”

¹⁹ Weiss (2009) finds seven occurrences of [tenwis] in Lucretius and an occurrence in Verg. *Georg.* 3.180

²⁰ Verg. *A.* V 432

will need to know where the morpheme boundaries were analysed; however, this is difficult with *genua*. It seems that, as with *cornu*, the *u* was analysed as a stem morpheme, placing the morpheme boundary immediately after the /n/. Interestingly, Vergil's analysis of disyllabic *genua* suggests a morpheme boundary following *u*. It is still unsatisfactory to call this rule one that licences free variation since the trend in Classical Latin was to analyse all /u/ following a consonant (except as described above) as [u] (Devine & Stephens, 1977, 63). Conversely, however, the general tendency in spoken Latin was to analyse vowels in hiatus as semivowels, and there was a far more general tendency to permit semivowels after consonants (Devine & Stephens, 1977, 81 & 61).

This leaves us with the anomalous trisyllabic forms, e.g. *silua*²¹, *suadeo*²², *suesco*²³. *Silua* can be dealt with in much the same way as *tenuis* and *genua*. Based on the tendency for Classical Latin not to accept semivowels after most consonants, this could be a hyper Classical form (Devine & Stephens, 1977, 63), and it could be licensed on the model of the Greek poets (Hopkinson, 1982, 173). We need not revise the rule given in 2.3 above to accommodate it.

Suadeo and *suesco*, following the rules given in 2.3 above, can also be treated as hyper-Classical forms, just as *silua*. However, we must take into account the argument offered by Devine and Stephens that there was a phoneme /s^w/. An analysis of [sw] as a monophonemic element does not lend itself to a reanalysis as two phonemes, especially since the variant [w] is being realised as [u] in the anomalous form. We could use this as evidence to reject the analysis of /s^w/ as a phonemic element in Latin, but this leaves us with the somewhat unsatisfactory conclusion that the special phonetic properties of /s/ in a syllable allowed this to be the only acceptable consonant + [w] onset in Classical Latin. However, a third possibility presents itself: in which /s^w/ was the analysis in higher registers of Latin and /sw/ was the analysis in popular Latin, where consonant + glide clusters were acceptable as syllable onsets (Devine & Stephens, , 80). The poet then would accept two variant forms of a word with a *su* onset: one with the Classical phonology, and one with the popular phonology.

For a poet in a pinch, it is no great feat then to reanalyse popular Latin /sw/ as /su/, by analogy with forms like *silūa*. This is somewhat convoluted, but I believe nonetheless satisfactory. This is supported by Devine and Stephens (p.80) where they state the following of *suesco*: 'The higher frequency of *su-* in this word [*suesco*] may reflect the fact that, unlike *suavis*, it presumably was *su-* (verbal root) in Proto-Latin'.

If it is the case that there is a difference in frequency between forms which presumably had a different phoneme (/s^w/) in Proto-Latin, then it seems more likely that there should be this division between the varieties of Latin. It is important to note though that the evidence supporting the existence of a phoneme /s^w/ is not conclusive.

5 The Claudian letters

Further evidence shows that it was rather well recognised that there was a difference in the realisation of these allophones. The emperor Claudius attempted to introduce three new characters into the Roman alphabet, among which was an upside-down digamma to represent [w] (ϝ).²⁴ It has been found in many inscriptions from Claudius's time as emperor. We see it used seven times in CIL VI.921:

(5) (a) DIϝI (*diui*) (3x)

²¹ Hor. *Car.* I.23.4, Epod. 13.2

²² Lucr. IV.1157

²³ E.g. Lucr. I.60.

²⁴ Suet. *Claud.* 41.3

- (b) IIR (*uir*) (2x)
- (c) IVIENTVT[IS (*iuuentutis*) (1x)
- (d) O]CTAIIAI (*octauiai*) (1x)

Clearly this symbol is being used to represent consonantal [w]. In another inscription, CIL VI.353.5, we see <I> being used to represent [w] following an /r/ in the conditions described above: SERIIIAI.²⁵ In these cases it is clear that whoever was inserting the inverted digammas knew how to recognise a syllabic [w].

