DEPARTMENT OF LINGUISTICS
FACULTY OF LANGUAGES AND CULTURES
SOAS UNIVERSITY OF LONDON

SOAS WORKING PAPERS IN LINGUISTICS

VOLUME 17, 2015

SOAS Working Papers in Linguistics

Volume 17
2015

Edited by

Charlotte Hemmings & Vivian Lee

Department of Linguistics
Faculty of Languages and Cultures
School of Oriental and African Studies
University of London
Thornhaugh Street, Russell Square
London WC1H 0XG
UK

Tel: +44 (0)20 7898 4044
Fax: +44 (0)20 7898 4009
E-mail: linguistics@soas.ac.uk
Website: www.soas.ac.uk/linguistics

Published by the Department of Linguistics of the School of Oriental and African Studies
University of London
Thornhaugh Street, Russell Square, London WC1H 0XG, UK

British Library Cataloguing in Publication Data
A Catalogue record for this book is available from the British Library

ISBN 0 7286 0404 3 9780728604049

Printed in Great Britain by The Print Room
SOAS

Contents

Editorial Note

July 2015

We are pleased to present the 17th Volume of the SOAS Working Papers in Linguistics. This volume features research in progress of staff, research students and research associates of the Department of Linguistics at SOAS.

The quality and variety of papers in this volume reflects, as always, not only the work of the department, particularly in the field of theoretical linguistics, but also the diversity of linguistic research and documentation currently being carried out at SOAS within the Faculty of Languages and Cultures.

As in previous volumes, a variety of languages from Africa, Asia and the Middle East are discussed, as well as languages and linguistic contexts in Australia, the Americas and Europe. Some papers provide synchronic analyses of particular linguistic phenomena, whilst others adopt a diachronic perspective. Hence the papers not only cover a wide geographical area but also a broad timespan from 7th century Arabic to Classical Tibetan and medieval Lingua Franca. Finally, the papers employ a wide range of methodologies, including corpus-based studies; experimental studies; questionnaires and linguistic landscapes.

The papers included in this volume are presented in the following sections: Phonetics and Phonology; Syntax and Semantics; and Historical Linguistics and Sociolinguistics.

We would like to thank everyone who helped and advised us in the editing process. In particular, thanks are due to Connor Youngberg and Lutz Marten for sharing their previous editorial experience; Lu Lu for her tireless help with copy editing; the anonymous reviewers for reviewing the papers in the volume; and the authors for their patience throughout the process. We would also like to thank the SOAS Linguistics Department for their generous support and Deborah Nelson and the Faculty staff for technical support in publishing and hosting the papers online. Last but not least, we are very grateful to Steve Redding; the staff of the SOAS Print Room, and Marcos Carbonell for their assistance in the printing process.

The Editors

Charlotte Hemmings
(ch33@soas.ac.uk)

Vivian Lee
(Vivian_Lee@soas.ac.uk)

Phonetics and Phonology

SOAS Working Papers in Linguistics Vol. 17 (2015): 3-31

Coalescence in Japanese Dialects is Diachronic[1]

Connor Youngberg
267787@soas.ac.uk

1. Introduction

In this paper, I examine the status of coalescence in the Owari dialect of Japanese and consider the following question: is coalescence in this dialect a synchronic hiatus resolution process? The Owari dialect exhibits the coalescence of the Middle Japanese vowel sequences /ai/ /oi/ and /ui/ to [æ:] [ø:] and [y:], exemplified in (1).

(1) **Examples of coalescence in Owari**

	Owari Japanese	Underlying form	Gloss
a)	[ræ:neN]	/raineN/	'next year'
b)	[osø:]	/oso-i/	'slow.NP'
c)	[fury:]	/furu-i/	'old.NP'

The synchronic status of coalescence is explicitly assumed in Terakawa (1985) and Yamada & Niwa (1989). This article critically examines data from the Owari dialect and offers a discussion and analysis of problematic facts. I extend the account of coalescence presented in Youngberg (2013) and argue that coalescence in Owari is no longer a synchronic process. Further evidence is drawn from dialectology works such as National Institute for Japanese Language and Linguistics (NINJAL 1968) and Ebata (2013) in addition to data recorded in the field[2].

Section 2 presents an introduction to Owari dialect of Japanese and coalescence. I examine the Owari data in further depth and point out problems forced by synchronic analysis of coalescence. I examine simple and compound nouns as well as adjectival and verbal conjugation. Exceptions are pointed out in both simple and compound nouns.

The article then discusses issues with a synchronic analysis in relation to exceptions and gives evidence that supports an analysis of Owari vowels in which they are fully specified in the lexicon due to the diachronic status of coalescence. Section 3 introduces Analytic and Non-Analytic domains (Kaye 1995, Yoshida Y. [1995] 1999) which is applied to compound exceptions. In section 4 I examine the facts regarding verbal conjugation which suggest that paradigms are listed in the lexicon with no active conjugation. Section 5 discusses pitch accent assignment and shift in nouns and verbs, focusing on the behaviour of the long vowels [æ:] [ø:] and [y:]. I conclude by claiming that all underlying representations are specified with the long vowel and that coalescence is no longer an active process in the phonology of speakers based on current data.

[1] I thank Monik Charette, Florian Breit, the editors and an anonymous reader for comments. All remaining errors are mine alone.
[2] Recordings were held in Ichinomiya City, April 2013.

2. The Owari dialect

The Owari dialect is spoken in Aichi Prefecture in central Japan. Coalescence is exhibited in the Nōbi plain between Nagoya City and Gifu City with locations to the south of Nagoya (e.g. the Chita peninsula) lacking coalescence. The Owari dialect is an Eastern dialect along with the Standard/Tokyo dialect (Katō 1977, Shibatani 1990:189) and has a Tokyo type accent pattern (Kindaichi 1977). The dialect does have some typically Western features, however, such as the usage of the negative suffix /-(a)N/ rather than Eastern /-(a)nai/ e.g. [kakaN] 'write-NEG.' (Keshikawa 1983, Yamada & Niwa 1989). See Hikosaka (2014) for recent discussion on the Western and Eastern aspects of Owari Japanese. Data below is drawn from NINJAL (1968), Terakawa (1985), Ebata (2013) and my own field notes. I assume for the moment that coalescence is synchronic, following Yamada & Niwa (1989). The vowel inventories for the two dialects are compared in (2). Assuming synchronic coalescence, underlying and surface representations of the vowel sequences are given for selected examples in (3), utilizing CV representations following Lowenstamm (1996) with Element Theory representations (Kaye, Lowenstamm & Vergnaud 1985, Backley 2011).[3]

(2) **Vowels in Tōkyō and Owari Japanese**

a. Tōkyō vowels

i i: ɯ ɯ:

e e: o o:

a a:

b. Owari vowels

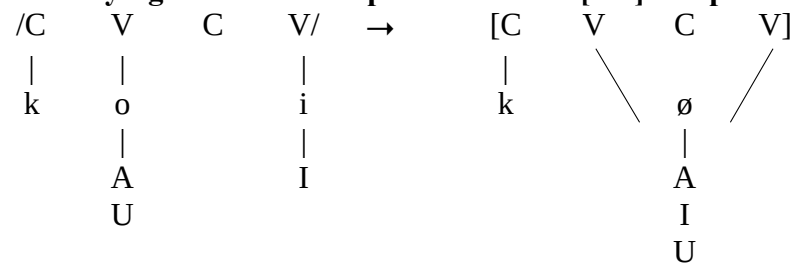
i i: y: u u:

e e: ø: o o:

æ:

a a:

(3) **Underlying and surface representation of [kø:] 'carp'**



[3] Previous arguments against branching constituents for Japanese are found in Yoshida S. (1996), Yoshida Y. (1999) and an extension of this argumentation can be found in the author's forthcoming thesis. See also Labrune (2012) for a similar mora based account.

The segmental phonology of the Owari dialect is identical to that found in Tokyo Japanese with the exception of the vowel inventory, exhibited in (2). Vowel coalescence has altered vowel sequences found in Middle Japanese, which are preserved in Tokyo and Kyoto Japanese. The vowel sequences /ai/ /oi/ and /ui/ are realized in Owari as [æ:] [ø:] and [y:] respectively through coalescence as exemplified in (3) with underlying vowel sequences.[4] I now examine coalescence as found in major word classes in Owari.

2.1. Nouns

I first examine nouns exhibiting coalescence. Consider the following data in (4).[5] Let us assume for now that coalescence is synchronic following Terakawa (1985) and Yamada & Niwa (1989). Data in (4) is drawn from fieldwork unless otherwise noted.

(4) **Owari Nouns**

	Owari	UR	Gloss
a.	[ræ:neN]	/raineN/	'next year'
b.	[ɕiharæ:]	/ɕiharai/	'payment'
c.	[æ:mæ:]	/aimai/	'unclear'
d.	[dæ:koN]	/daikon/	'daikon'
e.	[kawæ:so:]	/kawaiso:/	'poor thing'
f.	[æ:chi]	/aichi/	'Aichi (Pref.)'
g.	[æ:biki]	/aibiki/	'affair'
h.	[hæ:]	/hai/	'ashes'
i.	[kæ:]	/kai/	'clam'
j.	[ambæ:]	/ambai/	'seasoning'
k.	[sæ:ko:]	/saikoo/	'best'
l.	[hæ:]	/hai/	'fly'
m.	[ø:]	/oi/	'nephew'
n.	[kø:]	/koi/	'fish'
o.	[sy:ka]	/suika/	'watermelon' (Ebata 2013)
p.	[ky:]	/kui/	'post'(Ebata 2013)
q.	[ugy:su]	/uguisu/	'mockingbird' (Ebata 2013)
r.	[y:ro]	/uiro/	'steamed rice cake' (Yamada & Niwa 1989)

In simple nouns, coalescence occurs wherever there is a vowel sequence of the shape V_1V_2 where V_1 is {a, o, u} and V_2 is {i}. While it was hypothesized in Youngberg (2013) that /e/ seemed to be a trigger for coalescence, I now note evidence which suggests this assumption is incorrect. Direct comparison of Tokyo and Owari forms leads the observer to suppose that /e/ is a trigger for coalescence. I note that this correlation is likely due to reduction of /e/ to /i/ which then triggered coalescence, as in

[4] For more general introductions to coalescence and hiatus resolution, see de Haas (1988) and Casali (1996, 2011). I refer the reader to Youngberg (Forthcoming) for a formal treatment of hiatus resolution within GP.

[5] While field recordings (April 2013) attest to secondary palatalisation on the consonant preceding coalesced vowels, it seems to be variable and further investigation is necessary. For simplicity, I omit this from the transcriptions in this article.

the Owari word [hæ:] < *hai < Tokyo [hae] ‘fly’[6]. I note that vowel sequences with /e/ in the V_2 position are often reduced synchronically in modern Tokyo and Kyoto Japanese, such as the verb /kaer-u/ ‘to go home-NP’ which is often realized as [kairu]. This verb exhibits coalescence in Owari following a medial step of vowel reduction, with the Owari form [kæ:ru] corresponding to Tokyo [kaeru]. In addition to reduction evidence, exceptions to an /e/ trigger hypothesis are found in Ebata (2013) such as [ue] ‘above’. I therefore discard the hypothesis of /Ve/ as a context for coalescence and note that these words do not constitute exceptions.

2.1.1. Considering exceptions and other contexts

Loanwords do not exhibit coalescence for the speakers I have consulted as in the word [uisuki:] ‘whiskey’, nor do any coalesced loanwords appear in texts from NINJAL (1968), Terakawa (1985) or Ebata (2013). The evidence for coalescence in mimetic and reduplicated words is lacking, though some reduplicated words such as Tokyo [iroiro] ‘various’ derived historically from [iro] ‘color’ may exhibit coalescence upon further examination. Only one such word is found in Ebata (2013), with Tokyo [iroiro] ‘various’ being realized in Owari as [irø:ro].

A few simple nouns are found as exceptions which are problematic, such as the Sino-Japanese words [ai] ‘romantic love’ and [koi] ‘love’ which are identical in both Tokyo and Owari Japanese. These words could be considered non-dialect words, as [koi] and [ai] are rather literary, artistic or dramatic words for love. Another simplex word which also resists coalescence is [joip:ari] ‘nightowl’, drawn from Ebata (2013). This word, however, is rare in Tokyo Japanese and is not attested in the Owari materials from NINJAL (1968) or Terakawa (1985). [7]

Considering this, the above exceptions might simply be treated in the same way as loanword: there is no attestation of these words in the conversation transcribed in NINJAL (1968) or Terakawa (1985). I do note that though [ai] ‘love’ behaves exceptionally, the closely related Tokyo word [aikjo:] ‘emotions’ is attested in Ebata (2013) and realized as [æ:kjo:]. As simple exceptions are so few, they cannot truly be considered problematic for the positing of synchronic coalescence. We will see shortly, however, that these are not the only exceptions. I now move on to nouns with the locative suffix which exhibit coalescence.

2.1.2. Noun + locative

Nouns ending in the vowel {a, o, u} exhibit coalescence when suffixed with the locative particle /-i/ (Terakawa 1985:28). This is evidenced in texts from NINJAL (1976) and Terakawa (1985:28;34-49). Consider the citation forms [nagoja] ‘Nagoya’ and [furo] ‘bath’ and the locative forms /nagoja-i/ [nagojæ:] ‘Nagoya-LOC’ and /huro-i/ [furø:] ‘bath-LOC’. See the data in (5) and exceptional data in (6).

---

[6] I assume the general understanding in Element Theory (Kaye, Lowenstamm & Vergnaud 1985) and the work of Harris (1990, 2005) that reduction is loss or simplification of an elemental expression.

[7] In the Tokyo Japanese based Balanced Corpus of Contemporary Written Japanese (NINJAL 2012), this word has only 16 occurrences. This equates to roughly 0.16 tokens per million words, which is an extremely low frequency. If this word is rare in Standard Japanese, it may simply be treated by Owari speakers as a loanword of sorts.

(5) **Locative coalescence in Owari**

	Nominal stem	Locative /-i/	Gloss
a.	[nagoja]	[nagojæ:]	Nagoya
b.	[furo]	[furø:]	bath
c.	[jakuba]	[yakubæ]	town hall
d.	[zæ:ɕo]	[zæɕø:]	one's own residence/town

(6) **Locative exceptions**

a.	[dokojara]	[dokojarai]	somewhere
b.	[asoko]	[asokoi]	over there

As can be seen in (5) and (6), the locative suffix does not consistently trigger coalescence. Terakawa (1985:28) states that while coalescence does occur with vowel final nouns in the locative, the suffix also surfaces as [i], as in [dokojarai] /dokojara-i/ 'somewhere-LOC' and [asokoi] /asoko-i/ 'over there-LOC'. No further discussion is given by Terakawa (1985)[8]. I presume that deictics are protected from coalescence. Further fieldwork is necessary to confirm whether the locative alternation is productive beyond the few tokens found in the texts above. I leave this question open, but assume for the moment that coalescence is synchronic for non-deictic nouns in the locative.

2.2. Owari verbal forms

I now examine verb alternations found in the non-past (NP) and past tense verbs where stem-final segmental loss in past tense forms feeds the process of coalescence in the Owari dialect. Verbs are divided into consonant-final and vowel-final stems (Bloch 1946). Verbs exist in which coalescence has altered vowel sequences found *within* stems, such as /hair-/ 'enter' e.g. [hæ:ru] 'enter-NP'. Coalescence also occurs at the stem-suffix boundary when certain consonant-final stems lose their stem final consonant. The changes which delete or assimilate stem-final segments are known collectively as *onbin*. We focus here on the deletion of stem-final {k, g} in Tokyo and {k, g, s} in Owari.

Coalescence has been analyzed as occurring synchronically in verbs (Terakawa 1985, Yamada & Niwa 1989), with vowel sequences created at the boundary between a stem terminating in {k, g, s} and a consonant-initial suffix. Certain suffixes, such as the past /-ta/, trigger consonant assimilation or elision when adjacent to a stem-final consonant. An example is the stem /kak-/ 'write'. Consider the non-past (NP) of this stem with the NP suffix /-(r)u/, realized as [kaku] 'write-NP' with the initial consonant unrealized. The stem-final consonant is elided when the stem terminates in {k, g} when consonant initial morphemes such as the past tense morpheme /-ta/ are suffixed, with underlying /kak.ta/ thus surfacing as [kaita] 'write-PAST' in Tokyo Japanese.[9] This then feeds coalescence in Owari Japanese under a synchronic analysis, giving the Owari form [kæ:ta] 'write-PAST'.

This altered verbal stem is traditionally called the *onbin* or 'euphonic' stem (cf. the chapters in Iitoyo, Hino & Satō 1982-1986; Frellesvig 1995, 2010). Stems may elide

[8] Ebata (2013) claims that the locative particle in Owari is [e] as in Standard Japanese, but this may be due to shift in the speakers consulted.

[9] This assumes that the loss of the stem-final consonant is synchronic as in Davis & Tsujimura (1991).

their final segment as shown above or they may exhibit assimilation to an NC cluster or geminate, as with the stem /ɕaber-/ 'chat' /ɕaber-ta/ [ɕabet:a] 'chat-PAST' or /yom-/ 'read' /jom-ta/ [jonda] 'read-PAST'. *Onbin* are a set of three lenition and assimilation processes that affect the final segment of the suffix: elision, nasal assimilation and gemination.[10] First, consider the following Owari verbs in the Non-Past, Negative and Past forms in (7) and (8). Note the past tense verbs bolded in (8b) where coalescence occurs.

(7) **Owari verbs with Vowel-Final stems** (Ebata 2013)

Non-Past	Negative	Past	Gloss
/Stem-(r)u/	/Stem-(a)N/	/Stem-ta/	
iki-ru	iki-N	iki-ta	'live'
jame-ru	jame-N	jame-ta	'quit

(8) **Owari verbs with Consonant-final stems** (Ebata (2013)

a. Stems exhibiting nasal and voicing assimilation

/Stem-(r)u/	/Stem-(a)N/	/Stem-ta/	
ajob-u	ajob-aN	ayonda	'toddle'
am-u	am-aN	anda	'knit'
sin-u	sin-aN	sinda	'die'

b. Exhibiting gemination

a(w)-u	awa-N	at:a	'meet'
tats-u	tata-N	tat:ta	'stand'
jor-u	jor-aN	jot:a	'stop'

c. Exhibiting consonant elision and coalescence

aruk-u	aruk-aN	**ary:ta**	'walk'
isog-u	isog-aN	**isø:da**	'hurry'
kobos-u	kobos-aN	**kobø:ta**	'spill
hirak-u	hirak-aN	**hiræ:ta**	'open, intr.'

In both Owari and Tokyo, *onbin* changes to the stem-final segment are 'triggered' by certain /t/-initial suffixes, such as the gerund /-te/, past /-ta/, conjunctive /-tara/, concessive /-temo/ and assortative /-tari/. I assume for the moment that *onbin* changes are derived synchronically, following Davis & Tsujimura (1991) and Yoshida S. (1996).[11] Out of the three possible stem alterations, elision occurs in {k, g} final stems in Tokyo or {k, g, s} final stems in Owari.