However, it is clear from inscription CIL VI.2034 that consistency was not always the case; we have:

- (6) (a) BOIE (*boue*) l.11
- (b) IOII[(*ioui*] l. 16
- (c) ARIIALIUM (*arualium*) l. 18
- (d) ARIIALIUM (*arualium*) l. 21
- (e) BOIE (*boue*) l. 21

But also:

- (7) (a) IIVINI (*iuini*) l. 8
- (b) CONSERVES (*conserues*) l. 9
- (c) NERVA (*nerua*) l. 19
- (d) VERBA (*uerba*) l. 19

Particularly interesting however, is *uouimus* which is found as:

- (8) (a) VOIIMVS l. 13
- (b) IOIIMVS l. 17
- (c) VOVIMVS l. 20

Note also that *arualium* is spelt with a <V> in line 4. This data could represent two things: 1) the person who inserted the digammas could sometimes recognise [w] and sometimes could not or 2) that he was still getting used to the new letter and kept making mistakes. I think the latter is the more likely. There is no phonetic environment where he consistently either does or does not recognise [w]: *arualium* shows inconsistency after /r/ and *vouimus* shows inconsistency intervocalically and word initially. Rather, his usage seems primed – we tend to see inverted digammas in clumps and <V> in clumps.

We should note, however, that this writing convention did not last beyond the reign of Claudius. I think that a lack of meaningful contrast between [w] and [u] can be cited as a cause for the discontinued use of the inverted digamma after the reign of Claudius. Where a writing convention produced a meaningful contrast, as in <C> [k] versus <G> [g], the convention was found far more frequently and became a significant part of the writing system.

6 Vulgar Latin Developments

The data given above show fairly plainly that there was a clear understanding of the difference between

²⁵ Though in the line above, the author writes: INVICTAI

[w] and [u]. Daniel Jones (1950), following the tradition of Sapir, acknowledges that to one extent or another, native speakers have intuitions about what sounds are contrastive in their own language. How then do we explain the apparent acknowledgement of contrast between [w] and [u] if they are allophonic in Classical Latin? Peter Hawkins (1984, 17ff) explains that it is possible to perceive allophones of another language when they are phonemic in one's own language. However, if we assume that speakers emulating Classical Latin were well acquainted with more common registers, which at this point were beginning to distribute the allophone [w] with the phoneme /b/ (see below), then this is not problematic.

This evidence suggests that there must have been some sort of split in the pronunciation of [w] occurring at some period before Lucretius, etc. are using these variant forms. Matthew McCullagh (2011, 87) states that "In Indo-European, and possibly also in Early Latin, /i/ and /j/ were allophones, as were /u/ and /w/."²⁶ It seems that Classical Latin then, has preserved the older status as we would expect with a prestige variety, especially one which determined 'correctness' by the written word as opposed to the spoken word and which considered archaism a mark of erudition. With [u] and [w] being graphemically identical, it is not difficult to imagine that this reinforced the allophonic status of [w].

In popular speech however, as we stated above, [w] was much more prominent and by approximately the end of the republic period had lost its velar element (Herman, 2000, 39).²⁷ In a very general sense, the labial lost some of its sonority, strengthening to a fricative, though the outcome was different depending on region. I offer the following graphic to clarify in a very general and simple way what happened to [w] in some positions:

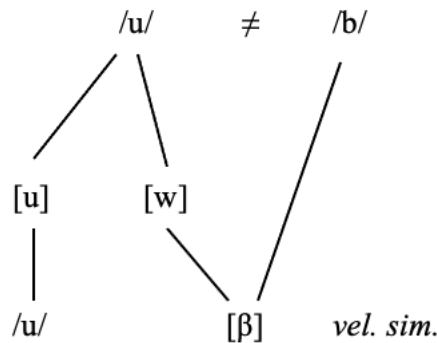


Figure 1

It should be noted that I differ from Adams in stating that there was no phoneme /w/ but rather that the [w] allophone began to be reanalysed as an allophone of the /b/ phoneme, though this analysis was not complete or universal.

7 Conclusion

In conclusion, we have established that [w] was in complementary distribution with [u], both of which were realisations of the phoneme /u/. Where there does seem to be the possibility of unclear distribution with regard to poetic usage, we can cite variant Vulgar Latin forms and morphophonemic analogy.

²⁶ He is of the opinion that /u/ and /w/ were separate phonemes during the classical period, though he does not distinguish between Classical Latin and Popular/Vulgar Latin.

²⁷ Though it should be noted, as Adams (2007) warns, "There was a degree of falling together of the original /b/ and /w/ but there are variations across the former Empire. The treatment of the two phonemes also varied according to their position in the word."

Where there is evidence that Latin speakers were able to discern a phonetic contrast between [w] and [u], we can conclude that the perceived contrast was based on the Vulgar variant of [w] which had lost its velar element and merged as an allophone of /b/. This explains the disyllabic and trisyllabic poetic variants, Claudius's intuition that the consonantal realisation of /u/ was a different sound and therefore should be represented with a different letter, and the separate development of [w] in the Romance languages. We can therefore state that there is significant evidence suggesting [w] was an allophone of /u/ and not a phoneme.

References

- Adams, J. N. (2007). *The regional diversification of Latin, 200 BC - AD 600*. Cambridge: Cambridge University Press.
- Ballester, X. (1996b). Fonemática de /u/ en latín. *Faventia*, 18, 137-140. Carr, P. (1993). *Phonology*. Basingstoke: Macmillan.
- Clackson, J. (2011). *A companion to the Latin language*. Oxford: Wiley-Blackwell.
- Clackson, J., & Horrocks, G. C. (2007). *The Blackwell history of the Latin language*. Malden, Oxford: Blackwell.
- Clark, J. Y., Colin, & Fletcher, J. (2007). *An introduction to phonetics and phonology* (3rd ed.). Malden, Oxford, Victoria: Blackwell Publishing.
- Devine, A. M., & Stephens, L. D. (1977). *Two studies in Latin phonology*. Saratoga: Anma Libri.
- Flobert, P. (1987). La réalité phonologique de /g^w/ en latin. In G. Serbat (Ed.), *Études de linguistique générale et de linguistique latine offertes en hommage à Guy Serbat* (pp. 103-112). Paris: Société pour l'Information grammaticale.
- Godel, R. (1953). Les semi-voyelles en latin. *Studia Linguistica*, 7, 90-99. Hawkins, P. (1984). *Introducing phonology*. London: Hutchinson.
- Herman, J. (2000). *Vulgar Latin*. University Park, Pa.: Pennsylvania State University Press. Hock, H. H. (1991). *Principles of historical linguistics* (2nd ed.). Mouton de Gruyter: Berlin.
- Hopkinson, N. (1982). Juxtaposed Prosodic Variants in Greek and Latin Poetry. *Glotta*, 60, 162-177.
- Jones, D. (1950). *The phoneme: its nature and use*. Cambridge: W. Heffer & Sons Ltd. McCullagh, M. (2011). The Sounds of Latin. In J. Clackson (Ed.), *A Companion to the Latin Language* (pp. 83-91). Oxford: Wiley-Blackwell.
- Palmer, L. R. (1954). *The Latin language*. London: Faber and Faber.
- Sapir, E. (1921). The psychological reality of phonemes (Le réalité psychologique des phonèmes). In D. G. Mandelbaum (Ed.), *Edward Sapir: Selected Writings in Language, Culture, and Personality* (pp. 46-60). Berkeley and Los Angeles, California: University of California Press.
- Serbat, G. (1987). *Études de linguistique générale et de linguistique latine : offertes en hommage à Guy Serbat*. Paris: Société pour l'information grammaticale.
- Sihler, A. L. (1995). *New comparative grammar of Greek and Latin*. New York ; Oxford: Oxford University Press.
- Touratier, C. (2005a). *Essais de phonologie latine : actes de l'atelier d'Aix-en-Provence, 12- 13 avril 2002*. Aix-en-Provence: Publications de l'Université de Provence.
- Touratier, C. (2005b). Système des consonnes. In C. Touratier (Ed.), *Essais de phonologie latine* (pp. 61-134). Aix-en-Provence: Publications de l'Université de Provence.
- Wallace, R. (1989b). The origins and development of the Latin alphabet. In W. M. Sennner (Ed.), *The Origins of Writing*. Lincoln, NE: University of Nebraska Press.
- Watbled, J.-P. (2005). Théories phonologiques et questions de phonologie latine. In C. Touratier (Ed.), *Essais de phonologie latine* (pp. 24-57). Aix-en-Provence: Publications de l'Université de