These changes have been captured by Davis & Tsujimura (1991) who approach the *onbin* changes in an Autosegmental rule-based account. I now discuss the application of

[10] For more on the diachrony of *onbin*, I refer the reader to Frellesvig (1995, 2010) for full discussion of its emergence in Old and Early Middle Japanese, see also Section 4. For an overview of the geolinguistic variation of *onbin* forms, see Uwano et al (1989), Iitoyo, Hino & Satō (1982-1986) and NINJAL (1989-2006)

[11] A synchronic process of vowel coalescence is assumed to be active by Terakawa (1985) and Yamada & Niwa (1989). Problems with such an analysis are discussed later in this article.

this proposal for Tokyo consonant stems compare Tokyo stems ending in {k, g} which exhibit lenition to the Owari stems where {k, g, s} elide. The account put forth by Davis & Tsujimura (1991) necessitates an ordered set of rules, as shown in (9). This rule ordering is applied to Tokyo and Owari {k, g, s} final stems in (10) and (11).

(9) **Tokyo process ordering proposed by Davis & Tsujimura** (1991)

a. Sonorant delinking (affects /r/, /w/ final stems)
b. Voicing assimilation (affects suffixes adjacent to [+voice] stems)
c. Nasal linking (affects /b/ final stems)
d. Labial delinking (affects /m/, /b/ final stems)
e. i-epenthesis (affects /s/, /k/, /g/ final stems)
f. Velar deletion (affects /k/, /g/ final stems)

(10) **Gerund form derivations in Tokyo Japanese**

Input	[+voice] spread	Epenthesis	Elision	Output	Gloss
/das.te/	-	/das.te/	-	[daɕite]	‘put-GER’
/kog.te/	/kog.de/	/kog.ide	/ko.ide/	[koide]	‘row-GER’
/kak.te/	-	/kak.ite/	/ka.ite/	[kaite]	‘write-GER’

(11) **Gerund form derivations in Owari Japanese**

Input	[+voice] spread	Epenthesis	Elision	Coalescence	Output	Gloss
/das.te/	-	/das.ite/	/da.ite/	/dæ:te/	[dæ:te]	‘put-GER’
/kog.te/	/kog.de/	/kog.ide	/ko.ide/	/kø:de/	[kø:de]	‘row-GER’
/kak.te/	-	/kak.ite/	/ka.ite/	/kæ:te/	[kæ:te]	‘write-GER’

Following (9) with alterations to Elision and the addition of a rule of Coalescence, we derive the Gerund forms found in (10) for {k, g, s} final stems in the following manner. Voicing assimilation first applies upon suffixation and then the vowel /i/ is epenthesised following Vowel Epenthesis as in (9e) to the consonant cluster formed by the stem and suffix. Velar deletion (9f), then takes place in this intervocalic cluster affecting {k, g} in Tokyo and {k, g, s} in Owari. For Owari, we expand the affected segments for consonant deletion and thus I consider this rule to be one of more general elision rather than velar deletion. This final process then results a vowel sequence as in Tokyo [kaite] from /kak+te/ ‘write-GER’ cf. Old/Middle Japanese *[kakite]. To derive the Owari forms in (11), we add and apply a rule of coalescence which applies last.

For reasons of space, I do not discuss a Government Phonology account of *onbin*, though see Yoshida S. (1996) for an analysis which utilizes similar processes without rule ordering. I note that the above account as well as that provided by Yoshida S.

(1996) utilizes arbitrary processes to give the derived verbal form.[12] The evidence for any synchronic *onbin* processes is somewhat thin on the ground. I return to this issue in Section 4.

Other consonant-final stems undergo either nasalization or gemination. Stems which terminate in /n/ or /t/ undergo no alterations. The stems terminating in {b, m} spread the feature [+voice] to the suffixes above as per voicing assimilation in (9b). These segments then undergo Nasal Linking in (9c) and Delabialization in (9d), with stems surfacing with the segment [n] e.g. /job-ta/ [jonda] 'read-PAST'. Stems terminating in the consonants {w, r} surface with a geminate consonant, in which the final segment delinks following the rule of Sonorant Delinking in (9a) with concomitant spread of the following consonant to the now empty consonantal position. This is seen in verbs such as /ɕaber-ta/ [ɕabet:a] 'talk-PAST'.

Returning to coalescence, I note that Yamada & Niwa (1989) assume that the vowel sequences exists in the underlying forms. Under a derivational rule-based analysis such as that pursued by Tsujimura & Davis (1991), this is plausible and I have provided an account above. The vowel sequence found in Tokyo Japanese simply feeds a process of coalescence e.g. /kak-/ 'write' /kak+ta/ > /kakita/ > /kaita/ > [kæ:ta] 'write-PAST'. A coalescence rule would apply to any vowel sequence of the shape /Vi/ in the grammar and would not be specific to verbal derivations. I later discuss evidence, however, which suggests that this account may not be correct.

2.3. Adjectival forms

I now examine coalescence in adjectives, where it is triggered by the adjectival non-past suffix /-i/. In line with the above analysis, a coalescence rule applies to underlying vowel sequences created by suffixation. Consider the data in (12), where forms exhibit coalescence in the Non-Past forms. The underlying stem vowel resurfaces in Transformative and Hypothetical forms

(12) **Adjectives** (Ebata 2013)

	Non-Past /Stem-i/	Transformative /Stem-naru/	Hypothetical /Stem-kerja/	Gloss
a.	ureɕi-i	ureɕi-naru	ureɕi-kerja	'happy'
b.	takæ:	taka-naru	taka-kerja	'tall'
c.	nukutø:	nukuto-naru	nukuto-kerja	'warm'
d.	akary:	akaru-naru	akaru-kerja	'bright
e.	atsy:	atsu-naru	atsu-kerja	'hot'
f.	samy:	samu-naru	samu-kerja	'cold'

All adjectives exhibit alternations where the coalesced vowel is not present. For example, the stems for 'hot' or /atsu-/ and 'cold' or /samu-/ are realized in the non-past form as [atsy:] 'hot-NP' and [samy:] 'cold-NP'. The stems are also realized without the coalesced vowel when other consonant-initial suffixes are added, such as the Past tense /kat:a/ e.g. [atsukat:a] 'hot-PAST' and [samukat:a] 'cold-PAST'.

[12] These processes are unattested elsewhere in the synchronic phonology of Japanese.

2.4. Compound Nouns

A number of exceptions to coalescence in compound nouns are found in Ebata (2013. If coalescence is a synchronic process, all compound nouns should have coalesced vowels. If compounds are somehow protected from coalescence, this may be due to their structure. See the example compounds from Owari in (13).

(13) **Compounds in Owari** (Ebata 2013)

a.	[koibi]	'little finger'	[ko]	'small'	+ [ibi] 'finger'
b.	[nakaibi]	'middle finger'	[naka]	'middle	+ [ibi] 'finger
c.	[yosoiki]	'going out'	[yoso]	'outside'	+ [iki] 'go'
d.	[wataire]	'cotton stuffed'	[wata]	'cotton'	+ [ire] 'stuff'
e.	[tsy:ri]	'rainy season'	[tsuju]	'monsoon'	+[iri] 'enter'

The above compounds are words where coalescence is absent with the exception of [tsy:ri]. The compound [tsy:ri] may be discounted as it is archaic and can be considered a simplex word originally derived from a compound. The other words above are protected from coalescence. I first hypothesize that the failure of coalescence to apply is due to the structure of compounds. In a Government Phonology analysis (Kaye, Lowenstamm & Vergnaud 1990) compound structure is captured with the use of analytic domains (Kaye 1995), which I discuss in Section 3.

2.5. Problems with a synchronic analysis of coalescence

At first glance, it seems that since there are clear morphological alternations, coalescence is a synchronic process which affects all Yamato and Sino-Japanese words where the context for coalescence is met. However, some exceptions to coalescence and a complicated process of derivation for verbs pose a challenge. An alternative analysis is also available: coalesced forms are lexically listed and coalescence is a diachronic sound change. In the remainder of this paper, I examine evidence which suggests that the latter view is correct.

In section 3, I refer to the structure of compound nouns with reference to analytic and non-analytic domains (Kaye 1995, Yoshida Y. 1999), discussed in the following section. Section 4 examines issues with the synchronic analysis of verbal derivation above. I examine the diachronic aspects of change, the behavior of new verbs, and experimental results which all suggest that verbal paradigms are no longer derived and are listed in the lexicon. Section 5 concludes the article by examining the interaction between pitch accent assignment and the vowels formed through coalescence.

3. Analytic and Non-Analytic Domains in English and Japanese

I now consider compounds and their structure utilizing the theory of analytic and non-analytic domains, proposed by Kaye (1995) to account for the interactions between morphological structure and phonological processes. The theory has been utilized previously for analysis of Tokyo Japanese by Yoshida Y. (1999). I first introduce the theory and then examine Owari compound nouns.

3.1. The visibility of morphology to phonological processes

It is uncontroversial that some morphological structure is visible to the phonology, but the question Kaye attempts to tackle is how much morphological structure is visible to

the phonology and what effect this structure has. The theory builds on the theories of morpho-phonology proposed by Kiparsky (1985) and Bromberger & Halle (1989).

Kaye (1995:302) states that "…Morphological structure has two effects on the phonology: little and none. These two interactions are called **Analytic** and **Non-analytic**." (Emphasis mine.) Analytic morphology has its own domain and thus analytic words have multiple domains to which phonology applies. Non-analytic words lack internal domains and non-analytic morphology brings no phonological domain to what may be understood neutrally to be a morphologically complex word. Analytic affixes are roughly equivalent to Level 2 affixes in Lexical Phonology (Kiparsky 1982, 1985) while non-analytic affixation is roughly equivalent to Level 1 affixation.

True compounds are composed of analytic domains. They are of the shape [[A][B]] or [[A]B] as in a word such as [[párent][hòod]] or [[dúst]mɨn] where each piece of morphological material has a domain in which phonology applies.[13] Non-analytic domains are words in which there are no visible domains with regards to phonology. They are of the shape [AB] as in [paréntal], which has penultimate stress assignment similar to the morphologically simpleword [agénda]. I unpack these notions in depth below.

Government Phonology (Kaye, Lowenstamm & Vergnaud (1990) eschews resyllabification and bans process ordering (Kaye 1992) as well as alteration to already computed structure. This is captured succinctly in the Projection Principle (Kaye, Lowenstamm & Vergnaud 1990), the Principle of Strict Cyclicity (Kean 1974, Kaye 1992) and the Minimality Hypothesis (Kaye 1992). The Principle of Strict Cyclicity (PSC) prevents alteration of previous phonological computation, such as stress assignment[14]. The Minimality Hypothesis bans any rule ordering or stratification and states that all processes may apply wherever the conditions are met. These statements are reproduced below.

(14) **The Projection Principle** (Kaye, Lowenstamm & Vergnaud 1990:221)
Governing relations are defined at the level of lexical representation and remain constant throughout a phonological derivation.

(15) **The Principle of Strict Cyclicity** (Kean 1974:179; Kaye 1992:142)
On any cycle A, no cyclic rule may apply to material within a previous cycle B without making crucial use of material uniquely in A.

(16) **The Minimalist Hypothesis** (Kaye 1992)
Phonological processes apply wherever their conditions are met.

[13] It is clear that 'dustman' is morphologically complex due to the preservation of the […st]m] cluster. In a simplex word, such a cluster would not be licensed as it is analysed within GP as a Rhyme-Onset-Onset cluster which is not possible in English. In this word, a domain-final nucleus is licensed in […st_] which provides Government Licensing (Charette 1990) to the preceding Coda-Onset cluster (Kaye 1990) with the following Onset being found in an independent domain. A lack of reduction of this cluster is a hallmark of a complex domain, yet no stress is found on *man* due to this portion of the compound being a dependent domain. See Kaye (1995) for more on licensing of positions in Analytic domains.
[14] See Scheer (2011) for critical discussion of Kaye (1992, 1995).

The Projection principle bans resyllabification operations. Following the PSC, alteration of an already processed domain is banned. The Minimality Hypothesis assumes that no strict ordering of processes is allowed, for the derivation of a well-formed phonological word. Processes apply without making reference to strata as in Lexical Phonology (Kiparsky 1985).

Given the above principles, exceptions to coalescence may then be explained as occurring within analytic domains. If coalescence is a process in the phonological grammar, it should occur wherever the context is met, unless the vowel sequences is in fact separated in the lexicon by a domain boundary e.g. [[ko][ibi]] 'little finger'.

The phonology portion of the morpho-phonology consists of two functions under Kaye's formalism: the Phonology function (φ) and the Concatenation function. The morphological domain in question is scanned by the phonology function with processes applying wherever their conditions are met. For us, this means a coalescence context where V1 is of the set {a, o, u} and V2 is /i/.

Once all possible processes have applied within a domain, this domain is concatenated with any neighboring domain and the process is repeated until all domains are processed and the word is ready for phonetic interpretation. More formally, we may consider a word ready for interpretation when all positions are licensed, following Kaye (1990) and Charette (1990).

3.2. [[A][B]]Analytic Domains

I now expand on the notion of analytic domains, which come in the shape [[A][B]] for a word composed of two independent domains or [[A]B] in a word with one independent domain and one dependent domain. The analytic domain shape of [[A][B]] is the structure of a compound noun. Consider the word 'blackboard' which is a compound consisting of two morphemes, [blák] and [bórd]. Kaye proposes a domain structure of [[blák][bórd]] for the resultant word [blákbòrd]. This word in fact has three domains: [blak], [bord] and [blakbord] or [A], [B] and [[A][B]], to which phonological processes apply. Note that the final result, [blákbòrd], has a primary and a secondary stress with no vowel reduction found in the unstressed vowel - this is caused by the assignment of stress to each domain [A] and [B] with the percolation of stress to the following concatenated domain, [AB]. The phonology then assigns one of the projected stresses as primary stress, namely the left domain in English.

3.2.1. Tokyo Japanese Analytic Compounds[15]

In Japanese analytic nominal compounds, Yoshida Y. (1999) proposes that the right-hand member of a compound determines the pitch accent of the resultant compound.[16] All words in Japanese come in both Accented and Unaccented variants. Following the GP analysis pursued by Yoshida Y. (1999), accented words have one accented nucleus, which is specified in the lexicon. Unaccented words are underlyingly unaccented with accent placed on the final nucleus during interpretation, wherever the

[15] I assume no distinction between Standard and Tokyo Japanese, though idiolectal variation does exist. I refer to the Standard as Tokyo Japanese, following Uwano (1977)

[16] I refer the reader to Yoshida Y. (1999) for more information on non-Yamato and longer compounds. See also Kubozono (2008), Nishiyama (2010) and Kawahara (2015) for other discussions of pitch accent.

end of the final domain occurs. Spread of the accent occurs until the initial nucleus in Tokyo Japanese. Consider the structure of Tokyo analytic compounds in (17). I follow Yoshida's notation below utilizing an asterisk for accented nuclei and an overline to represent high pitch.

(17) **Tokyo Analytic Compounds** (Yoshida Y. 1999:134-136)

a.	[ta ke]	+	[ha ɕi]	-> [ta ke ba ɕi][17]
	'bamboo'		'chopsticks'	'bamboo chop sticks'
b.	[a ta ma]	+	[ka zu]	-> [a ta ma ka zu]
	'head'		'number'	'head count'
c.	[sa sa]	+	[mi do ri]	-> [sa sa mi do ri]
	'bamboo grass'		'green'	'green of bamboo grass'
d.	[na ma]	+	[ta ma go]	-> [na ma ta ma go]
	'raw'		'egg'	'raw egg'
e.	[ka ni]	+	[ta ma go]	-> [ka ni ta ma go]
	'crab'		'egg'	'crab omelette'
f.	[u zu ra]	+	[ta ma go]	-> [u zu ra ta ma go]
	'quail'		'egg'	'quail omelette'

Above, the resultant compound retains the accent pattern of the righthand term. Yoshida Y. (1999) proposes that in analytic compounds the [B] domain determines the accent of the final [AB] domain and is the head of the compound [[A][B]]. Unlike stress languages, no secondary accents are possible in Tokyo Japanese pitch accent and thus no trace of the [A] domain accent remains[18]. If Owari compounds are analytic, we expect the [B] domain pitch accent to be preserved.

3.3. [[A]B] Analytic Domains

[[A]B] is also a proposed domain structure for compounds. Consider the words 'superman' and 'postman' in Southern British English discussed by Kaye (1995). The word [súpə̀màn] has a structure of [[super][man]] and thus has secondary stress on the ultimate nucleus. The stress is preserved in domain [B] in the final from but is realized as secondary stress. No vowel reduction occurs in [man].

[17] Some compounds exhibit *rendaku* or 'sequential voicing' where voiceless initial consonants {k, s, t, g} are realised as their voiced counterparts {g, z, d, b}. I refer the reader to works such as Itō & Mester (1986), Nasukawa (2005) and Labrune (2012) for further discussion of this phenomenon.

[18] An alternative interpretation is that pitch spread masks any realisation of the accent and that the projected nuclear structure in domain [A], posited by Yoshida Y. (1999), simply fails to be interpreted.

Compare this to the compound ‘postman’ which has the proposed structure of [[post]man]. The resultant form [póstmɨn] has no secondary stress on the ultimate nucleus and reduced vowel in the ultimate nucleus, which points to a domain shape of [[A]B]. Reduction processes and lack of secondary stress is a signal for a dependent domain in English.

This domain shape is also found in verbs and other constructions as with the past tense morpheme /-d/ or plural morpheme /-z/. Words with dependent morphemes are proposed to have the structure [[A]B]. Kaye (1995) notes that this structure is signaled by phonotactically odd words. This analytic structure creates otherwise unacceptable clusters within a language. Kaye (1995) notes that in a word such as [[dri:m]z], /m/ is not homorganic to /z/, while nasal homorganicity within a regular mono-morphemic word would be necessary for a nasal-obstruent cluster to hold. The [A] domain final empty nucleus is preserved and change to the domain final nasal /m/ is prevented through the PSC, where no alteration to domain [A] is permitted in an analytic word of the shape [[A][B]] or [[A]B] in the final concatenated [AB] domain and no vowel shortening is evidenced.

3.3.1. [[A]B] domains in Tokyo Japanese

This domain shape has only been proposed by Yoshida Y. (1999:146-151) for noun-particle combinations as in Accented [[namida]ga] ‘tear-NOM’ and Unaccented [[sakura]ga] ‘cherry-NOM’.

(18) **Representation of nominative derivations in [[A]B] domains**

a. [sa ku ra] + ga = [sa ku ra ga]

b. [na mi da] + ga = [na mi da ga]

I note that the same [[A]B] structure is possible for Japanese verbs and adjectives[19]. In Owari compounds, I expect compounds to exhibit preservation of the pitch of domain [B].