Provence.

Weiss, M. L. (2009). *Outline of the historical and comparative grammar of Latin*. Ann Arbor: Beech Stave Press.

A Psycholinguistic Experimental Approach to Affect and Vertical Position

Thomas Williams

University of the West of England, Bristol

Abstract. In this project I replicate ‘Study 1’ of Meier and Robinson’s (2004) experiment. This study presented participants with positive and negative words either at the top or the bottom of a screen, and recorded the time taken to evaluate each word. A shorter evaluation time implies a relation between the type of word and its position. This replication study differs from the original in that the words used are cross-checked across multiple corpus sources, and not from the researchers’ own whim, with the expectation that the word meanings are much clearer. The findings in this replication study partially supported Meier and Robinson’s original findings, namely Lakoff and Johnson’s (2003) proposal that positive is up. However, it was not found that negative is down, this is most likely due to the fact that the results were found to not be statistically significant.

Keywords: psycholinguistics; affect; vertical position

1 Introduction

The use of positivity and negativity being associated with, respectively, a higher vertical position and a lower vertical position can be observed in many different contexts. For example, recreational drug users ‘get high’ until the drugs wear off, where they experience a ‘comedown’. People with a happy disposition may say ‘things are looking up’; people who are sad may feel ‘low’. If you are alive, you are most likely to be found standing upright, if you are dead you are more likely to be found lying down below ground. The use of positivity and negativity corresponding with, respectively, a higher or lower vertical position can be observed cross-culturally and across a variety of contexts (Lakoff & Johnson, 2006; Li, 2010, p. 75; Meier & Robinson, 2009, p. 240).

Meier and Robinson (2004) suggest there is a ‘surprisingly physical basis’ for the association of spatial metaphor and its corresponding vertical position. More surprisingly, from the overview of related literature, this often-cited study appears to have had its findings unchallenged.

In the literature review I will argue that one flaw of the study were the positive and negative words selected for testing and will give an overview of similar findings which overall appear to support Lakoff and Johnson’s (2003) proposal that POSITIVE IS UP and NEGATIVE IS DOWN.

My project aims to conduct a modified replication of Meier and Robinson’s Study 1 of their 2004 paper, not selecting words on what appears to be intuition, but by selecting words taken from real-world data: *The British National Corpus*, the corpus-based *Collins Dictionary*, and publicly accessible messages sent through the social networking website *Twitter*.

Research into text processing carries implication into work on reading skills for children and L2 learners, and into signage and advertising discourse which incorporates the use of orientation in order to convey a message in the most effective way possible. It is a simple study, but it could provide a small insight into the relationship between language and mind.

2 Context

2.1 Overview of Literature

This project will centre around the concept of the orientational metaphor on a vertical axis. Lakoff and Johnson (2003, p. 5) state that ‘the essence of metaphor is understanding and experiencing one kind of thing in terms of another’ and that it defines our ‘human conceptual system’ (p. 6). From this, it is assumed there is a non-arbitrary link between the ‘human physical and cultural experience’ and the mental spatialization of orientational metaphors (p. 14). They cite a number of examples of orientational metaphors.

(1) HAPPY IS UP; SAD IS DOWN

I’m feeling up. That boosted my spirits. My spirits rose [...] I’m feeling down. I’m depressed. He’s really low these days.

(2) CONSCIOUS IS UP; UNCONSCIOUS IS DOWN

Get up. Wake up. [...] He rises early in the morning. He fell asleep. He’s under hypnosis. He sank into a coma.

(3) HEALTH AND LIFE ARE UP; SICKNESS AND DEATH ARE DOWN

He’s at the peak of health. Lazarus rose from the dead [...] He came down with the flu. His health is declining. He dropped dead. (Lakoff & Johnson, 2003, p. 15)

Lakoff and Johnson (2003, pp. 15-17) state that the reason for this correlation is that humans are subject to the same experiences of, say, standing upright when happy, awake or alive, and of being physically down (or perhaps horizontal) when sad, asleep or dead. Additionally, in film reviews, good films are given a ‘thumbs up’ and bad films are given a ‘thumbs down’ (Webster, 1996 in Meier & Robinson, 2009, p. 240).

These claims are supported by Nayak and Gibbs (1990, Experiment 6), who claim their study shows that ‘conceptual systems are organised around metaphors and those metaphors are used in comprehension’ (cited in Langston, 2002, p. 283). This is also supported in Li’s work on ‘the spatial self’ in Chinese autobiography. For example, looking at target domains and examples of manner verbs, Li finds that ‘sink’, ‘fall’ and ‘droop’ are used when describing ‘loss of control’ (2010, p. 75). Henceforth the concept defined as ‘expressions which have meanings involving position or motion with respect to a vertical scale’ (Nagy, 1974, p. 3) will be referred to as *verticality*.

In their study on eye movement and figurative language, Matlock and Richardson (2004) found that ‘fictive motion sentences’ (for example ‘the palm trees run along the highway’, where a motion verb is used but it does not literally express motion) can influence the eye movements of a participant. In this study it was found that participants would gaze for a longer duration on the subject of a sentence if it was described with a fictive motion verb (Matlock & Richardson, 2004, p. 913). Therefore, it is possible that there is some kind of underlying cognitive relation between literal movement and linguistic knowledge.

In a study conducted by Richardson et al. (2003) participants were asked to listen to short sentences while completing a ‘visual discrimination task’ and a ‘picture memory experiment’ in order to test spatial representations were cognitively activated during verb comprehension (Richardson et al, 2003, p. 767). They presented evidence that verb comprehension interacts with perceptual-spatial processes, at least with verbs that imply literal or metaphorical spatial relationships. The verbs were

categorised empirically as having either horizontal or vertical image schemas. The spatial orientation of the verbs' image schemas exerted influences on spatial perception and memory, interfering with performance on a visual discrimination task, and facilitating performance in the encoding of a visual memory (Richardson et al., 2003, p. 776).

This interference is also suggested in Paley's (1978) paper, where she states that both pictorial and linguistic information commanding 'up' and 'down' is interference. She continues:

'When subjects respond to word meaning, it is assumed that two stages of processing are involved: formation of a linguistic code for the word, and initiation of the motor response.' (Paley, 1978, p. 71)

In the same publication Paley cites a study where:

'[A] reduction of interference from irrelevant words [was shown] to colour judgements when the response keys were labelled with colours rather than colour words.' (Paley, 1978, p. 71)

This is discussed further in the 'Limitations' section below.

2.2 Why the Sunny Side is Up

The idea of verticality and affect is taken further in Meier and Robinson's (2004) study, where participants were presented with 'positive' or 'negative' words, shown either at the top or bottom of a computer screen and the participants 'were instructed to evaluate each word as quickly and as accurately as possible' (Meier & Robinson, 2004, p. 244) with their response times recorded. The researchers found that the position of a word and the affect of the word correlated – thus adding evidence to Lakoff and Johnson's proposal (2003) that HAPPY IS UP; SAD IS DOWN.