3.4. Non-Analytic morphology

Morphology which carries no domain information to the phonology, or in which no boundaries are visible to phonology, is **non-analytic morphology**, with a structure [A B] (Kaye 1995:308). An example of this cited by Kaye (1995) is the English suffix *-al* as in *parental*. *Paréntal* exhibits ‘stress shift’ which is in actuality the result of a non-analytic domain structure. If the word *parental* were analytic e.g. [[parent]al], stress would be assigned to the [A] domain e.g. [párɨnt] followed by computation of the full word giving *[párɨntɨl], as expected if the word were of the domain shape [[A]B]. Kaye notes that this stress pattern is the same as with a simplex word such as [agénda].

[19] Yoshida Y. (1999) proposes that verbs are Non-Analytic but this is dependent upon the assumption that accent is default antepenultimate for nouns as well as verbs. The current consensus, however, is that accent in verbs is different, with penultimate accent being the default. See Nishiyama (2010).

Non-analytic behavior in Japanese can also be seen in the following compounds drawn from Yoshida Y. (1999:98).

(19) **Non-analytic compounds in Tokyo Japanese**

a. [haná] ‘flower’ + [katá] ‘shape] = [hanágata] ‘star/popular person’

b. [uɛiro] ‘behind’ + [aɛí] ‘leg’ = [uɛiróaɛi] ‘hind legs’

c. [yamá] ‘mountain’ + [mitɛi] ‘street’ = [yamámitɛi] ‘mountain foot path’

d. [sakura] ‘cherry’ + [mitɛi] ‘street’ = [sakurámitɛi] ‘cherry road’

In the above compounds, preservation of the [B] domain accent as seen earlier is absent. Antepenultimate accent is found, identical to simple Japanese words which are longer than three nuclei such as Tokyo [murásaki] or [hototógisu] ‘cuckoo’.

Non-analytic forms are largely idiosyncratic or irregular forms. Kaye (1995) claims that any non-analytic form is listed separately in the lexicon, as these are not synchronically produced (Kaye 1995:311). He further states that *all irregular paradigms*, such as English strong verbs e.g. *wept*, ought to be listed in the lexicon separately and that the relation between irregular forms is diachronic and thus not relative to the phonology. This could easily be the case for Owari verb forms and would predict a lexicalist account of verbal paradigms.

3.5. Domain elision: from Analytic to Non-Analytic

Kaye (1995) further notes cases where analytic compounds may eventually become non-analytic as well as cases of speaker variation captured as variable domain structure. Domains may eventually fade, such as in the word *cupboard* [kʌ́bɨd] , which has elided its domains and undergone phonological change. The vowel [ɔ:] in ‘board’ is reduced to [bɨd] and no secondary stress is assigned as would be expected in an analytic domain. Also telling is the medial cluster not being realized as it might be in other compounds such as the nonsense word *capboard* [[kǽp][bɔ̀:d]].

Variation may also be explained through domain elision. This is noted by Kaye (1995) in compounds utilizing the word [metre]. Compare non-analytic [àltímətə̀] ‘altimeter’ to analytic [[áltĭ][mítər]] and analytic [[kílə̀][mítĕ]] versus non-analytic [kə̀lómətə̀].

3.6. Analytic to Non-Analytic domains in Japanese

Yoshida Y. (1999:144-146) also notes that there is domain elision found in the Tokyo Japanese interpretation of compound nouns. This is clear when a compound behaves as a simple word in its failure to retain the accent of an Accented [B] domain and exhibit antepenultimate accent.

(20) **Analytic variants of compounds** (Older speakers, Yoshida Y. 1999)

a.
*		*		(*) _____*
[na ma]	+	[ta ma go]	->	[na ma ta ma go]
‘raw’		‘egg’		‘raw egg’

b.
_		*		_____*
[ka ni]	+	[ta ma go]	->	[ka ni ta ma go]
‘crab’		‘egg’		‘crab omelette’

c.
___		*		_____*
[u zu ra]	+	[ta ma go]	->	[u zu ra ta ma go]
‘quail’		‘egg’		‘quail omelette’

(21) **Non-Analytic variants of compounds** (Younger speakers, Yoshida Y. 1999)

a.
*		*		____*
[na ma]	+	[ta ma go]	->	[na ma ta ma go]
‘raw’		‘egg’		‘raw egg’

b.
_		*		____*
[ka ni]	+	[ta ma go]	->	[ka ni ta ma go]
‘crab’		‘egg’		‘crab omelette’

c.
___		*		_____*
[u zu ra]	+	[ta ma go]	->	[u zu ra ta ma go]
‘quail’		‘egg’		‘quail omelette’

The above compounds in (21) show no preservation of the [B] domain pitch accent and instead pattern with simplex words with more than three nuclei such as [hototógisu] ‘cuckoo’ in exhibiting antepenultimate accent assignment. Elision is noted to exist in Japanese as well as in English, with Yoshida Y (1999:144-146) noting shift from analytic to non-analytic compounding found in some younger speakers. Compare the Japanese analytic [[uzura][tamágo]] to non-analytic [uzuratámago] with English analytic [[kílə][mítĕ]] versus non-analytic [kə̆lómətə̆]. This exemplifies the elision of domains in both Japanese and English.

3.7. Domains theory exceptionality and predictions

In sum, analytic compounds are protected from phonological processes by their structure as per the PSC: the domains have already been processed by phonology and thus are impermeable to alteration in following domains. This analysis has also been applied to harmony in Turkish (Ploch 1996) and Basque (Cobb 1996) compounds, with application of harmony failing to apply across the [B] domain of analytic nominal compounds [20]. What we expect to find is for the same facts regarding accent preservation in an analytic compound to hold for Owari Japanese.

[20] A fuller discussion of Analytic and Non-Analytic domains and exemplification for other languages can be found in Da Silva (1992), Gussmann & Kaye (1993), Charette (2000, 2004), Kula (2002) and Scheer (2011) among others. For a recent analysis of Japanese accent couched in Optimality Theory (Prince & Smolensky 2002) utilizing Domains, see Poppe (2012).

3.8. Owari Compound nouns revisited

Recall the data presented earlier in this article, presented again for convenience. I note that pitch accent assignment in Owari is nearly identical to that of Tokyo words (Uwano 1977, Keshikawa 1983, Terakawa 1985) and the same analytic patterning is expected in Owari Japanese. Consider the compounds in (22) with proposed domain structures.

(22) **Compounds in Owari** (Ebata 2013)

a.	[[koi][bi]] 'little finger'	[ko] 'small' + [ibi] 'finger'
b.	[[naka][ibi] 'middle finger'	[naka] 'middle +[ibi] 'finger
c.	[[yoso][iki]] 'going out'	[yoso] 'outside' + [iki] 'go'
d.	[[wata][ire]] 'cotton stuffed'	[wata] 'cotton' + [ire] 'stuff'
e.	[tsy:ri] 'rainy season'	[tsuju] 'monsoon'+ [iri] 'enter'

We can hypothesize that the above exceptions to coalescence are analytic compounds, while compounds exhibiting coalescence have become non-analytic as in [tsy:ri]. Compare [tsy:ri] to the English word 'cupboard'. The compound [tsy:ri] is likely a product of domain elision, with an original structure of *[[tsuyu][iri]], as with other non-analytic compounds discussed above. Ebata (2013) also notes that this word is slightly archaic. An analytic analysis would account for a lack of coalescence in a word such as [nakaibi] 'middle finger', with a structure of [[naka][ibi]].

Following Yoshida Y. (1999), an analytic domain analysis for the compounds above which resist coalescence also predicts that the accent of domain [B] in the compounds above takes precedence in analytic compounds of the shape [[A][B]], e.g. [[ko][ibi]]. While a full accent analysis of the dialect and all possible compounds is beyond the scope of this paper, we now examine the accent patterns found in the compounds containing [ibi] 'finger' for which comparable data is available. What is expected is that all finger names in (23) will have final accent as with the original constituent word [ibí], which appears in the [B] domain. The data shows that this is not the case. I represent the accented nucleus with an acute accent. Unaccented words represented with no acute accent. Spreading is not noted below.

(23) **Compound nouns for fingers** (Ebata 2013)

a.	[i bí]	'finger'
b.	[ó ja]	'parent'
c.	[o ja í bi]	'big finger, thumb'
d.	[ku su ri]	'medicine'
e.	[ku su ri í bi]	'ring finger'
f.	[ná ka]	'middle'
g.	[na ká i bi]	'middle finger'
h.	[hi to sa ɕi í bi]	'index finger'
i.	[ko í bi]	'little finger'

One can observe that the final accent of [ibí] is not preserved in any of the forms above. The resultant compounds either exhibit penultimate accent, as is found with verbs and adjectives in many Owari words (cf. Section 5) or antepenultimate accent, found in longer non-analytic compounds and simplex words as in Tokyo Japanese above and in Owari the words [ɕaembátake] 'vegetable garden' or [murásaki] 'purple' (Ebata 2013).

The patterns in (23) are not expected if these compounds are proposed to be analytic compounds with the domain shape of [[A][B]].

If we propose the alternate domain shape of [[A]B], this analysis also fails: accent would not be found on the [ibi] portion of the compound and the accent of domain [A] would be retained. I represent the relevant data in (24).

(24) **Owari compounds (Ebata 2013) with comparable [A] domains**

a.	[ó ja]	‘parent’	
b.	[o ja í bi]	‘big finger, thumb’	*[ó ja i bi]
c.	[ku su ri]	‘medicine’	
d.	[ku su ri í bi]	‘ring finger’	*[ku su ri i bi]
e.	[ná ka]	‘middle’	
f.	[na ká i bi]	‘middle finger’	*[ná ka i bi]

Note that for those compounds where the separate [A] domains own accent is attested in Ebata (2013), I note that the accent pattern of the [A] domain does not predict the final shape of the compound. The above compound forms are unpredictable in their accent assignment, which is a hallmark of a non-compound word. GP predicts that these compounds must be listed with their accent specified in the lexicon like morphologically simplex words.

What is crucial here is that the lack of coalescence in these forms is not predicted by their non-analytic structure. With respect to vowel coalescence these forms are clearly non-analytic. If these forms are non-analytic, they should be assessed by the phonology as one domain and coalescence should occur if synchronic. There is no prevention of coalescence predicted by the GP model. Two options are available at this point. We say the preceding domains analysis is incorrect to admit a few exceptions. Alternatively, we investigate the possible inactivity of coalescence as a synchronic phonological process and assume that all vowels which seem to be outputs of coalescence are in fact underlyingly specified as long vowels rather than vowel sequences. I now take the stance that coalescence is diachronic and long vowels ought to be specified underlyingly. I discuss issues with a synchronic analysis of elision and coalescence in verbs in Section 4 and the behavior of Owari pitch accent spread in Section 5 which points to the coalesced vowels in Owari as having a long vowel structure underlyingly.

4. Problems with verbal coalescence

I now re-examine the issue of synchronic verbal coalescence in verbal forms. A transparent analysis of synchronic coalescence in verbs is dependent upon the synchronic activity of consonant elision, which affects stems in segments {k, g, s}. I recall the relevant data in (25) below.

(25) **Consonant-final stems exhibiting coalescence**

	Non-Past	Negative	Past	Gloss
a.	aruk-u	aruk-aN	**ary:ta**	‘walk’
b.	isog-u	isog-aN	**isø:da**	‘hurry’
c.	das-u	das-aN	**dæ:ta**	‘put out’

To posit a synchronic case for coalescence in verb forms, it must be clear that the elision and assimilation processes are synchronic, with the above forms being formed through deletion of the stem-final segments {k, g, s}. This in turn gives rise to a vowel sequence, feeding coalescence. The alternatives are either a full listing of the paradigm in the lexicon or the listing of a suppletive *onbin* stem with a vowel sequence specified in the lexicon, such as /arui-/ 'walk-PAST'. The existence of a separate stem has been assumed in previous literature where synchronic *onbin* has been denied (as in Frellesvig 2010, Iitoyo, Hino and Satō 1982-1986). I note that this analysis in the Owari dialect would posit the existence of a stem with a vowel sequence, one which never surfaces and which cannot be independently verified. The existence of a lexically separate stem is possible for Owari, but the coalesced vowel would be specified underlyingly as part of the *onbin* stem e.g. /kak-/ 'write' would be associated with the *onbin* stem /kæ:-/. In the remainder of Section 4, I discuss evidence which supports lexical listing for verb forms.

4.1. The diachronic facts of *onbin*

We must first consider the age of *onbin* processes. By *onbin* I refer to the lenition and assimilation processes noted in 2.2, though other variants besides those found in Tokyo and Owari Japanese are found within the dialects. Frellesvig (2010:191-195) notes that words affected by *onbin* were widely attested by the Late Middle Japanese period (1200-1600 C.E.) and also gives evidence that *onbin* dates to the Old Japanese period, dating the birth of *onbin* to roughly before 800 C.E. *Onbin* forms begin to appear in texts through the Early Middle Japanese period (800 C.E.-1200 C.E.). The forms with *onbin* changes are widely attested in the Late Middle Japanese period (1200-1600 C.E.) in both native and foreign texts. This is especially transparent in accounts from foreign grammarians, such as those from the Portuguese missionary Rodrigues (1603-4, 1604-8). Both *onbin* variants and the original unaltered verbal stems were still in use during this period, but *onbin* forms were characteristic of speech and non-*onbin* forms characteristic of the written word. This evidenced in both prose texts and commentary by native philologists (Frellesvig 2010:194).

The *onbin* changes affected all categories of words, not just verbs. Frellesvig (2010:192) notes that the *onbin* processes were not entirely regular, unlike the automatic and regular lenition of consonants such as /p/ to /w/ or /f/ e.g. [kapu]>[kawu] 'buy-NP' and [kepu]>[kefu] 'today' found in the Early Middle Japanese. Frellesvig states that "…The *onbin* changes give rise to variant shapes of individual words or morphemes which coexisted for a considerable period…" (Frellesvig:2010:231). Importantly, I point out the fact than *onbin* were not regular. "They were non-automatic sound changes; that is to say, there were certain phonological conditions which had to be met for *onbin* to take place, but no identifiable phonological conditions under which the change **must** occur." (Frellesvig 2010:231, emphasis mine). Variation in modern dialects occurs largely due to the various forms of a Middle Japanese word which survived in certain areas and not others. Compare Eastern Japanese /kaw.ta/ [kat:a] 'buy-PAST' to Western Japanese [ko:ta] 'buy-PAST'. There is now no synchronic derivation at work, but rather variant *onbin* forms were adopted by different populations.

In addition to the age of the changes, Frellesvig (2010:230-2) claims that not all *onbin* sound changes affecting verb stems were equal. He notes a division between core *onbin* and analogical *onbin*. The core *onbin* were found in verbs, nouns and other words and affected the word medial segments {p, b, m, n, k, g}. The analogical *onbin* changes affected only verbal stems terminating in {t, r, s} as the result of analogy from the core changes preceding certain suffixes.

While I note that analogy and universality of *onbin* forms may have led speakers to generalise a verbal paradigm, there is also the issue of suffixes failing to create new *onbin* forms in more recent stages of the language. Frellesvig (2010:130-1) notes that later suffixes such as the LMJ desiderative /-(i)ta-/ did not trigger *onbin* though earlier suffixes of a similar shape such as the EMJ perfect suffix /-(i)te/ regularly trigger *onbin*.

Synchronically these changes are explained by analysing *onbin* triggering suffixes as consonant-initial. Other suffixes are analysed as vowel initial though it must be noted that consonant-initial suffixes such as /-(r)u/ for the Non-Past simply exhibit their initial consonant following Vowel-final stems and suppress this consonant when adjoined to a consonant-final stem. Why does the initial /t/ of the Gerund suffix /-te/ never exhibit such alternations? It must be stipulated that *onbin* are only triggered by a specific set of suffixes. While synchronically speakers know nothing of diachrony, this still seems to be a somewhat arbitrary analysis. I note that while a synchronic analysis of these forms is possible, the irregular *onbin* changes are not independently attested in Modern Japanese dialects and there seems to be little motivation for such a synchronic analysis.

I argue that a lack of recent *onbin* forms is due to *onbin* failing to be a synchronic process in the Late Middle Japanese (1200-1600 C.E.) period, with speakers lexicalising these forms in Modern Japanese. Once we take the step to posit full suppletion for *onbin* forms, arbitrary aspects of deriving a synchronic analysis vanish. Dialect variation found in modern Japanese dialects is not a result of synchronic *onbin* rule variation but rather simply variation in learned paradigms. Coalescence is simply another diachronic rule which applied following the Late Middle Japanese period.[21]

I support the lexical analysis in the following discussions of new verbs and experimental results which point towards a stored verbal paradigm with no productivity of *onbin* processes.

4.2. New verbs

I note that new verbs with stems from clipped loan words never form consonant-final stems affected by consonant elision. Verb stems terminating in the segments {s, k, g, b, m, n, t} are a closed class. Tsujimura & Davis (2011) offer an overview of new coinages which fall into two classes. New verbs are formed either by usage of the light verb construction /X-suru/ ‘to do X’, such as [memo-suru] ‘to make a memo’ or by addition of the same suffixes as the r-final stems to a novel stem formed from a clipped word with /r/ added. An example of the latter is the verb stem /sabor-/ ‘cut–class’ from

[21] I briefly note that evidence of /Vi/ coalescence is absent in the work of Rodrigues (1604) which roughly places coalescence occurring after the Middle Japanese period. Keshikawa (1971) and Hikosaka (1997) note that works from the early 1800’s show orthographical errors which are perhaps evidence of coalescence.

French ‘sabotage’ conjugated as /sabor-u/ ‘cut class.NP’ and /sabo-t:a/ ‘cut.class.PAST’. This occurs productively with commonly known concepts (e.g. ‘to google’ [guguru]) or coinages specific to certain groups but, Tsujimura & Davis (2011) note that the light verb construction is more common[22]. I know of no new verbs which utilise any other consonant-final stems productively. The class of consonant-final stems seems to be a closed class. Experimental evidence also shows that speakers are not able to recognize and derive novel consonant-final nonsense verb stems reliably.

4.3. Experimental evidence against synchronic derivation

Experiments by Vance (1991) utilise nonsense verbs to test Tokyo Japanese speakers on their ability to derive consonant-final verbs based on the existing paradigms. I discuss the results of Vance (1991) below. The results in fact fail to show that speakers are able to generate verb forms synchronically or even utilise analogy effectively, pointing towards lexical listing for verb paradigms.

Vance (1991) held a number of experiments (building on an earlier experiment with design flaws in Vance 1987) designed to test the ability of Tokyo Japanese speakers to conjugate novel consonant-final verb forms.

The task was a multiple choice pen-and-paper forced choice experiment with no time limit. The design disguised nonsense word stimuli in the introduction as slang words for standard verb forms (*hok-u* as a slang nonsense word for the *mi-ru* ‘see’). Subjects were given two examples of each nonsense verb with the non-past /-(r)u-/ and desiderative /-(i)tai/ suffixes. The nonsense verb stems given were /hom-/ /mur-/, /hok-/ and /kap-/ which were given associated real verbs as semantic counter parts e.g. [hoku] was given as a ‘real’ fashionable word for common verb [mir-u] ‘look-NP’ .[23] Each set of stimuli was followed by a multiple-choice task where a sentence with the verb blank was given, with the respondent choosing from three possible verb forms, one of which is analogically correct e.g. [hota], [hokutta] and [hoita] were all possible Past tense responses for the nonsense stem /hok-/ with [hoita] being the analogically correct response compared to other real verbs such as [kaita] ‘write-PAST]. The participants were tested on the conjugation of the stem in combination with the hypothetical /-(r)eba/, volitional /-(j)o:/, negative /-(a)nai/ and past tense /-(i)ta/ forms.