However, the words used in this study may not have been appropriately selected. To test this, I searched for a sample of the words used in the British National Corpus, finding many had collocates to suggest that some words categorised either as positive or negative could equally be argued to suggest the opposite. In particular, 'critical' is listed as a negative word, and while examples exist of it being used negatively, a search of the British National Corpus also returns collocates of critical with 'theory', 'mass' and 'thinking' – phrases which are not necessarily negative at all. And given the participants in this study were in fact undergraduate students, should they not have viewed critical as a positive quality? Another questionable example is sleep, which is listed as a positive word — yet, if Lakoff and Johnson (2003) are correct, 'CONSCIOUS IS UP; UNCONSCIOUS IS DOWN'. Additionally, sleep is sometimes used as a euphemism for death ('putting the dog to sleep'), which is most likely to be perceived as negative — again, this is ambiguous and could have led to incorrect trials in Meier and Robinson's study. In my opinion, it is essential that the words used in my project are crystal clear to the participants. If the words used are too open-ended, participants may become uninterested or confused during the procedure, thus giving unreliable results.

3 The Study

At the time of the study, I had been unable to find any published replication studies of this experiment. By replacing the word list with one based on 'real life' data, it is hoped this replication should reduce any incorrect trials and provide another insight into verticality and affect.

3.1 Methodology

Twenty participants were selected from a convenience sample of students at the University of the West of England, Bristol.

For this experiment, one hundred words were selected from multiple corpus sources. This will enable the words to be selected from their real-life usage. The identified sources were: the *Brigham Young University British National Corpus* [accessed from <http://corpus.byu.edu/bnc/>], which while it arguably does not represent up to date modern usage, it is however a very large database of verbal and written data; *Collins Dictionary* [accessed from <http://www.collinsdictionary.com/>] as this is a corpus-based dictionary; and *Twitter* [accessed from <http://twitter.com/search-home>] although (from personal experience) the search tool is more limited, collecting a sample of ‘tweets’ will help identify very recent, publicly accessible usages of each word. The words are then categorised into either positive or negative, justified from the corpus data collection²⁸.

Participants are seated in front of a computer with a standard QWERTY keyboard. The ‘1’ key represents POSITIVE, ‘5’ represents NEGATIVE. Using the exact same program as in Meier and Robinson’s original study, a fixation cue (displayed as +++) appears at the centre of the screen before moving towards the top or bottom of the screen. This allows participants to focus on the position of the word before it appears. After 300 milliseconds (ms), the randomly selected word will appear in place of the fixation cue. Participants evaluate each word and press the corresponding button with as much accuracy and speed as possible. The next word is displayed after a pause of 300ms. The experiment continues until all words have been displayed. The experiment lasts around five minutes.

Tests where participants give an incorrect response to the displayed word are not counted. All accurate results will be processed through SPSS to work out and compare mean response times.

3.2 Limitations

The procedure is intended to be as close to the original study as possible. The only modification in this experiment will be the words that appear on the screen, and, due to technical limitations, the use of a QWERTY keyboard instead of a response box. This also poses another problem — I would have liked to have replaced the response keys in a repeat study with some kind of pictorial representation of positive and negative as opposed to the use of arbitrary keys, in order to test if Paley’s claims that arbitrary keys interfere with perception times (1978, p. 71) are reproducible in this study. Unfortunately, technical limitations also mean I will only be able to reproduce the first experiment of Meier and Robinson (2004).

It is difficult to control for cultural interpretation of each word. However, the wide cross-checking of each word from multiple corpus sources should control for as minimal misinterpretation of each word as possible.