A separate task further analysed the ability of respondents to conjugate nonsense /r/ final stems. I do not discuss the results of the nonsense-word /kap-/ or the results of the second experiment and focus instead on the results relevant for *onbin* forms.

The results of 50 subjects, chosen at random, were selected for analysis. The results are poor for the three verbs with real counterpart paradigms (/hom-/ /hok-/ /mur-/). Only 68% of the total number responses, or 410/600, were correct among all responses.

[22] An example of the former is ‘Google’ forming the verbs /gugur-u/ ‘google.NP’ /gugu-t:a/ ‘google.PAST’. An example of the latter is specific to students at Kwansei Gakuin University , with students using the verbs [paparu] and [mamaru] to lightly refer to eating at the on-campus cafeterias Big Papa and Big Mama

[23] Consonant-final verb stems terminating in /p/ do not exist in Japanese. Vance (1991) utilised this novel verb to test the speakers’ ability to analogically create a paradigm from existing consonant-final stems. Speakers did not succeed in accurately performing this task, see Vance (1991) for details.

While the correct response percentage is above chance, it would be expected that speakers would perform more accurately if they had an active ability to conjugate consonant-final verb paradigms or if speakers actively used analogy to form new paradigms.

Respondents performed most poorly on responses to the verb stem /hok-/, with the mean correct per subject for /hok/ being 2.42/4 compared to 2.92 correct responses per subject for the stem /hom-/.The given forms for *hok*- were [hoku] 'hoku.NP' and [hokitai] 'hoku-DES' with analologically correct responses for the other forms being [hokeba] (Hypothetical), [hoko:] (Volitional), [hokanai] (Negative) and [hoita] (Past). The paradigm is clear given the non-past and desiderative priming sentences, so it is unclear why all speakers did not perform well given the lack of timing restrictions.

Vance (1991) claims that the subjects had difficulty recognising analogically correct verbal forms, and that the results tentatively support the hypothesis that regular inflectional forms are stored in the lexicon in addition to previous experiments reported in (Vance 1987). While the regular paradigm of consonant-final verb stems and a synchronic phonological analysis such as that by Davis & Tsujimura (1991) would predict the correct recognition and production of these verbs, subjects do not seem to accurately analyse nonsense stems as predicted nor do speakers conjugate them correctly and reliably in a forced choice task, even given adequate stimuli and time in which to respond. De Chene (1982), Batchelder & Ōta (2000), Klafehn (2003) and Sugaya (2011) present further evidence from nonsense word studies which support the lexical storage hypothesis for native Japanese speakers.

What the above results predict is that the paradigms of consonant-final stems are stored in the lexicon for Tokyo Japanese speakers. It is reasonable to assume that Owari speakers would show similar results. Having given evidence that forms might be lexically listed, I now discuss in the following section evidence from Owari pitch accent which supports the lexical listing of coalesced vowels. The evidence suggests that rather than vowel sequences being present underlyingly in nominal and verbal forms, long vowels are found in the lexical structures.

5. The behaviour of coalesced vowels and pitch accent

I now discuss the behaviour of coalesced vowels in relation to pitch accent spread. Accent patterns found in Tokyo and Owari are identical with the exception of the regressive spread of accent from the accented nucleus. Accent spreading is analysed by Yoshida Y. (1992, 1999) as operating regressively until the initial nucleus in Tokyo. I note that in Owari pitch spread operates until the pen-initial nucleus, yet this rule may be broken in certain contexts. This has been noted previously by Mizutani (1960) and Uwano (1977). I discuss the facts below.

5.1. The unique characteristics of Owari spreading

Mizutani (1960) observes that Tokyo accent and Owari accent are not quite identical when one considers the pitch patterns found. Consider the following data from Tokyo and Owari which focuses on nouns in plain and nominative form with three nuclei, drawing on data from Uwano (1977). Notation from Yoshida Y. (1999) is applied to differentiate Accented and Unaccented words.

(26) **Tokyo and Owari nominal accent**

a. **Tokyo nominal forms** (Uwano 1977:290)

Citation form	Noun.NOM	Gloss
ka bu to	ka bu to ga	‘helmet’
ko ko ro	ko ko ro ga	‘heart’
ka ga mi	ka ga mi ga	‘mirror’
sa ku ra	sa ku ra ga	‘cherry blossom’

b. **Owari forms** (Uwano 1977:290)

Citation form	Noun.NOM	Gloss
ka bu to	ka bu to ga	‘helmet’
ko ko ro	ko ko ro ga	‘heart’
ka ga mi	ka ga mi ga	‘mirror’
sa ku ra	sa ku ra ga	‘cherry blossom’

c. **Longer words** (Ebata 2013)

u ta ga u	‘doubt’
o ku ja mi	‘mourning’
sa ki o to to ɕi	‘three years ago’

In (26a), pitch spread occurs wherever possible until the initial nucleus in Tokyo Japanese. In (26b), we see identical Accented and Unaccented patterns represented in Owari but we find that spread does not affect the initial nor peninitial nucleus. Extending the analysis of Tokyo Japanese in Yoshida Y. (1999), I claim that there is inaccessibility for spreading of the two initial nuclei, rather than the sole initial nucleus as in Tokyo Japanese. This explains why spread is nonexistent in an Unaccented word with three nuclei such as /sakura/. Spread is exhibited only when the nominative particle /ga/ is added. Mizutani’s (1960) is correct in his intuition that Owari accent assignment is different. In a GP understanding of pitch accent utilized here, it is obvious that the difference between the dialects can be captured by a more restrictive spreading process. However, Mizutani (1960) also notes another difference: a pitch attraction rule in Owari Japanese exists which is sensitive to the occurrence of long vowels and nasal-obstruent clusters.

5.2. Pitch attraction in Owari Japanese

Mizutani (1960) then examines words with long vowel or a nasal-obstruent cluster in either the initial or pen-initial position. In these words, pitch attraction may operate and violate the ban on spreading to the initial and second nucleus. He notes this data from recordings and examines various contexts in which this process occurs. Data in Ebata (2013) supports this assertion. I discuss only long vowels below. I note that with regards to pitch attraction, both the five standard long vowels of the set {a, i, u, e, o} and the coalesced vowels {æ, ø, y} behave in an identical manner.

(27) **Data showing spread to pen-initial nucleus** (Ebata 2013)

a. o se: bo: ‘end of the year’

b. ɕi ø: ga ri ‘shell gathering/clamming’

c. e ræ: ga o ‘triumphant look’

d. o ɕo: ga ttsama ‘the new year’

(28) **Data showing spread to initial nucleus** (Ebata 2013)

a. to: mo ‘ricefield’

b. dæ: do ko ‘kitchen’

c. sy: ka ‘watermelon’

d. kø:tsu ‘this person’

e. tæ: fu: ‘typhoon’

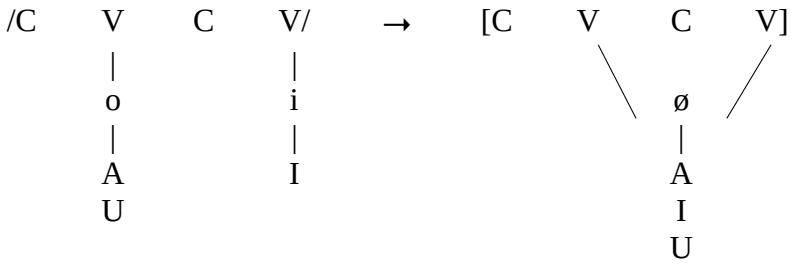
f. dʑo:dan ‘joke’

g. kʲo:kaɕo ‘textbook’

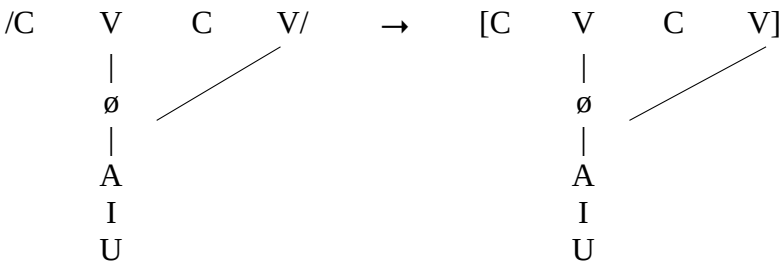
h. o: dæ: dʑi N ‘rich person’

Above, we see that the accent is attracted to the site of a long vowel.[24] Only the spread of high pitch is affected, with spread to either the initial or second nucleus. If a coalesced long vowel was underlying a nuclear sequence rather than an underlying long vowel, we would not expect coalesced vowels to behave in the same manner as the five standard long vowels. I exemplify the two possible underlying structures and their interpretation in (29).

(29) **Representation assuming underlying vowel sequence**



(30) **Representation assuming underlying long vowel with dependent V_2**



The unified behavior of standard and coalesced vowels points to the same underlying syllabic structure, whatever representational proposal is assumed. Long vowels must be specified underlyingly to explain this fact. I now examine the behavior of past tense verbal forms and pitch accent to see if this hypothesis is supported.

5.3. Verbal Accent in Owari

We now examine the pitch patterns found in Owari verbs, with particular attention given to the Past tense form. Description and data are drawn from Terakawa (1985) with further data and pitch accent drawn from Ebata (2013).

As with Tokyo verbs, Owari verbs have both Accented and Unaccented variants. Unaccented verbs are finally accented while Accented verbs receive penultimate accent in the Non-Past form as in Tokyo. I discuss only Accented verbs in this section, as Unaccented verbs have universal final accent and tell us nothing about the underlying structure of a word. Consider the data in (31) below.

[24] It should be noted that this can be paralleled with the attraction of stress in Latin or English, increasing the similarity of accent assignment in Japanese to that of the Latin stress rule, discussed by Kubozono (2008) and Kawahara (2015) in relation to Tokyo Japanese. Ebata (2013) does discuss informally that Owari is rather similar in terms of intonation to a stress language. However, unlike in Latin stress assignment, accent is not reassigned in Owari Japanese in the presence of a long vowel.

(31) **Accented Vowel final-Stems**

a. Accented verbs exhibiting penultimate accent in the Past

NP	Neg	Past	Gloss
i ki* ru	i ki* n	i ki* ta	‘live’
ya me* ru	ya me* n	ya me* ta	‘quit’
a zu ke* ru	a zu ke* n	a zu ke* ta	‘put away’

b. Accented verbs exhibiting penultimate accent in the Past

NP	Neg	Past	Gloss
a wa te* ru	a wa te* n	a wa te* ta	‘foam’
i ta me* ru	i ta me* n	i ta me* ta	‘cook/heat’
a ma e* ru	a ma e* n	a ma e* ta	‘coddle’

I note that in the above vowel-final stem data, accent is universally penultimate for Non-Past and Negative forms while Past forms show some variability, with some forms exhibiting a penultimate accent and others an antepenultimate accent. We expect to find the same pattern for Consonant-final stems where an underlying stem with a vowel sequence has previously been assumed. Consider the data in (32).

(32) **Accented C-stems**

NP	Neg	Past	Gloss
a jo* bu	a jo* ba N	a jo* n da	‘toddle’
sa so* (w)u	sa so* wa N	sa so* t ta	‘invite’
a ru* ku	a ru* ka N	a ry* : ta	‘walk’
i so* gu	i so* ga N	i sø* : da	‘hurry’
ko bo* su	ko bo* sa N	ko bø* : ta	‘spill’

In the non-past and negative forms, we find universal penultimate accent. In Past tense forms, we find universal accent on the antepenultimate nucleus with no penultimate

accent found in any of the verbal forms attested in Ebata (2013). While this is no surprise for those forms with a geminate or nasal consonant, it is unexpected for any synchronic analysis of coalescence.[25] If coalescence is a synchronic surface phenomenon, accent should theoretically also be found on the penultimate or antepenultimate nucleus as with verb-final stems, which is not attested. We must then claim that this nucleus is in the same position as those found in typical long vowel constructions: this nucleus is not an appropriate accent site. The same result is found whether one applies a GP analysis (Yoshida Y. 1999) or a Degenerate Mora analysis (Labrune 2012) of the above long vowels: it is clear that there is an underlying long vowel structure for coalesced vowels based on patterning with the standard long vowels in terms of pitch accent spread and a lack of ability to receive an accent on the second half of a long vowel (or the second nucleus in a sequence) in past tense verbal forms.

6. Conclusion

I note that further investigation is necessary to confirm the nature of coalesced vowels and their structure specified in Owari Japanese which is beyond the scope of this article. Forthcoming work will examine the dialect further and also examine the nature of coalescence in Niigata and Akita Japanese while discussing the typology and diachrony of coalescence further. Compounds and verbal pitch accent assignment also merits further investigation when field trips are feasible. For now, I note that the above discussion has given evidence to support the idea that coalescence is not a synchronically active process. In contrast to previous work, I propose that the dialect of Owari has diverged and the underlying as well as surface vowel inventory have been augmented in comparison to that found in Tokyo.

References

Batchelder, E-O. & Ōta, A. (2000) Rule vs. rote in Japanese verb inflection. In Melby A.K. & Lommel A.R. (ed.) Fullerton. *LACUS Vol.* 26. Presented at the 26th Linguistic Association of Canada and the United States Forum, Edmonton, Alberta, Canada, August 3-7, 1999. pp. 39-50.

Backley, P. (2011) *An Introduction to Element Theory*. Edinburgh: Edinburgh University Press.

Bromberger, S. & Halle, M. (1989) Why Phonology is different. *Linguistic Inquiry* 20, pp. 51-70.

Charette, M. (1990) Licence to Govern. *Phonology* 7, pp. 233–253.

Charette, M. (1991) *Conditions on phonological government*. Cambridge: Cambridge University Press.

Charette, M. (2000) When p-licensing fails: the final high vowels of Turkish. *SOAS Working Papers in Linguistics* 10, pp. 3-18.

Charette, M. (2004) Defining the structure of Turkish words. *SOAS Working Papers in Linguistics* 13, pp. 49-79.

[25] I lack space to discuss the proposals for representation of these words. From a GP point of view, Yoshida Y. (1999) proposes that the nucleus in a long vowel, nasal-obstruent cluster and geminate is governed and fails to project, thus unable to support accent assignment. A foot in GP is understood to be two nuclei in a licensing relation at the nuclear projection. Labrune (2012) alternatively analyses these structures as being composed of degenerate moras within a foot. See also Kawahara (2015) who assumes that the syllable plays a role in accent assignment shift to the antepenultimate mora.

Cobb, M. (1996) Clitics in Basque: a phonological approach. *SOAS Working Papers in Linguistics and Phonetics* 6, pp. 6-39

Da Silva, T.C.A. (1992) *Nuclear Phenomena in Brazilian Portuguese.* SOAS University of London: PhD Thesis.

Davis S. & Tsujimura N. (1991) An Autosegmental account of Japanese verbal conjugation. *Journal of Japanese Linguistics* 13, pp. 91-116.

De Chene, B. (1982) The segmentation of Japanese verbs: experimental evidence. *Papers in Japanese Linguistics* 8, pp. 29-61.

Ebata Y. (2013) Owari dictionary. In Ebata Y. (ed.) *Aichi-ken no kotoba*. Tōkyō: Meiji Shoin.

Frellesvig, B. (1995) *A History of the Japanese language.* Cambridge: Cambridge University Press.

Frellesvig, B. (2010) *A Case study in diachronic phonology: the Japanese Onbin Sound Changes*. Aarhus: Aarhus University Press.

Gussmann, E. & Kaye, J.D. (1993) Polish notes from a Dubrovnik café. *SOAS Working Papers in Linguistics and Phonetics* 3, pp. 427-462.

Harris, J. (1994) *English sound structure*. Oxford: Blackwell.

Harris, J. (2005) Vowel reduction as information loss. In P. Carr, J. Durand & C. J. Ewen (eds.) *Headhood, elements, specification and contrastivity*. Amsterdam: Benjamins. pp. 119-132.

Hikosaka, Y. (1997) *Owari kinpen o shu to suru kinseiki hōgen no kenkyū***.** Ōsaka: Izumi Shoin.

Hikosaka, Y. (2014) Kansai to Kantō, nidaihōgen no [aida]. [Kansai and Kantō, between the two large dialects.] *Nihongogaku* 33.1.

Iitoyo, K., S. Hino and R. Satō (eds.) (1982-1986) *Kōza Hōgengaku [The Dialectology Series]*. Tōkyō: Kokusho Kankōkai.

Itō, J. & Mester, A. (1986) The phonology of voicing in Japanese: theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17.1, pp. 49-73.

Itō, J. & Mester, A. (2003). *Japanese Morphophonemics: Markedness and Word Structure*. MIT Press Linguistic Inquiry Monograph Series 41. Cambridge, MA: MIT Press.

Katō, M. (1977). Hōgen kukakuron. [Theories of dialect division.] in Ōno & Shibata (1977b). pp. 41-82.

Kawahara, S. (2015) The phonology of Japanese accent. In Haruo Kubozono (ed.) *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology*. Berlin: De Gruyter Mouton. pp. 445-492.

Kaye, J.D. (1990) "Coda" Licensing. *Phonology* 7, pp. 301–330.

Kaye, J.D. (1992) On the interaction of theories of Lexical Phonology and theories of phonological phenomena. In: Dressler, U., Luschützky H. Pfeiffer O. & Rennison J. (eds.) *Phonologica 1988*, pp.141-155. Cambridge: Cambridge University Press.

Kaye, J.D. (1995) Derivations and Interfaces. In Durand J. & Katamba J. (eds.) *Frontiers of Phonology*, pp. 289-332. London & New York: Longman.

Kaye, J.D., J. Lowenstamm, and J-R. Vergnaud. (1985) The Internal Structure of Phonological Elements: A Theory of Charm and Government, *Phonology Yearbook 2*, pp. 305-328.

Kaye, J.D., J. Lowenstamm, and J-R. Vergnaud. (1990) Constituent Structure and Government in Phonology. *Phonology* 7, pp. 193-231.

Kean, M-L. (1974) The Strict cycle in phonology. *Linguistic Inquiry* 5.2, pp. 179-203.

Keshikawa, R. (1971) *Nagoya hōgen no kenkyū. [Research on the Nagoya dialect]*. Nagoya: Taibundō

Keshikawa, R. (1983) Aichi-ken no hōgen. [The dialects of Aichi Prefecture]. In Iitoyo et al. (1983) Vol. 6, pp. 207-241.

Kindaichi, H. (1977) Akusento no bunseki to hensen. [Accent analysis and development.] In Ōno & Shibata (1977b). pp. 129-180.

Kiparsky, P. (1982) Lexical Morphology and Phonology. In Yang I.S. (ed) *Linguistics in the morning calm: selected papers from SICOL-1981 Vol. 1*. Seoul: Hanshin Publishing Company. pp. 3-91.

Kiparsky, P. (1985) Some consequences of Lexical Phonology. *Phonology Yearbook* 2. pp. 85-138.

Kubozono H. (2008) Japanese Accent. In Miyagawa S. and Saito M. (eds.) *The Oxford Handbook of Japanese Linguistics*. Oxford: Oxford University Press. pp.165-191

Kula, N.C. (2002) *The Phonology of Verbal Derivation in Bemba*. PhD dissertation, Universiteit Leiden, LOT dissertation series 65.

Labrune, L. (2012) *The Phonology of Japanese*. Oxford: Oxford University Press.

Lowenstamm, J. (1996) CV as the Only Syllable Type. In Durand, J. & Laks, B. (eds.) *Current Trends in Phonology Models and Methods*. Salford: European Studies Research Institute, University of Salford. pp. 419-442.

Mizutani, O. (1960) Report on One Characteristic of Accent of the Nagoya Dialect, Part Two. *Nihon onsē gakkai kaihō [Bulletin of the Phonetic Society of Japan]* 102, pp. 8-10.

Nasukawa, K. (2000) *A Unified Approach to Nasality and Voicing*. PhD Thesis. University College London.

Nasukawa, K. (2010) No consonant-final stems in Japanese verb morphology. *Lingua* 120, pp. 2336-2352.

National Institute for Japanese Language & Linguistics. (1968) *Aichi-ken Komaki-shi Fujishima Hōgen. [The dialect of Fujishima, Komaki City, Aichi Prefecture]*. http://www.ninjal.ac.jp/publication/catalogue/hogenrokuon_siryo/ Accessed 28th of May, 2015.

National Institute for Japanese Language & Linguistics. (1989-2006) *Hōgen bunpō zenkoku chizu. [Grammar Atlas of Japan]*. http://www2.ninjal.ac.jp/hogen/dp/gaj-pdf/gaj-pdf_index.html Accessed 28th of May, 2015.

Nishiyama, K. (2010) Penultimate accent in Japanese predicates and the verb-noun distinction. *Lingua* 120, pp. 2353-2366.

Ōno S. & Shibata T., eds. (1977a) *Iwanami Kōza Nihongo 5: On'in. [Iwanami Japanese Series 5: Phonology]*. Tōkyō: Iwanami Shoten.

Ōno S. & Shibata T., eds. (1977b) *Iwanami Kōza Nihongo 11: Hōgen. [Iwanami Japanese Series 11: Dialects]*. Tōkyō: Iwanami Shoten.

Ploch, S. (1996) The role of parsing. *SOAS Working Papers in Linguistics and Phonetics* 6, pg. 76-105.

Poppe, C. (2012) Accent and domain structure in Japanese: generalizing across lexical entries. *Phonological Studies* 15, pp.43-50.

Prince, A. & Smolensky, P. (2002) *Optimality Theory: Constraint Interaction in Generative Grammar.* http://roa.rutgers.edu/files/537-0802/537-0802-PRINCE-0-0.PDF Accessed 29th of May, 2015.

Scheer, T. (2011) *A Guide to Morphosyntax-Phonology Interface Theories.* Berlin: Mouton de Gruyter.

Shibatani, M. (1990) *The Languages of Japan*. Cambridge: Cambridge University Press.

Sugaya, N. (2011) L2 Learners' Outperformance of Native Japanese: Evidence from Nonce Verb and Real Verb tests. In Plonsky L. & Schierloh M. eds. *Selected Proceedings of the 2009 Second Language Research Forum*. Somerville, MA: Cascadilla Proceedings Project. pp.142-151.

Terakawa, M. (1985) Nishikasugai District. In Aichi-ken Kyōiku Iinkai. (ed.) *Aichi no kotoba : Aichi-ken hōgen kinkyū chōsa hōkokusho.* Nagoya: Aichi-ken Kyōiku Iinkai.

Tsujimura N. & Davis, S. (2011) A Construction Approach to Innovative Verbs in Japanese. *Cognitive Linguistics* 22 (4), pp. 797-823.

Uwano Z. (1977) Nihongo no akusento [Japanese Accent.]. In Ōno S. & Shibata T. (1977a). pp. 281-321.

Uwano Z., Masao, A. Kazuo K. & Motoei, S. (1989) On'in sōran [Survey of phonology]. In Tokugawa M. (ed.) *Nihon hōgen dai jiten [Encyclopedia of Japanese dialects.]* Tōkyō: Shogakukan. pp. 1-77.

Vance, T. (1987) *An Introduction to Japanese Phonology*. Albany NY: SUNY Press.

Vance, T. (1991) A New Experimental Study of Japanese Verb Morphology. *Journal of Japanese Linguistics* 13, pp. 145-166

Yamada T. & Niwa K., eds. (1989) *Aichi-ken no hōgen. [The dialect of Aichi Prefecture]*. Nagoya: Bunkazai Tosho Kankōkai.

Yoshida, S. (1996) *Phonological government in Japanese*. Australian National University, Canberra.

Yoshida, Y. (1992) Pitch accent and tone harmony in Japanese. *SOAS Working Papers in Linguistics and Phonetics* 2, pp. 339–364.

Yoshida Y. (1999) *On pitch accent phenomena in Standard Japanese*. The Hague: Holland Academic Graphics. Published version of 1995 PhD Thesis, SOAS University of London.

Youngberg, C. (2013) Vocalic Coalescence in Owari Japanese. *SOAS Working Papers in Linguistics* 16.

Syntax and Semantics

SOAS Working Papers in Linguistics Vol. 17 (2015): 35-48

Constituent order in the Tibetan noun phrase

Edward Garrett and Nathan W. Hill
eg15@soas.ac.uk and nh36@soas.ac.uk

1. Introduction

The majority of Tibetan grammars lack explicit discussion of constituent order inside the Tibetan noun phrase (e.g. Schwieger 2006). Beyer's grammar is an exception (1992: 204-233). He distinguishes seven positions in a noun phrase: nominal, determiner, reflexive, numeral, plural, totalizer, and selector. These terms he defines semantically, e.g. "[s]electors, unlike plurals, do not specify simply that there is more than one entity referred to; instead, selectors specify what we can call the RANGE of entities referred to in the set denoted by the nominal" (1992: 232 emphasis in original). Because his analysis is illustrated anecdotally it is often difficult to know how Beyer would analyze a particular Tibetan NP. In some details Beyer's analysis is clearly wrong. For example, he treats *dag* as a 'plural' but *rnams* as a 'selector'.[1] In support of this analysis Beyer claims that *rta dag rnams* 'horses' is a possible order but not **rta rnams dag* 'horses' (1992: 205). However, Michael Hahn draws attention to the occurrence of both the orders *rnams dag* and *dag rnams* in the Tibetan translation of the *Buddhacarita* (2003[1978]). Thus, Beyer's syntactic argument for placing these two words in distinct part-of-speech categories does not hold. This case compels one to fear that Beyer arrived at his analysis impressionistically rather than empirically. A part-of-speech tagged corpus of Tibetan texts provides a convenient way to empirically investigate Tibetan noun phrase structure.

2. The Tibetan corpus and tag set

This paper gives a schematic presentation of the order of constituents in the Tibetan noun phrase, as revealed by an investigation of a part-of-speech tagged corpus of Tibetan texts. The corpus was produced by a project based at the School of Oriental and African Studies, University of London,[2] taking advantage of the profusion of digitised Tibetan texts that are now available online.[3]

The corpus can be divided into two parts, the reference corpus and the extended corpus. The reference corpus includes only highly secure analyses, namely those materials that have been hand-tagged and checked by a human being. By contrast, the extended corpus includes additional materials that have been tagged automatically by a computer. On the basis of insights gained by studying the reference corpus, we have created software that assigns a part-of-speech tag to each word in an unanalysed Tibetan text, after first segmenting the text into words. Garrett et al. (2014) describes Version 1.0 of the rule-based part-of-speech tagger, designed for Classical Tibetan materials.

---

[1] We tag both *dag* and *rnams* as [d.plural].

[2] The project 'Tibetan in Digital Communication' is funded by the U.K's Arts and Humanities Research Council.

[3] Available digital corpora include the entire Derge Kanjur (www.thlib.org/encyclopedias/literary/canons/kt/catalog.php#cat=d on 10 June 2014) and over 60 Dunhuang documents (otdo.aa.tufs.ac.jp on 10 June 2014).

For a complete description of the tag set used in the corpus, including a discussion of nominal heads and case markers, see Garrett et al. (2015). The following list focuses only on the part-of-speech categories for modifiers that occur within the noun phrase.

[adj] Adjectives: Those words that can occur in attributive position and that are not verbal nouns, we define as adjectives. Tibetan adjectives are almost always multisyllablic, with the paragon varieties ending in *-po, -mo, -pa, -ma, -can, -ldan*, and *-med*. Etymologically *-can, -ldan* 'with', and *-med* 'without' are verbs, but when they occur in forms that are functioning nominally they cannot be analyzed as verbs. So, we treat them together with the preceding syllable as an adjective. Although words such as *nag* 'black', *gsar* 'new', and *che* 'big' are frequently treated as adjectives for pedagogical purposes, a single syllable in predicate position before verbal suffixes is a verb. These words may appear to occur attributively, but we see this as an instance of the formation of compounds. For example, the compound *blon-chen* 'prime minister' contrasts with noun and adjective pair *blon-po chen-po* 'great minister'.

[d.dem] Demonstratives: This tag is used for the demonstratives *ḥdi* 'this' and *de* 'that'.

[d.det] Determiner: The most frequent determiner is *gźan* 'other'. In addition, we identify *ya-re* 'each one (of two)' as a determiner on the basis of the following sentence: *Brgya-byin daṅ Tshaṅs-paḥi rgyal-pos lag-pa ya-re nas zin te* 'The kings Indra and Brahma each took him by one of his hands'. We reckon *ḥbaḥ* 'sole' as a determiner on the basis of sentences such as *rus-pa daṅ khrag ḥbaḥ źig gis sa rtsog-rtsog ltar ḥdug-pa mthoṅ* 'They saw the ground besmirched with only bone and blood'.

[d.emph] Emphasis: This category was initially created for *ñid* in phrases such as *rgyal-po ñid* 'that very king' or *lus ñid* 'this body'. This syntactic use of *ñid* must be distinguished from the use in Buddhist terminology of *-ñid* inside words, e.g. *stoṅ-pa-ñid* 'emptiness'. Apart from *ñid,* we have categorized *kho-na* 'the very, same' and *re-re* 'each' as emphatics. This use of *kho-na* should not be confused with its function as a third person pronoun in Old Tibetan.

[d.indef] Indefinite: This category is used for the allomorphs of the indefinite marker *cig, źig,* and *śig* as in *pho-ña cig* 'a messenger'.

[d.plural] Plurals: The plural markers *rnams, dag, kun, thams-cad, ḥo-cog* (and its variants) and *tsho* are tagged as their own category 'plural'. However, plural pronouns (*bdag-cag, khyed-cag, ḥu-bu-cag*) are treated as one word. The plural marker *-cag* is not removed because to do so would result in pronominal stems which are not mutually comparable (viz. *bdag* is a singular pronoun, *khyed* a plural pronoun, and *ḥu-bu* has no independent life outside of *ḥu-bu-cag*).

[num.card] Cardinal numbers: We distinguish cardinals (*gcig* 'one', *gñis* 'two', *gsum* 'three', etc.) and ordinals (*daṅ-po* 'first', *gñis-pa* 'second', *gsum-pa* 'third', etc.). Other derivatives of numerals are treated according to their respective syntax, thus *gcig-pa* 'sole' is an adjective, *gñi-ga* 'both' is an indefinite pronoun, etc. In higher numbers each component digit is tagged separately.

[num.ord] Ordinal numbers: As just discussed, in numbers we distinguish cardinals (*gcig* 'one', *gñis* 'two', *gsum* 'three', etc.) and ordinals (*daṅ-po* 'first', *gñis-pa* 'second', *gsum-pa* 'third', etc.). Some adjectives can be distinguished from ordinal numbers only in context. For example, in the phrase *rdo-rje rtse-lṅa-pa* 'a five-pronged vajra' the noun phrase syntax and the overall meaning of the passage dictates that *lṅa-pa* is part of the word *rtse-lṅa-pa* 'five-pronged' rather than a word 'fifth'.

To explore the relative order of constituents within the Tibetan noun phrase, we queried the corpus for sequences of words with particular part-of-speech categories, irrespective of the orthographic content of the words themselves. In order to broaden our investigation beyond the patterns noticed by human taggers, we used the machine-tagged extended corpus as our dataset. By setting the 'exclude ambiguous tags' option in the search interface, we excluded from our search results all words of insecure part-of-speech assignment, that is, those cases where, unsure what the tag for a word should be, the rule tagger gives an ambiguous answer.[4] The resulting 'shingles' or sequences of parts-of-speech were then analysed and compared.[5]

3. Summary of Findings

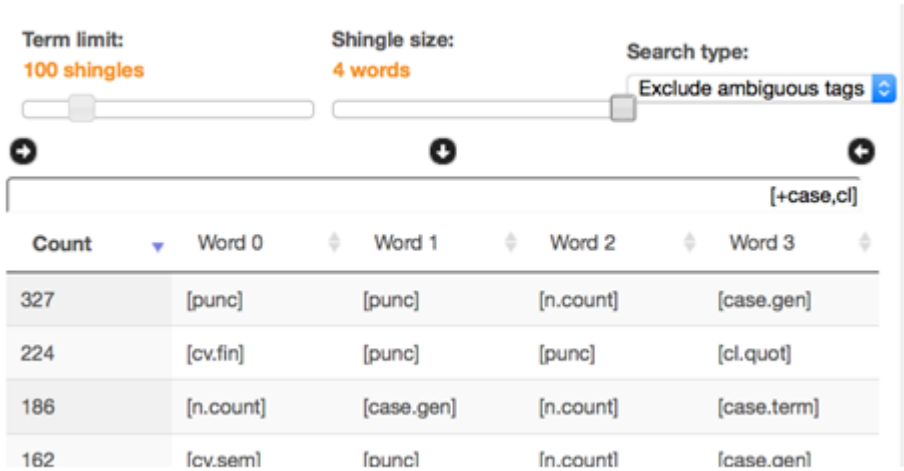
There is a convenient way to find the attested patterns of three element noun phrases (and shorter NPs) by specifying that the fourth element of the shingles window includes those elements that are known to end noun phrases. The elements we chose to require as the fourth word in the search window are case markers [case.xxx] and clitics [cl.xxx]. Case markers end a noun phrase. Clitics come even after case markers. The reason to specify clitics as one option as the fourth constituent is that a noun phrase in the absolutive case (i.e. is not followed by an overt case marker) that is followed by a clitic will be specified by such a search. The shingles interface allows search from the right. We put [+case,cl] into the search window, searching from the right, with a window of four words and returning 500 shingles (cf. Illustration 1). This search captures a series of four words in which the fourth word is a case marker or a clitic. From the results of such a search it is necessary to disregard cases, converbs, clitics, negation, finite verbs,

[4] This has disadvantages such as precluding exploring the relative frequency of the orders *rnams dag* and *dag rnams* because the sequence of letters *dag* is ambiguous between the plural marker and a verb 'be pure'. Nonetheless, the corpus permits such investigations, which we shall benefit from returning to with an improved understanding of Tibetan noun phrase structure on the basis of words of unambiguous part-of-speech.

[5] Normally, shingle search returns POS tags only. By selecting the "show word forms" option, word forms are returned instead. This is important in order to see the specific words in context that underlie the generalized part-of-speech patterns. Preceding a POS item with "+" indicates that only a partial match is required. [+v] will match every tag that starts with "v", and [+v,n.v] will match every tag that starts with either "v" or "n.v". Note that both [+v,+n.v] and [v,n.v] are ill-formed searches.

and punctuation occurring among the first three constituents because these elements are not part of a noun phrase. For example, the most frequent pattern is [punc] [punc] [n.count] [case.gen], which for the purposes of analyzing noun phrase structure counts simply as [n.count].

Illustration 1: rightward rightward search for [+case_cl] (1 January 2015)



The process of disregarding irrelevant entries, combining compatible patterns, and supplementing with additional searches is described in the appendix. Table 1 presents the overall patterns discovered.

Table 1: Summary of Tibetan noun phrase structures

	[n.count]	[adj]	[d.det]	[d.plural]	[d.dem]	[d.emph]
	[n.count]	[adj]	[d.det]	[d.plural]	[d.plural]	[d.emph]
	[n.count]	[adj]	[d.det]	[d.indef]		
	[n.count]	[adj]	[num.card]	[d.indef]		
	[n.count]	[num.card]	[num.card]	[num.card]	[num.ord]	
[n.v.invar]	[n.count]	[n.prop]	[n.count]	[d.dem]		
	[p.pers]	[n.count]				
	[p.pers]	[p.refl]	[d.plural]			

These patterns fall into three groups: noun phrases headed by common nouns, noun phrases headed by proper nouns, and noun phrases headed by personal pronouns.

Common nouns: The first five rows of Table 1 give patterns headed by common nouns. One might be tempted to further combine the patterns [d.plural] [d.dem] [d.emph] and [d.plural] [d.plural] [d.emph] as [d.plural] [d.plural] [d.dem] [d.emph], but this is not possible because [d.plural] [d.plural] [d.dem] is not an attested pattern. If we look in more detail at the pattern [d.plural] [d.plural] it becomes clear that certain words sit more comfortably as the first of two plurals whereas others sit more comfortably as the second of two plurals. In particular, the patterns *rnams thams-cad*, *rnams kun*, and

rnams sna-tshogs suggest that the plural marker *rnams* is of quite a different nature than these others words, which imply something about quantification. Thus, Beyer may be right to divide 'plurals' from 'selectors' and to put *thams-cad* and *kun* in the latter category (1992: 232). If so, *rnams* should count as a 'plural' and not a 'selector' as Beyer has it. The rare patterns *thams-cad kun* and *sna-tshogs kun* suggest that even a category of 'selectors' would not be homogenous.

A surprising omission in Table 1 is the pattern [d.dem] [d.plural], which occurs in the corpus 399 times, much more frequently than the 72 attestations of the pattern [d.plural] [d.dem]. The reason that this pattern was not included in Table 1 is that it did not combine felicitously with the other patterns. Noting that 154 of the occurrences of pattern [d.dem] [d.plural] occur after the *śad* punctuation mark, it seems likely that this is a pattern typical of pronominal use of *de* and *ḥdi*.

The first four rows of Table 1 make clear that noun phrases are either marked as indefinite, with [d.indef], or they are marked as definite, with a plural marker [d.plural], a demonstrative [d.dem] or both. Table 1 might also hasten the conclusion that that [d.emph] is only compatible with definite marked noun phrases, but the absence of the pattern [d.indef] [d.emph] may be an accidental gap.

As a final remark about noun phrases headed by common nouns, the patterns [adj] [num.ord] and [adj] [num.card] [num.ord] are not attested. This potentially suggests that ordinal numbers are syntactically treated the same as adjectives.