²⁸ The words selected were as follows: Positive: happy, enthusiasm, good, healthy, courage, beauty, friendship, wonderful, freedom, joy, kindness, paradise, rewarding, bright, delight, nice, safety, brilliant, fantastic, celebrate, lovely, achievement, awesome, relax, pleasant, exciting, thrive, cheerful, shine, thrilled, excellent, fortunate, lucky, prosperous, peaceful, harmony, trust, congratulate, welcomed, warmth, clarity, satisfied, calm, enjoy, lush, win, amazing, glad, grateful ; Negative: heartless, deceived, hate, sad, despair, grim, depressed, low, sickness, doomed, calamity, disaster, catastrophe, exhausted, upset, apprehension, fear, grief, disgust, inhumanity, terror, tragedy, gloom, downcast, misery, mournful, impoverish, dismay, sorrow, anguish, regret, terrible, suspect, anger, hateful, death, lonely, greed, helpless, stink, wrong, panic, broken, faulty, frustration, crisis, boredom, murder, abuse.

The complete data set alongside the corpus justification has been omitted for space reasons.

Another limitation to the study itself which is not controlled for is whether the position of a word, regardless of its affect, affects the reaction time of the processing of the word.

3.3 Results

After incorrect trials were removed, a univariate analysis of variance (ANOVA) was carried out to find and compare the mean response times of all participants. These findings were put against word valence and position as shown in Figure 1. In tests of between-subjects effects there were no significant mean differences between groups for valence ($p = 0.302$), position ($p = 0.223$) or for the comparison between both results (valence and position: $p = 0.159$). For valence and position $F = 1.988$ – this is not significant. An independent samples t-test (Levene's test for equality of variances) was then carried out between only the positive words and vertical position. It was found the differences between the higher and lower position for positive words only is significant ($t = 2.022$, $df: 978$, $p < 0.05$).

The same test was carried out again, this time on the negative words only. It was found the difference in results between the higher and lower position for negative words is not significant ($t = 126$, $df = 978$, $p > 0.05$).

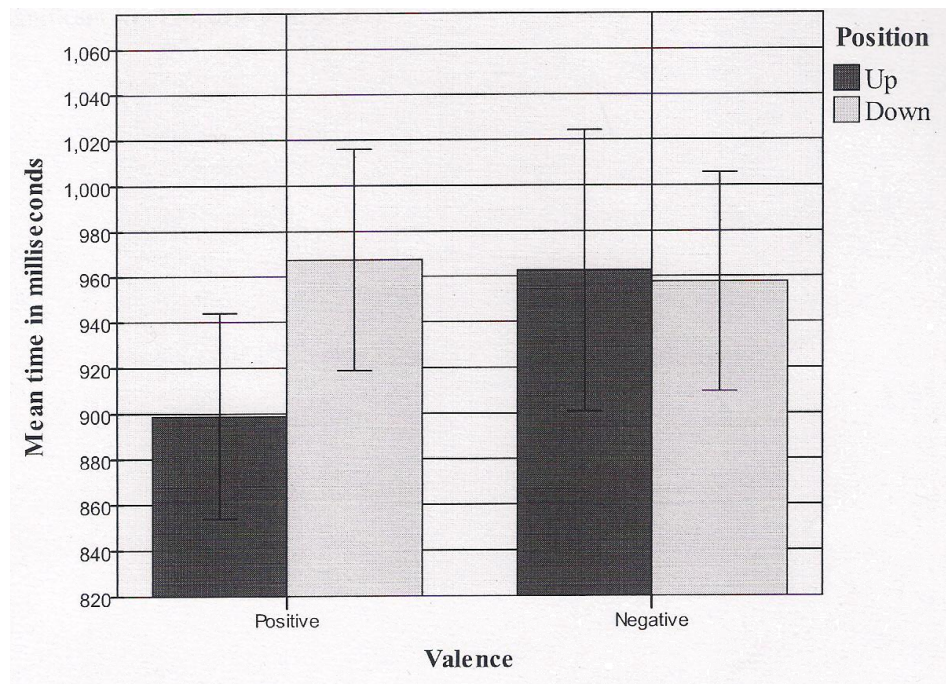


Figure 1: Mean response times against valence and position. Error bars: 95% confidence interval.