Proper nouns: The sixth row of Table 1 gives the one overarching pattern of noun phrases headed by a proper noun, viz. [n.v.invar] [n.count] [n.prop] [n.count] [d.dem]. Although though modifiers typically follow their heads in Tibetan noun phrases, as seen in this pattern, modifiers both precede and follow proper nouns. The initial [n.v.invar] is typically *ḥphags-pa* 'exhaulted' or *dam-pa* 'sacred'. The remaining elements or the pattern are illustrated by the phrases *pha-rgan Mar-pa lo-tstsha* 'old father, Marpa the translator' and *Mar-pa lo-tstsha ḥdi* 'this Marpa the translator'.

Personal pronouns: The final two rows of Table 1 give the two patterns of noun phrases headed by personal pronouns: [p.pers] [n.count] as seen in *ṅed ma-smad* 'we, mother and children' and [p.pers] [p.refl] [d.plural] as seen in *ṅed raṅ tsho* 'we'.

Although this study has merely touched upon the mysteries of the Tibetan noun phrase a few tentative conclusions are possible. First, the words that we tag [d.plural] can be further sub-classified with *rnams* in one category and *thams-cad, kun,* and *sna-tshogs* in another category. Second, proper nouns are peculiar in allowing preceding modifiers. Third, adjectives and ordinals appear to be incompatible.

References

Beyer, S. (1992) *The Classical Tibetan language*. New York: State University of New York.

Garrett, E., N. W. Hill, and A. Zadoks (2014) A Rule-based Part-of-speech Tagger for Classical Tibetan. *Himalayan Linguistics* 13.1, 9-57.

Garrett, E., N. W. Hill, A. Kilgarriff, R. Vadlapudi, and A. Zadoks (2015) The

contribution of corpus linguistics to lexicography and the future of Tibetan dictionaries. *Revue d'Etudes Tibétaines* 32, 51-86.

Hahn, M. (2003[1978]) On the function and origin of the particle dag. *Schluessel zum Lehrbuch der klassischen tibetischen Schriftsprache und Beitrage zur tibetischen Wortkunde* (*Miscellanea etymologica tibetica I – VI*). Marburg: Indica et Tibetica Verlag, pp. 95-104.

4. Appendix 1: early stages in the analysis

We isolated the patterns presented in Table 1 (above) by querying the 'shingles' search interface of the corpus. However, the limits of this interface and the presence of patterns including more than one noun phrase in the search results meant that a certain amount of cleanup was necessary in order to isolate the patterns that drew our attention. We put [+case,cl] into the search window, searching from the right, with a window of four words and returning 500 shingles. This search captures a series of four words in which the fourth word is a case marker or a clitic. From the results of such a search it is necessary to disregard cases, converbs, clitics, negation, finite verbs, and punctuation, occurring among the first three constituents because these elements are not part of a noun phrase. For example, the most frequent pattern is [punc] [punc] [n.count] [case.gen], which for the purposes of analyzing noun phrase structure counts simply as [n.count]. When such entities are disregarded and the frequency calculations are recalculated, the results achieved are what can be seen in Table 5 in appendix 2. These data still contain many patterns that are not legitimate noun phrases. We propose to simplify this picture by disregarding two types of phenomena.

In the first case, in a sequence [n.count] [n.count] we disregard the first noun. The pattern [n.count] [n.count] is a very frequent pattern, occurring 348 times in the corpus. However, according to the protocols of the project this sequence of tags is only possible when two nouns are in apposition or when they belong to two different noun phrases. For example, the most common sequence of words that match the pattern [n.count] [n.count] [+case,cl] is *bcom-ldan-ḥdas rgyal-po ḥi*, occurring three times in the corpus. In this example *bcom-ldan-ḥdas* 'Bhagavan' and *rgyal-po* 'king' belong to two different noun phrases and *bcom-ldan-ḥdas* should be understood as in the absolutive case. A scroll through the results of the search [n.count] [n.count] [+case,cl] does not yield any obvious examples of apposition. Consequently, it is safe to assume that most of the 348 examples of [n.count] [n.count] in the corpus are best taken as examples of the noun phrase structure that consists of a single noun. Put more generally, in those patterns in Table 5 that yield a sequence of two nouns, the first noun can be omitted from analysis, similar to how we ignored nouns in position one followed by a case marker in position two.[6]

In the second case, we disregard all of those noun phrase structures in which the third element is a verbal noun [n.v.xxx]. Although the distinction between an adjective and a nominalized intransitive verb is often somewhat arbitrary, because our project does not encode a distinction between transitive and intransitive verbs, allowing [n.v.xxx] as the right constituent of a noun phrase includes many structures that are not wanted. The

[6]As an aside, because it is not strictly speaking incorrect to analyze an apposition of the structure [n.count] [n.count] – for example *yab rgyal-po* 'the father, the king' as an example of the noun phrase structure [n.count], there is in fact no problem at all with disregarding the first noun in such cases.

pattern [n.count] [n.v.invar] may seem like an ordinary noun phrase in a case like *chos dam-pa* ‘sacred dharma’ (cf. *rgyal-po chen-po* [n.count] [adj] ‘great king’), but the pattern [n.count] [n.v.past] is not intuitively amendable to analysis as a noun phrase in a pattern like (*rgyal-po s*) *dgra bsad-pa* ‘(the king) killed the enemy’. Thus, patterns of the type [n.count] [n.v.xxx] are best addressed within the context of an overall investigation of the relationship of noun phrase structure with verb nominalization and subordination structures. These are issues that deserve attention, but are beyond the scope of this study.

Disregarding the first [n.count] in a sequence of two and all [n.v.xxx] in third position results in the noun phrase structures seen in Table 6 (appendix 2). Some of these patterns are unlikely to be actual noun phrase patterns. Specifically, there are six patterns that have [n.count] in the third position (cf. Table 2).

Table 2: noun phrase patterns with [n.count] in position three

32	[n.count]	[num.card]	[n.count]
22	[n.count]	[d.plural]	[n.count]
21	[n.count]	[adj]	[n.count]
15	[adj]	[d.plural]	[n.count]
10	[n.count]	[n.prop]	[n.count]

Since a noun often heads a noun phrase, these patterns look as if they may have more than one head. We have investigated the specific examples underlying each of these patterns, and many have a second constituent in the absolutive case, e.g. *lan gcig yul du* ‘once, in a land’, *bya rnams nam-mkhaḥ la* ‘birds, to the sky’, *rgyal-po chen-po nad kyis* ‘the great king, by illness’, *maṅ-po rnams saṅs-rgyas kyis* ‘many, by the buddha’.

The final pattern [n.count] [n.prop] [n.count] includes phrases that consist of two noun phrases, such as *dam-chos Rgya-gar yul nas* ‘sacred dharma from the land [of] India’ but also phrases that are a single noun phrase such as *pha-rgan Mar-pa lo-tstsha daṅ* ‘Old father, Marpa, the translator, and’. Thus, we exclude from further consideration the first four patterns in Table 2, but maintain the fifth.

There are seven patterns that have [n.count] in the second position. These patterns seem *a priori* unlikely to be capturing genuine noun phrases, because an [n.count] typically heads a noun phrase and is followed by, rather than proceeded by, dependent constituents.

Table 3: noun phrase patterns with [n.count] in position two

12 [n.v.fut]	[n.count]	
9 [d.plural]	[n.count]	[d.plural]
8 [num.ord]	[n.count]	
8 [n.v.past]	[n.count]	[num.card]
10 [n.prop]	[n.count]	
10 [n.v.invar]	[n.count]	
9 [p.pers]	[n.count]	

Of the patterns in Table 3 the first three appear not to permit legitimate noun phrases. We provide a common example of each pattern: *bya-ba stan las* ‘called, from the seat’, *rnams yul-mi rnams daṅ* ‘plr., neighbors, and’, *daṅ-po theg-chen gyi* ‘first, of the Mahāyāna’.[7] The examples underlying the fourth pattern [n.v.past] [n.count] [num.card] often do not consist of single noun phrases, e.g. *ḥkhrugs-pa lan gñis su* ‘disturbed, a second time’. However, some cases, such as *bsdus-pa bam-po gsum daṅ* ‘collected, as (?) section three’ are perhaps amenable to analysis as a single noun phrase. Nonetheless, because such a noun phrase involves a nominalized verb that is governing additional constituents to the left, outside of the shingle frame, such examples are not considered here. The final three patterns in Table 3 do include cases of single noun phrases: *Rgya-gar yul du* ‘in the land, India’, *ḥphags-pa raṅ-saṅs-rgyas las* ‘to the noble Pratyekabuddha’, *ṅed ma-smad kyi* ‘of us, mother and children’. However, the patterns [n.v.invar] [n.count] and [p.pers] [n.count] also include examples that must be analyzed as two separate noun phrases: *sogs-pa saṅs-rgyas kyis* ‘etc., the buddha’, *khyod stan la* ‘you, on the seat’. Consequently, these patterns as such cannot be used to detect noun phrase boundaries.

A final pattern that came up in the search, but does not select a single noun phrase is [p.pers] [p.pers]. In all examples the two pronouns are different grammatical persons, e.g. *khyod ṅa la* ‘thou, to me’ or *bdag-cag khyod la* ‘we, to thou’, and thus obviously are not part of the same noun phrase.[8]

Table 7 (appendix 2) omits those patterns that appear incompatible with a single noun phrase, and also omits single word noun phrases; such words reveal nothing about noun phrase structure. One formalization of Tibetan noun phrase structure would simply consist of the structures present in Table 7 (appendix 2), however this presentation can be simplified, by presuming that trailing members of a noun phrase are optional; with this possibility in mind several of the two and three constituent patterns are combinable into one or more overarching four constituent pattern. For example, [n.count] [adj], [n.count] [d.plural], [n.count] [d.dem], [n.count] [adj] [d.plural], and [n.count] [d.plural] [d.dem] may all be conceptualized as sub-cases of the four constituent patter [n.count] [adj] [d.plural] [d.dem]. If all of the possibilities discussed so far are combined in this way the result is Table 4.

[7] It is so difficult to imagine the pattern [d.plural] [n.count] [d.plural] even not partially including two noun phrases; this pattern could thus be used to identify absolutive case marking after the first plural.

[8] The pattern [p.pers] [p.pers] could be used to isolate some noun phrases marked in the absolutive case.

Table 4: Summary of Tibetan noun phrase structures

	[n.count]	[adj]	[d.det]	[d.plural]	[d.dem]	[n.rel]
	[n.count]	[adj]	[d.det]	[d.plural]	[d.dem]	[d.emph]
	[n.count]	[adj]	[d.det]	[d.plural]	[d.plural]	[d.emph]
	[n.count]	[adj]	[d.det]	[d.indef]		
	[n.count]	[adj]	[num.card]	[d.indef]		
	[n.count]	[num.card]	[num.card]	[num.card]	[num.ord]	
[n.v.invar]	[n.count]	[n.prop]	[n.count]	[d.dem]		
	[n.count]	[p.interrog]	[n.rel]			
	[p.pers]	[n.count]				
	[p.pers]	[p.refl]	[d.plural]			

This combined presentation takes advantage of some patters that were not revealed in the initial search, but only discovered during the process of attempting the over schematization. In particular the following were helpful: *ñis brgya rtsa bdun-pa* [num.card] [num.card] [num.card] [num.ord] ‘207th’, *kha-che paṇ-chen de la* [n.prop] [n.count] [d.dem] [case.all] ‘to that great-pundit Kashmiri’, *btsun-mo chuṅ-ṅu gźan dag* [n.count] [adj] [d.det] [d.plural] ‘other small queens’, *dam-pa rje-btsun Mi-la* [n.v.invar] [n.count] [n.prop] ‘holy lord Mila’, *srin-po chen-po lṅa źig* [n.count] [adj] [num.card] [d.indef] ‘a five great demons’, *rnams thams-cad ñid* [d.plural] [d.plural] [d.emph].[9]

Table 1 (above) is a recapitulation of Table 4, excluding the two patterns that end with [n.rel], since it is convenient to see the relator noun as acting on rather than as a member of the preceding noun phrase.

5. Appendix 2: Large tables

6. Table 5: potential noun phrase structures of up to three elements

3620	[n.count]		
490	[n.count]	[d.plural]	
45	[n.count]	[d.plural]	[n.v.invar]
22	[n.count]	[d.plural]	[n.count]
12	[n.count]	[d.plural]	[d.dem]
12	[n.count]	[d.plural]	[d.plural]
9	[n.count]	[d.plural]	[n.v.past]
348	[n.count]	[n.count]	
91	[n.count]	[n.count]	[d.plural]
38	[n.count]	[n.count]	[d.dem]

[9] The patterns [adj] [num.ord] and [adj] [num.card] [num.ord] are not attested. This potentially suggests that ordinal numbers are syntactically treated the same as adjectives.

27	[n.count]	[n.count]	[num.card]
19	[n.count]	[n.count]	[adj]
13	[n.count]	[n.count]	[n.v.invar]
9	[n.count]	[n.count]	[n.v.neg]
9	[n.count]	[n.count]	[n.v.past]
8	[n.count]	[n.count]	[n.prop]
264	[n.count]	[n.v.past]	
261	[n.count]	[n.v.invar]	
83	[n.count]	[n.v.fut.n.v.pres]	
54	[n.count]	[n.v.fut]	
53	[n.count]	[n.v.neg]	
41	[n.count]	[n.v.past.n.v.pres]	
27	[n.count]	[n.v.cop]	
17	[n.count]	[n.v.fut.n.v.past]	
250	[n.count]	[d.dem]	
22	[n.count]	[d.dem]	[n.rel]
18	[n.count]	[d.dem]	[d.emph]
212	[n.count]	[adj]	
22	[n.count]	[adj]	[n.v.invar]
13	[n.count]	[adj]	[d.dem]
11	[n.count]	[adj]	[n.count]
10	[n.count]	[adj]	[d.plural]
10	[n.count]	[adj]	[n.count]
202	[n.count]	[n.prop]	
10	[n.count]	[n.prop]	[n.count]
172	[n.count]	[num.card]	
32	[n.count]	[num.card]	[n.count]
27	[n.count]	[num.card]	[num.card]
15	[n.count]	[num.card]	[n.v.invar]
13	[n.count]	[num.card]	[n.v.past]
42	[n.count]	[p.pers]	
20	[n.count]	[num.ord]	
16	[n.count]	[d.det]	

9 [n.count] [d.det] [d.plural]

14 [n.count] [p.interrog] [n.rel]

8 [n.count] [d.indef]
8 [n.count] [p.indef]

824 [n.v.invar]
10 [n.v.invar] [n.count]
8 [n.v.invar] [n.v.pres]
804 [n.v.past]
8 [n.v.past] [n.count] [num.card]
414 [n.rel]
396 [n.v.fut.n.v.pres]
8 [n.v.fut.n.v.pres] [n.v.past]
164 [n.v.fut]
12 [n.v.fut] [n.count]
127 [n.v.past.n.v.pres]
99 [n.v.pres]
9 [n.v.fut.n.v.past]

254 [n.prop]
19 [n.prop] [d.dem]
10 [n.prop] [n.count]

253 [p.pers]
10 [p.pers] [n.v.invar]
9 [p.pers] [n.count]
9 [p.pers] [p.refl]
8 [p.pers] [d.plural]
8 [p.pers] [p.pers]

93 [d.dem]
167 [d.dem] [n.rel]

70 [num.card] [num.card] [num.card]

62 [num.ord]
8 [num.ord] [n.count]

48	[adj]		
15	[adj]	[d.plural]	[n.count]
47	[d.det]		
37	[p.interrog]	[n.rel]	
26	[d.plural]		
9	[d.plural]	[n.count]	[d.plural]

Table 6: noun phrase structures excluding [n.count] [n.count] and [n.v.xxx]

3968	[n.count]		
581	[n.count]	[d.plural]	
22	[n.count]	[d.plural]	[n.count]
12	[n.count]	[d.plural]	[d.dem]
12	[n.count]	[d.plural]	[d.plural]
288	[n.count]	[d.dem]	
22	[n.count]	[d.dem]	[n.rel]
18	[n.count]	[d.dem]	[d.emph]
231	[n.count]	[adj]	
13	[n.count]	[adj]	[d.dem]
21	[n.count]	[adj]	[n.count]
10	[n.count]	[adj]	[d.plural]
210	[n.count]	[n.prop]	
10	[n.count]	[n.prop]	[n.count]
199	[n.count]	[num.card]	
32	[n.count]	[num.card]	[n.count]
27	[n.count]	[num.card]	[num.card]
42	[n.count]	[p.pers]	
20	[n.count]	[num.ord]	
16	[n.count]	[d.det]	
9	[n.count]	[d.det]	[d.plural]
14	[n.count]	[p.interrog]	[n.rel]

8	[n.count]	[d.indef]	
8	[n.count]	[p.indef]	
10	[n.v.invar]	[n.count]	
8	[n.v.past]	[n.count]	[num.card]
414	[n.rel]		
12	[n.v.fut]	[n.count]	
254	[n.prop]		
19	[n.prop]	[d.dem]	
10	[n.prop]	[n.count]	
253	[p.pers]		
9	[p.pers]	[n.count]	
9	[p.pers]	[p.refl]	
8	[p.pers]	[d.plural]	
8	[p.pers]	[p.pers]	
93	[d.dem]		
167	[d.dem]	[n.rel]	
70	[num.card]	[num.card]	[num.card]
62	[num.ord]		
8	[num.ord]	[n.count]	
48	[adj]		
15	[adj]	[d.plural]	[n.count]
47	[d.det]		
37	[p.interrog]	[n.rel]	
26	[d.plural]		
9	[d.plural]	[n.count]	[d.plural]

Table 7: noun phrase structures with all suspect and uninformative structures removed

581	[n.count]	[d.plural]	
12	[n.count]	[d.plural]	[d.dem]
12	[n.count]	[d.plural]	[d.plural]
288	[n.count]	[d.dem]	

22	[n.count]	[d.dem]	[n.rel]
18	[n.count]	[d.dem]	[d.emph]
231	[n.count]	[adj]	
13	[n.count]	[adj]	[d.dem]
10	[n.count]	[adj]	[d.plural]
210	[n.count]	[n.prop]	
10	[n.count]	[n.prop]	[n.count]
199	[n.count]	[num.card]	
27	[n.count]	[num.card]	[num.card]
42	[n.count]	[p.pers]	
20	[n.count]	[num.ord]	
16	[n.count]	[d.det]	
9	[n.count]	[d.det]	[d.plural]
14	[n.count]	[p.interrog]	[n.rel]
8	[n.count]	[d.indef]	
8	[n.count]	[p.indef]	
10	[n.v.invar]	[n.count]	
19	[n.prop]	[d.dem]	
10	[n.prop]	[n.count]	
9	[p.pers]	[n.count]	
9	[p.pers]	[p.refl]	
8	[p.pers]	[d.plural]	
167	[d.dem]	[n.rel]	
70	[num.card]	[num.card]	[num.card]
37	[p.interrog]	[n.rel]	

SOAS Working Papers in Linguistics Vol. 17 (2015): 49-76

Associated motion with deictic directionals: A comparative overview

Aicha Belkadi
ab105@soas.ac.uk

1. Introduction

This paper focusses on deictic directionals and their rarely discussed function as markers of Associated Motion (henceforth AM). AM is a term generally used to refer to a category of verbal affixes exhibited by a number of indigenous Australian, North American and South American languages, whose function is to indicate that the event encoded by a verb is framed with respect to a motion co-event (Koch, 1984; Wilkins, 2005; Guillaume, 2009 amongst others). The most characteristic systems of AM involve complex paradigms, where affixes are each paired with specific and quite sophisticated types of motion. Specifications may encode information about the direction and orientation of the motion event, a defined temporal relation with the verb's event, and particular aspectual notions. The examples below from the Australian languages Mparntwe Arrernte and Adnyamathanha illustrate some of the possibilities.