4 Conclusion and Discussion

The mean reaction times of the participants appear to be roughly in line with the results of Meier and Robinson (2004, p. 255), however in my study participants were on average slower by around 100ms compared to the original study.

As the words were selected to be more overtly positive or negative, I expected to find a more marked difference in the mean reaction times. However, the overall results were not statistically significant, and this could most likely be put down to the reduced number of participants in this study.

It is possible that in order to gain more participants for the study, an incentive could have been offered in exchange for their time and participation.

It is interesting to note that there is significance between verticality and the positive words only, and it appears that the results are following the same trend as found in the original study conducted by Meier and Robinson (2004). Additionally, where reaction times were overall much faster for positive words in the higher position, this study appears to have replicated similar results as in the original study. I speculate that there probably is a relation between verticality and affect, and it appears more marked for positive words. However, the underlying reasons as to why this happens are not clear. It is with pure speculation that I suggest this marked difference in the results for the positive words may arise as participants see and pay more attention to words placed at the top of a screen as they are deemed ‘more important’ — also, positive words could be looked upon more favourably, hence the shorter reaction times — while it be stressed this is mere speculation and requires further study to back up these claims, this could be linked with Paley’s (1978) findings, if it is assumed participants are unconsciously biased toward positivity.

References

- Lakoff, G. & Johnson, M. (2006). *Metaphors We Live By*. Chicago and London: The University of Chicago Press.
- Langston, W. (2002). Violating Orientational Metaphors Slows Reading. *Discourse Processes*, 34(3), 281-310.
- Li, J. (2010). Verticality, Horizontality, and States of the Self: Cognitive Metaphors for the ‘Spatial Self’ in Chinese Autobiographical Writings. *Metaphor and Symbol*, 26(1), 68- 95.
- Matlock, T. & Richardson, D. C. (2004). Do eye movements go with fictive motion? In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, 909-914.
- Meier, B. & Robinson, M. (2004). Why the Sunny Side is Up: Associations Between Affect and Vertical Position. *Psychological Science*, 15(4), 243-247
- Meier, B & Robinson, M (2009). The Metaphorical Representation of Affect. *Metaphor and Symbol*, 20(4), 239-257.
- Nagy, W (1974). *Figurative Patterns and Redundancy in the Lexicon*. [Unpublished doctoral dissertation]. University of California at San Diego.
- Paley, S. A. (1978). Judging pictorial and linguistic aspects of space. *Memory & Cognition*, 6(1), 70-75.
- Richardson, D. C. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, 27(1), 767 – 780.
- Ritchie, L. D (2008). X IS A JOURNEY: Embodied Simulation in Metaphor Interpretation. *Metaphor and Symbol*, 23(1), 174 – 199.

Acknowledgements

I would like to thank Professor Brian Meier of Gettysburg College for allowing me access to the original computer program used for his 2004 study with Michael Robinson.

contents

ULAB National Committee 2012	4
ULAB Local Committee 2012	4
Introduction to the ULAB II Proceedings <i>Lydia Wiernik and Caitlin Wilson, University of Edinburgh</i>	5
Old Frisian Gemination-Degemination through Strata <i>Jurij Božič, University of Ljubljana</i>	7
On Tocharian Exceptionality to the Centum-Satem Isogloss <i>Richard Littauer, University of Malta</i>	18
Reanalysis and Regularisation: The Development of Morphological Structure <i>Gwilym Lockwood, School of Oriental and African Studies, University of London</i>	26
‘English corrupted by Singaporeans’: a Study of Speaker Perceptions of Colloquial Singapore English in 2012 <i>Nicholas Marsh, University of Oxford</i>	33
The Grapheme <V> in Classical Latin: Did it Represent Two Phonemes or Three? <i>Devon Smith, University of Cambridge</i>	49
A Psycholinguistic Experimental Approach to Affect and Vertical Position <i>Thomas Williams, University of the West of England, Bristol</i>	59

2012