(1) **Mparntwe Arrernte** (Australia: Wilkins, 2006)
angk-***artn***.alpe-ke
speak-QUICK:DO&GO.BACK-PC
'Quickly spoke and then went back.'

(2) **Adnyamathanha** (Australia, Tunbridge, 1988)
yarta veni witi-nali-***angg***-alu
ground very pierce-CONT.COMING(sic)-PERF-3SG.ERG
'It (a piece of iron hanging down from a truck) poked into the ground all the way (to here).'

The affix *-artn-* in (1) marks that the event of speaking, lexicalised by the verb, is followed by a motion of the subject directed away from the deictic anchor. In addition, it also provides some information about the speed in which this event enfolds. The affix *-angg-* in (2) marks that the verb's event occurs continuously, as part of a motion oriented towards a deictic anchor.

A similar phenomenon has been described in a number of African languages from the Niger-Congo, Nilotic and Afro-Asiatic phyla (Frajzyngier, 1989; Bourdin, 2006; Alamin et al., 2012; Belkadi, 2014 amongst others) and for some South American languages, such as Quechua (Weber, 1989). In those languages, the motion co-event is marked by deictic directionals whose primary function is to encode ventive or itive path semantics rather than dedicated inflections. However the range of features that can be attributed to the added motion component often mirror those encoded by AM affixes. The example which follows from Taqbaylit Berber illustrates this.

(3) **Taqbaylit** (Algeria, Belkadi: 2014)
t-ɣra ***=d*** taktaf
3SGF-read.PRF =VEN book
'She read the book and came back.'

In (3), the event described by the verb *ɣra* 'to read' is understood as being followed by a motion event of the subject of the verb directed towards the location of the speaker. This additional motion event is, in this example, encoded by the clitic *=d*, which is more commonly used in Berber as a deictic directional marking ventive path semantics.

AM interpretations of deictic directionals have been discussed in the literature under different terminologies, such as 'alloying', 'roundtrip motion' or 'coerced motion' (cf. Alamin et al., 2012; Bourdin, 2006; Belkadi, 2014), but overall remain largely underdescribed. For the most part, sources present passing examples in more general descriptions of deictic directionals and their range of interpretations in specific languages. The only detailed description of the phenomenon is provided by Bourdin (2006) for a range of AM uses of the ventive and itive particles in Somali. As will be discussed in various parts of this paper, Bourdin finds that pragmatic factors ─ such as the context of utterance, the TAM carried by the main verb, and its inherent lexical meaning ─ have an effect on how these various meanings are derived in the language. This scarcity and narrow focus of the literature available on the topic means that particular aspects of AM marked by deictic directionals (which I will refer to as D-AM), and related issues are not really addressed. First, the uses of directionals as D-AM markers are never systematically contrasted with their basic deictic path semantics uses. If examples are always shown to occur with verbs describing non motion events, no source has made the stance to explicitly separate the two functions. However, given that the two functions of directionals tend not to occur with similar classes of verbs and, more importantly, trigger different entailments or interpretations, it seems crucial to establish a distinction between them. Second, none of the previous literature on the topic has compared or linked the AM uses of directionals with the types of AM affixes described in examples (1) and (2). Given the similarities between the two systems, such a comparative discussion seems crucial to an understanding of D-AM.

The main aims of the present paper are precisely to address these questions. The first goal is thus to highlight some characteristic typological features of D-AM and further emphasize the role of pragmatics (as mentioned by Bourdin), such as implicature/inference and relevance, in the derivation of these interpretations. Its second aim is to draw some comparison with AM expressed by inflectional affixes (which I will refer to as I-AM).

The descriptive discussion of D-AM presented in this paper is based on a modest and preliminary survey[1] of six languages. Somali (Cushitic), Pero (Chadic) and a few Berber languages represent the Afroasiatic group. Tima and Päri represent respectively

[1]This paper summarises on-going research. The author acknowledges that the languages considered (except Quechua) are closely related, but the discussion here is not intended as a fully-fledged cross-linguistic typology of AM expressed by directionals. The aim is rather to lay out some characteristics shared by the phenomenon in some languages and propose some criteria, which can be used as points of comparison for other languages.

Niger-Congo and Nilotic. Finally, Hualaga Quechua, from Peru, offers a non-African point of view. The data presented comes from two types of sources: general grammars of the relevant languages, and particular descriptions of deictic directionals and their uses in those languages. These works and their sources are referenced after relevant examples. The Taqbaylit data, unless stipulated otherwise, is data collected by the author during various fieldworks in Algeria. Given that D-AM is under-studied, it is not exactly known at this stage how widespread the phenomenon is. The languages for which data is available are all spoken in Africa, except Huallaga Quechua which is spoken in South America. Most of those languages, such as Pero, Berber and Somali, belong to the Afro-Asiatic phylum, and are spoken in Northwestern Africa, Nigeria and Somalia. The other African languages mentioned, Tima and Päri, although they belong to different language phyla, are both spoken in Sudan.

The paper is organised as follows. Section 2 introduces D-AM uses, and discusses how those differ from basic deictic directional uses. Section 3 is devoted to the characteristics of D-AM and the relation between the main verb's semantics and derived motion components. Sections 4 and 5 introduce the phenomenon of I-AM, and draw some comparison with D-AM.

2. Directionality and D-AM

Deictic directionals consist of verbal affixes, particles, clitics and sometimes prepositions which occur predominantly in the descriptions of motion events where they contribute a deictic path. Many languages have systems contrasting ventive morphemes, which orient the motion event to or towards the location of the deictic centre, with itive morphemes, which specify motion away or not in the direction of the deictic centre[2]. In discourse, the deictic centre very often corresponds to the speaker, but in narratives or with certain verbs directionals can be anchored with respect to other participants. The following Somali examples, where the particles *soo* and *sii* direct the motion event expressed by the verb *soc* 'walk' respectively toward and away from the speaker illustrate this.

(4) **Somali** (Somalia; Claudi, 2012)

a. wuu ***soo*** soc-eyaa
FOC: 3SGM VEN walk-PRSPRG:3SGM
'He is walking towards me.'

b. wuu ***sii*** soc-eyaa
FOC:3SGM ITV walk-PRSPRG:3SGM
'He is walking away from me.'

Deictic directionals also regularly modify verbs that do not belong to the semantic class of motion. Often, the interpretations they get derive from their spatial semantics and metaphorical associations. They are, for instance, frequently found in the descriptions of

[2] Languages differ as to the type of system of deictic directionals they use. Some languages have systems involving more morphological and semantic oppositions. Some languages only overtly mark ventive semantics. Amongst the languages surveyed Somali, Berber (except some varieties of Taqbaylit), and Päri formally encode a distinction between ventive and itive directions. Pero, Tima and Quechua only display ventive markers.

natural phenomena and bodily secretions, and descriptions of vision or perception events. These uses can be illustrated with the ventive clitic *=d* in Berber, which canonically occurs in descriptions of motion events (5a), but is also often found in clauses which describe events relating to bodily secretions, such as (5b), and those which describe events of perception, such as (5c).

(5) a. **Ait Seghrouchen Berber** (Morocco; Bentolila, 1969)
i-ṛaḥ =***d*** ġr i
3SGM-go,PRF =VEN to POSS.1SG
'He came to my house.'

b. **Ait Seghrouchen Berber** (Morocco; Bentolila, 1969)
imṬawn La Ṭaṛn ***=id*** zG^{w}aLn-inw
tears PROG go.down =VEN eyes-POSS.1SG
'Tears came down from my eyes.'

c. **Ghadames Berber** (Libya; Kossman, 2013)
a ***ddu***_t-əswăḍ ba_tănăt
when VENT_3S:F-look:P be.not:P_3P:F:ACC
'When she looked, they were not there.'

In some of these contexts, deictic directionals can mark a contrast of visibility (see Fortis & Fagard, 2010). Ventive morphemes mark the fact that a particular object or abstract entity comes into the visibility of the speaker. Itive morphemes, on the other hand, specify the disappearance of an entity from the speaker's visibility. In (5b), for instance, the tears move from the inside to the outside of the body, where they become visible to the speaker and the ventive is used (Bentolila, 1969; El Mountassir, 2000 for a complete discussion relating to Berber). In other contexts, the deictic directionals may be used simply because the events described are perceived as involving motion, metaphorically. This is the case with events of perception and vision, which have been shown to be conceptualised in terms of the 'fictive' motion (i.e. abstract motion) of a perceived stimulus or of a sensory function of the perceiver (Talmy, 2000; Slobin, 2004). The event of looking in sentence (5c), from Ghadamsi Berber, is an instance of such fictive motion. It involves an abstract path, which the ventive deictically modifies[3] (Kossman, 2013; Belkadi, In Press). There would be, of course, much more to say about the uses of deictic directionals in such contexts. The point of this very brief discussion is that whether they mark visibility or fictive motion, ventive and itive morphemes are here used with their basic deictic path semantics: they specify the event encoded by the verb they modify as directed toward or away from the space of a deictic centre.

Another well-documented and fairly common route of extension for deictic directional morphemes is their grammaticalization into tense, aspect or modality markers (Bourdin, 1999; Heine & Kuteva, 2002). Grammaticalization into TAM markers reflects more general conceptual metaphors found in languages, associating the domains of space or motion with the domain of time (Bourdin, 2006; Evans, 2013; Lakoff & Johnson, 1980 amongst others). The Somali ventive particle *soo* discussed in (4) has, in addition to its

[3] The deictic anchor here seems to be the experiencer in subject function rather than the speaker.

spatial semantics, aspectual and temporal uses. This is illustrated by the contrast in interpretation between the pairs of sentences below. Example (6b), which contains the particle, has a recent past interpretation not available in (6a), where the particle is absent. Similarly, the event described in (7b) is construed as continuing until present time, while it is not in (7a) (Bourdin, 2006: 25-26).

(6) **Somali** (Bourdin, 2006)

a. wuu iga bax-ay
FOCUS:3MASC.SG 1SG.OBJECT:from go.out-PAST:3MASC.SG
'He left my place (and went to do some errands).'

b. wuu iga ***soo*** bax-ay
FOCUS:3MASC.SG 1SG.OBJECT:from VEN go.out-PAST:3MASC.SG
'He has left my place not long ago.'

(7) a. toban sanadood uun baan bariis cun-ayey
Ten years only FOCUS.1SG rice eat-PAST.PROG:1SG
'During ten years (that I spent there) all I ate was rice.'

b. toban sanadood uun baan bariis ***soo*** cun-ayey
Ten years only FOCUS.1SG rice VEN eat-PAST.PROG:1SG
'[Enough is enough] for ten years now all I have eaten is rice.'

D-AM, the extension of deictic directionals as markers of Associated Motion, although more frequent than it appears, is typologically rarer than the types of extensions aforementioned. It also differs in several respects. The main distinction is of course that, in D-AM contexts, deictic directionals presuppose a motion event which occurs in addition to the event encoded by the verb stem they modify. Very frequently the verb stem does not encode motion, however the interpretation can arise with some verbs of motion, as in the following example from Taqbaylit Berber.

(8) **Taqbaylit**

i-ʕum =***d***.
3SGM-swim.PRF =VEN
'He (went somewhere) swam and came back (to the location of the speaker or to his house).'
'*He swam (towards or to the location of the speaker).

Despite containing just one verb, the previous example describes two events. The main event, encoded by the verb form *iʕum* 'he swam', occurs against the backdrop of a motion event which is encoded by the deictic directional *=d* and encodes the path [go and come back]. The swimming event is understood as taking place in a location different to that of the speaker, but is not the event directly anchored deictically by the directional. Instead it seems that the AM event is directed with respect to a deictic anchor. Contrast this example with (9) in which the event encoded by the verb 'to jump' is the one being deictically anchored.

(9) **Taqbaylit**
t-jjlb =***d*** γr tabla.
3SGF-jump.PRF =VEN to table
'She jumped on the table (in the direction of the speaker).

The crucial difference between the interpretations of the ventive in (8) and (9) is that in the former the deictic path is a subcomponent of the motion event encoded by the verb, while in the latter, deictic path is external to that event. Adopting Talmy's terminology (2000), the verb in (9) lexicalises both a motion component and a manner co-event. The clitic =*d* contributes a deictic path semantics to that complex event. The resulting interpretation is one in which the motion referred to by 'jump' in this particular context is translational. In (8), while the verb 'swim' encodes motion and manner, the deictic clitic does not intersect with these components. It is the AM event which is oriented deictically, rather than the event expressed by the verb.

Since the deictic path component in AM contexts is not a subpart of the event, it also differs from directional interpretations of deictics in terms of TAM sharing. In some rare contexts, seemingly when the verb carries present tense or imperfective semantics, the additional motion event can be understood with a future tense reference. The following data from Taqbaylit and Somali illustrates this.

(10) a. **Taqbaylit**
lla =***d*** i-tt-awum
PROG =VEN 3SGM-IMPRF-swim.IMPRF
'He is swimming and then he will come back.'

b. **Somali** (Bourdin, 2006)
wuxu ***sii*** joog-aa London
FOCUS.3MASC.SG ITV stop/stand-PRES:3MASC.SG London
'He is in London right now and will continue on his way (towards a place where I am not)

Finally, unlike other meaning extensions, such as TAM ones, D-AM extensions cannot be said to derive from conceptual metaphor. The reanalysis of deictic directionals as TAM markers involves a transfer of their spatial properties to the domain of time (Heine, Claudi & Hünnemeyer 1991). D-AM, on the other hand, does not involve such transfer between different domains of conceptualisation. In fact, both D-AM and deictic path uses of directionals belong to the same domain of space.

All these idiosyncrasies of D-AM show that it has to be set apart from other uses of deictic directionals, and be considered as a distinct route of grammaticalisation for these morphemes. A preliminary and tentative scenario for how D-AM may develop from the basic deictic path semantics of directionals is left until section 5. A brief typological survey of D-AM is first given in section 3 below, focussing on the range of properties attributed to the AM event.

3. Descriptive properties of D-AM

The motion event presupposed in D-AM occurs in addition to the event encoded by the verb modified by the deictic directionals, however the two events are semantically closely related. These semantic relations vary in different ways. Variations seem to be caused by the different constraints imposed by particular languages, but also often depend on the inherent semantics of the main verb and relevance given a particular pragmatic context. As mentioned in section 1, the role of verb meaning and pragmatics in deriving various D-AM interpretations has been highlighted by Bourdin (2006) in Somali. What follows differs from Bourdin's work in two ways. First, the properties of D-AM are here identified from a comparative perspective. Second, it focusses on how variations affect different aspects or components of the motion event: its time relation with the main verb's event, the shape of its path and the identity of the entity in motion. It also identifies a more prominent role to the lexical semantics of the verb and the type of events it refers to in deriving D-AM in the first place. These different aspects related to D-AM are discussed in turn from sections 3.1 to 3.4 below.

3.1. Time relations

Three time relations can obtain between the main verb's event and the motion event triggered by deictic directionals: precedence, concomitance and subsequence. Precedence obtains when the D-AM event occurs prior to the main verb's event. This is illustrated in (11) with two examples from Päri.

(11) **Päri** (Sudan; Andersen, 1988)

a. ùbúr á-***ŋùt***-ò
Ubur C-cut+ITV+AP+INTR
'Ubur went to cut.'

b. ùbúr á-***ŋùnn***-ò
Ubur C-cut+VEN+AP-INTR
'Ubur came to cut.'

In (11a), the cutting event lexicalised by the verb is understood as being preceded by a motion of the subject *Ubur* towards a location which is not that of the speaker. This prior motion event is marked by the itive directional, incorporated into the verb. Notice that the two events are related semantically, since the motion event occurs for the purposes of the event encoded by the verb. A similar presupposition is made for (11b), where the event of cutting is preceded by a motion of the subject, this time directed towards the location of the speaker. Again, motion is marked by the ventive directional on the verb.

Concomitance obtains when the two events take place at simultaneous time intervals. In (12) below, from Tamasheq Berber, the event described by the verb glossed as 'to work' occurs over the duration of the motion event implied by the ventive marker *ə́dd*.

(12) **Tamasheq Berber** (Heath, 2006)
i-šššəγæ-***ədd***
3SGM-work.PERF-VEN
'He came working (he was working as he came).'

This temporal relation seems to be mainly derived from pragmatic factors. Indeed, as will be discussed in more detail later, concomitance is often found in particular situational contexts or with verbs which describe events more likely to occur as part of an overall motion event.

The third temporal relation that can hold between a D-AM event and the main verb's event is subsequence. In such instances, the added motion occurs after the event described by the main verb. The examples in (13) illustrate this.

(13) a. **Tima** (Sudan, Alamin et al., 2012)
kɔ́yɔ̀-***ɔ̀ŋ*** k-ùrtú
build.IMP:SG-VENT NC.SG-house
'Build the house and come.

b. **Pero** (Nigeria, Frajzyngier, 1989)
cúg-***ínà*** tù púccù
fall-COMPL.VENT PREP there
'He fell there and came.'

c. **Somali** (Bourdin, 2006)
soo joogso
VEN come.to.a.halt:IMPER.2SG
'Stop fooling around and come here!'

In the examples above from Tima, Pero and Somali, the motion events are marked by the language's respective ventive directionals, and occur after the event encoded by the main verb.

Overall, time relations seem to follow consistent patterns given a particular language and a particular directional. In Berber the ventive directional primarily triggers subsequent motion interpretations, while the itive is used primarily for precedence. Kossman (2013) describes (but does not give examples) Associated Motion interpretations triggered by the itive clitic with verbs of verbal interaction such as *ăn* 'to say', *aslil* 'to call', *ăsmağği* 'to speak' in Ghadamsi. In addition to describing that the agent is not in the same location as the deictic centre, the itive may presuppose a prior motion of the verb's subject. The following examples from Tamasheq Berber show further the link between ventive and subsequence and the link between itive and precedence in Berber.

(14) **Tamasheq** (Mali, Heath, 2006)
a. i-jræw-***ədd*** áẓrəf y a-d ǝqqən-æγ
3SGM.S-find.PERF-VEN money DAT DEM-COMIT build.SHIMPRF-1SG.S
'He got (literally found) money and brought it in order that I build.'

b. i-kfa-***hín*** ázṛəf è mæssi-s
3SGM.S-give.PERF-ITV money DAT master-3SG.POSS
'He went and gave the money to his master.'

The temporal connections that arise are not necessarily strict. D-AM events may occur at other times depending on the main verb's meaning and situation referred to, or the particular context of utterance. Bourdin (2006) shows that D-AM involves concomitance instead or in addition to subsequence in Somali, if the pragmatic context allows it. To illustrate this variation in interpretations, he discusses the examples in (15), which involve the verb *seex* 'to sleep'. The event that the verb describes can be understood as occurring before the motion event marked by the ventive *soo*, as in (15a.i). Alternatively, if the subject of the verb is known by the speech participants to have travelled prior to the time of utterance – for instance by bus – the verb's event can be presupposed to have occurred during this particular motion situation. This is what happens in the second interpretation of (15a). The example in (15b) contains a specification of the location of the sleeping event described by *seex*. There the event is explicitly specified as having occurred inside a bus. The presence of the ventive marker implies that this bus journey took place toward the location of the speaker.

(15) **Somali** (Bourdin, 2006)
a. waan ***soo*** seex-day
FOCUS:1SG VEN sleep-PST:1SG
(i) 'I took a nap before coming here.'
(ii)'I took a nap on my way here (on the bus).'

b. bas-kuu ku ***soo*** dhex seex-day
bus-DEF:3SGM in VEN inside sleep-PST:3SGM
'He slept on the bus on his way here.'

As the following examples show, similar contextual coercions are found in Berber. In (16b), the event of working is understood as occurring during a journey to the deictically anchored location. Example (15a) is slightly different. Concomitance is inferred because of the agent of the verb in this particular sentence, and the typical ways in which it acts: a fire typically spreads out while destroying (hence eating) everything on its path.

(16) **Tamasheq** (Heath, 2006)
a. i-kša-***hin***
3SGM.S-eat.PERF-ITV
'It (=bush fire) ate up (the vegetation) going away that way.'

b. i-ššəγæ-***ədd***
3SGM-work.PERF-VEN
'He came working (he was working as he came).'

As mentioned before, the inherent meaning of a verb and the type of event it lexicalises also participate in triggering temporal adaptations. Bourdin (2006), again, discusses some examples from Somali, which are provided in (17) below. The examples show the

interpretations on the motion event which arise when the ventive *soo* is used with the verbs glossed as 'to sleep' and 'to sit down' in face to face conversations. In these contexts, the directional derives an anterior motion, rather than the most common subsequent motion. The reason for these coercions is that these events usually imply the impossibility of motion of their subject while they unfold. What's more, the result states they trigger usually last for some time. Thus a motion following these events, in the short time interval presupposed by a face to face conversation is rather unlikely. Given this, the most relevant type of motion is one which is prior to the verb's event.

(17) **Somali** (Bourdin, 2006)

a. ***soo*** seexo
VEN sleep:IMPER.2SG
'Come sleep over here!' [face to face conversation]
'*Sleep and come over here.'

b. ***soo*** fadhiso
VEN sit.down:IMPER.2SG
'Come here and sit down!' [face to face conversation]
'*Sit down and come here!'

Due to the scarcity of the data on AM interpretations of deictic directionals available in the literature, it was not possible to find more examples of such changes in the various languages consulted. A possible similar instance was however found in Tamasheq Berber involving a verb describing a drinking event. In this example, provided in (18) below, a causative motion event (translated as 'to give' by the source) seems to be presupposed by the presence of the ventive directional.

(18) **Tamasheq Berber** (Heath, 2006)

Ø-òlæs-à-s æ-næ̀s-bahù t-ənna
3SGM-repeat.PERF-DAT-3SG SG-AGENT-lying 3SGF-say.VblN

i-ffúd a-tt-***ədd***
3SGM-be.thristy.RESLT DEM-3SGM.O-VEN

ì-s-əsə́w
3SGM-CAUS-drink.AOR
'The lying man proceeded to say (that) he was thirsty, and he (=midget) should give him (something) to drink.'

Example (18) is interesting because, as mentioned previously, the ventive directional in Berber, including Tamasheq, usually derives a subsequent motion. However, in this particular context and with this particular verb, the motion event occurs prior to the drinking event. As in Somali, precedence seems to be triggered here by the interaction between the inherent semantics of the verb and pragmatic relevance. Indeed for an agent to make a patient drink, which is essentially what the example describes, it makes more sense for the former to move towards the latter before rather than after. This example and the few preceding ones demonstrate the crucial role of the semantic and pragmatic contexts in determining the time in which the presupposed motion occurs with regard to

the verb's event. As will be shown in the following sections, pragmatics and semantics also play a role in fixing other features of the motion event.

3.2. Shape of the path

The shape of the path associated with the presupposed motion event varies too. Most of the data discussed in the previous sections involve motion whose paths are simplex: in each case, the event encoded by the verb occurs in a particular location and is either preceded or followed by a motion event directed towards or away from the deictic centre. AM encoded by ventive elements can however involve complex paths of the type [go and come back/ return]. The first part of the path is directed towards a location distinct from the deictic centre, where the main verb's event takes place. The second part of the path is directed toward the deictic anchor. The first part of the path may not be overtly given in translations, but instead implied by the return part of the path. This 'return-shaped' path (after Wilkins, 2006) arises in many contexts and in most of the languages surveyed. It is this particular shape of the path and its frequency which has led scholars like Bourdin (2006) to use the term 'roundtrip motion'. The following examples from Quechua, Taqbaylit and Tashelhit Berber illustrate this particularity.

(19) **Huallaga Quechua** (Peru; Weber, 1989)

a. tanta-ta ranti-Ri-***mu***-y
bread-OBJ buy-sud-VENT-2IMP
'Go buy bread (and return quickly).'

b. wanu-chi-***mu***-sha-: aycha-ta
'The meat that I killed (i.e. which I went off and killed and brought back here).'

(20) a. **Taqbaylit**
tlm-γ =***d*** taqbaylit
learn.PRF-1SG =VEN Taqbaylit
'She learned Taqbaylit and came back.'

b. **Tachelhit** (Morocco; El Mountassir, 2000)
i-kka =***d*** Brahim Franṣa
3SGM-pass.PRF =VEN Brahim France
'He passed in France and came back.'

A return-shaped path seems to be the default interpretation of D-AM in most northern Berber dialects which present the trait, in Quechua and Somali. The frequency of these interpretations may be attributed to particular elicitation techniques. Informants may need to first construct a motion away from the location of utterance, in order to then construct a motion event whose path can be directed toward the deictic centre. Data from Tima discussed in Alamin et al. (2012) seem to support this hypothesis. In this language, complex paths are limited to discourse situations where both the speaker and their addressee are in the same location. In such contexts, the first part of the path has to be overtly expressed by the Tima counterpart of the English verb 'to go'.

(21) **Tima** (Alamin et al., 2012)

a. àyí mɔ́ɔ̀k-***iŋ***
go.IMP:SG drink.IMP:SG-VENT
'Go, drink, and come (back).'

b. àyí kɔ́yɔ̀-***ɔ̀ŋ*** k-ùrtú
go.IMP:SG build.IMP:SG-VENT NC.SG-house
'Go, build the house and come (back).'

Alternatively, return-shaped motion might be argued to be triggered specifically by the 'toward deictic centre' semantics of ventive directionals. It is possible that ventive morphemes are interpreted in those instances as directing the motion event to a prominent location, rather than simply the location of a deictic centre. A location already visited by a subject is more prominent, and this would make the return meaning more frequent.

Complex D-AM gives rise to a subtle range of interpretations. In the following example from Somali (Bourdin, 2006), D-AM motion involving a motion of the subject away from the speech location and a return to the speech location underlies the meaning of that sentence, even though it is not overtly expressed as such.

(22) aad baan u ***soo*** cun-i
much FOCUS:1SG to/for VEN eat-INFNT
'I'm going to stuff myself while you wait for me here.'

The precise shape attributed to the path of the presupposed motion event is context-dependent. Concomitant motion, for instance, is more likely to involve a simple path, while subsequent motion tends to involve a return-shaped path. The semantics of the verb modified by the clitic may also prompt specific path semantics. In Taqbaylit Berber, for instance, verbs lexicalising states or events involving the appearance of a new state, coerce a path which is more likely to be simplex, than complex (although these interpretations are also possible). In such contexts, the predominant interpretation is one in which the new state comes about in a distal location, followed by motion of the theme to a goal location, where the speaker is. This is illustrated with example (23).

(23) **Taqbaylit**
t-zz-bzəgg ***=d*** iman =is yawək.
3SGF-CAUS-be.wet.PRF =VENT self =POSS.3SG all
'She arrived soaking wet (to the location of the speaker)'.

Like any other motion event, D-AM presupposes the displacement of an entity, which after Talmy (2000) I will refer to as the figure of the event. In all the examples so far, the figure systematically corresponds to the subject of the main verb. This, however, is not necessarily the case. These variations are covered in the next section.

3.3. Identity of the figure

As mentioned above, in most cases of D-AM, the figure of the presupposed motion event coincides with the subject of the verb. However, some languages display more

complex patterns. In Pero, Ait Seghrouchen and Tamasheq Berber, the figure can be the object of a transitive verb. In such cases D-AM derived from transitive verbs is causative. There must be contemporaneity between the motion of the verb's subject and its object. The figure is the object of the verb, while the subject of the verb is understood as the agent of the causative motion.

(24) **Pero** (Frajzyngier, 1985)

a. ní -íp ***-nà*** túję̀
1SG catch COMP.VEN horse
'I caught a horse (and brought it).'

b. nì -íp ***-nà*** póngè-ì
1SG catch COMP.VEN the money
'I caught the money (and brought it).'

(25) **Ait Seghrouchen Berber** (Bentolila, 1969)
aṛaḥ ažm ***=d*** aman
go.AOR draw.out.AOR VENT water
'Go draw out some water (and bring it here).'

(26) **Tamasheq** (Heath, 2006)
i-jræ̀w-***ədd*** áẓrəf y a-d
3SGM.S-find.PERF-VEN money DAT DEM-COMIT

ə̀qqən-æγ
build.SH.IMPRF -1SG.S
'He got (literally found) money and brought it in order that I build.'

The figure of AM may not be an argument of the verb, and instead be a speech participant or some prominent discourse entity. This is quite rare and seems to be triggered by both pragmatic and semantic factors. In the corpus surveyed only two examples were found in Taqbaylit Berber and Quechua, two languages which follow a consistent (almost strict) subsequent motion pattern. Both examples involved stative verbs translated as 'to remain' modified by a ventive directional. They are presented in (27) and (28).

(27) **Huallaga Quechua** (Weber, 1989)
tayta-: ospital-chaw keeda-kU-***mu***-sha
father-1P hospital-LOC remain-REFL-VENT-3PRF
'Father remained (over there) in hospital.'

(28) **Taqbaylit** (Mettouchi, 2011)
tə-qqim ***=dd*** faṭima tuhṛiʃt-nni
3SGF-sit.PRF =VEN Fatima clever-DEM
'(There) remained Fatima.'

In the above examples, the deictic expressions 'over there' and 'there' are not overtly expressed but implicitly understood and provided in the translations. These deictic

expressions, which serve to locate the place of the states unfolding as distinct from the deictic centre, clash with the ventive semantics of the directionals used. These unexpected inferences are easily explained by assuming that there is motion expressed there, but one where the figure is not the subject of the verb — who is described as remaining in a distinct location — but the speaker or another participant.

More rarely, the directional itself forces subject-figure 'disjointness' (Bourdin, 2006). Amongst the languages surveyed, only Somali seems to follow this pattern. Precisely Bourdin (2006) attributes a function of 'switch-reference' to the itive *sii*. Where a motion event whose endpoint is distinct from the deictic centre cannot be inferred, a motion whose figure is a discourse participant, rather than the main verb's subject is constructed (see also Claudi, 2012).

(29) **Somali** (Claudi, 2012)

a. ***sii*** seexo
ITV sleep.IMPRF.2SG
'Have a rest until I come back.'

b. aad baan u ***sii*** cun-i
much FOCUS:1SG to/for ITV eat-INFN
'I'm going to stuff my face, while you are out.'

c. muus wuu ***sii*** cun-eyaa
banana FOCUS ITV eat-PRSPRG:3SGM
'He is eating a banana while I'm absent.'

It is quite evident that a range of factors play a role in the different properties of an additional motion event. One of these factors is the particular semantics of the main verb. Indeed it was shown at different points of the previous discussion that the inherent semantics of a verb affect the time relation of an AM or the identity of its figure. The particular semantics of the main verb may actually play a more crucial role. It may be a pivot not only in the types of AM we get, but more importantly on whether we get it at all. This issue is covered in section 3.4.

3.4. Main verb's semantics

The two functions of deictic directionals distinguished in section 2 — i.e. basic directionality and D-AM — rarely overlap. Each function is associated with coherent semantic classes of verbs. Most of the languages surveyed seem to have a clear-cut dichotomy whereby motion verbs trigger directional uses, while other classes of verbs trigger D-AM uses. This seems to be the option followed by Somali (Bourdin, 2006; Claudi, 2012). Other systems are a bit more complex. In Tima (Alamin, et al., 2012), for instance, verbs describing motion events as well as verbs describing natural phenomena and bodily secretions trigger directional uses of the ventive affix, while other classes of verbs derive Associated Motion interpretations.

(30) **Tima** (Alamin et al., 2012)

a. kɨ̀-címbʌ́rí ɛ́n-díík-***íŋ*** áỳɪnʈ́ɔ́ mádɔ̀ràsà
NC.SG-child TAM-walk-VEN DIR:SPEAKER.THERE school
'The child is walking towards the school (where I am).'

b. kw-àà̰ʈɔ́k ɛ́-w-ɔ̀dànà-y-***íŋ***
NC.SG-sky TAM-EE-cry-EE-VEN
'It was thundering (lit. the sky cried towards the speaker)'.

(31) a. kɔ́yɔ-***ɔ̀ŋ*** k-ùrtú
build.IMP:SG-VENT NC.SG-house
'Build the house and come.'

b. mɔ́ɔ̀k-***íŋ***
drink.IMP:SG-VEN
'Drink and come.'

In Qechua (Weber, 1989), D-AM is more likely to be triggered by verbs that do not describe motion events. The only exception seems to be with verbs of emitting and perception of some stimulus, which do not involve D-AM as long as the speech participants are separated by some distance (Weber, 1989). Compare, as an example, the interpretations of (32a) and (32b).

(32) **Huallaga Quechua** (Weber, 1989)

a. rika-ykU-***mu***-nki wasi-nchi:-ta ima-shi ka-yka:-n
see-IMPACT-VEN-2IMP house-12P:-OBJ what-IND be-IMPRF-3
'Go see what there might be in our house (and come back).'

b. qam-ta rika-***mu***-:
you-OBJ see-VEN-I
'I see you (over there).'

Both clauses provided in (32) are headed by the verb of perception *rika* 'to see', which in both instances is modified by the ventive affix *-mu-*. In (32a), the speaker and subject of the verb, which happens to be the addressee, seem to be located in the same space and a motion event interpretation is triggered. In (32b), the speech participants are not in the same location and the ventive does not give rise to an AM interpretation.

Taqbaylit Berber and a number of northern Berber languages have an overall clear-cut distinction between D-AM verbs and others, but the dichotomy is more difficult to establish. Deictic directional readings are found with verbs of motion, as long as they encode an orientation or imply translational motion, verbs of perception and emission of a stimulus, verbs of transfer, degree achievements, inchoative verbs and their causative counterparts (Belkadi, 2014; in press). All other activity verbs, statives or motion verbs which describe events not conceptualised as involving traversal of space (such as the verb 'to dance' in 34) derive D-AM interpretations (Ibid). Some examples are provided in (33) and (34).

(33) **Taqbaylit Berber: directional readings** (Belkadi, 2014)

a. t-jjlb =***d*** γr tabla.
3SGF-jump.PRF =VEN to table
'She jumped on the table (in the direction of the speaker).'

b. t-γni =***d*** taʕcit kaml!
3SGF-sing.PRF =VEN afternoon whole
'She sang the entire afternoon!'

c. y-vna =***d*** uxxam.
3SGM-build.PRF =VEN house
'The house got built'.

d. y-ḥma =***d*** yimnsi.
3SGM-be.hot.PRF =VEN dinner
'The dinner became hot.'

e. fka-n =as =***d*** i tqcict snat n tibwadin n lggatu.
give.PRF-3PLM =3SG.DAT =VEN DAT girl two of boxes of sweets
'They gave her two boxes of sweets.'

(34) **Taqbaylit Berber: D-AM readings** (Belkadi, 2014)

a. t-zz-bzgg =***d*** iman =is yawk
3SGF-CAUS-be.wet.PRF =VEN self =POSS.3SG all
'She went somewhere, soaked herself and came back' or 'She arrived soaking wet (to the location of the speaker)'.

b. t-γra =***d*** taktaf
3SGF-read.PRF =VEN book
'She read the book somewhere else and came back (to the location of the speaker).'

c. t-cdḥ =***d*** di tamγra
3SGF-dance.PRF =VEN in wedding
'She danced at the party and came back.'

Based on these facts, Belkadi (2014) proposes that directionals in Taqbaylit deictically modify any verb which lexicalises a path or encodes an event conceptualised as involving translational motion[4]. Events which do not fit the relevant criteria derive D-AM readings. The fact that all languages surveyed display D-AM with verbs not encoding motion, or for some with verbs expressing events which do not involve fictive motion (e.g. events involving vision, perception and emission of stimuli), do not describe bodily secretions, natural phenomena or change of states, corroborates this

[4] Belkadi (2014)'s analysis is based on Beavers (2008)'s analysis of the goal marker *–ni* in Japanese, and work by Beavers, Levin & Tham (2009, and references therein) identifying the role of pragmatics in directional interpretations of locative prepositions with some manner verbs in Romance and Germanic languages.

analysis. The exclusion of D-AM with motion verbs, particularly those which encode orientations or paths, suggests that the phenomenon is primarily derived pragmatically, as a kind of 'last-resort' interpretation. This pragmatic derivation further explains the range of D-AM interpretations: it occurs in contexts where motion is required and speakers construct the most appropriate motion event given the context.

This analysis of D-AM as a pragmatic reconstruction based on inference and relevance is further supported by the fact that various verbs in some of the languages surveyed can trigger either D-AM readings or basic directional readings depending on the actual situation they refer to. An example is given below from Quechua, where the verb translated as 'to work' is deictically modified by the ventive affix *mu* if it describes an event which can be interpreted as involving motion (35a). But if it cannot be interpreted as such, it presupposes the occurrence of a separate motion event (35b).

(35) **Huallaga Quechua** (Weber, 1989)

a. sikya-ta aru-***mu***-shaq
ditch-OBJ work-VEN-1FUT
'I will work the irrigation ditch (proceeding toward here).'

b. wasi-ta aru-***mu***-shaq
house-OBJ work-VEN-1FUT
'I will work the house (going over there to do so and then returning).'

The properties of D-AM highlighted by this small survey are summarized below.

(i) D-AM involves consistent time relations between the added motion event and the event encoded by the verb. This relation is either one of precedence or subsequence. Those time relations can, however, be reversed if the pragmatic or syntactic context requires it.

(ii) D-AM seems to consistently also imply concomitance between the two events involved, if this interpretation is the most relevant given the context.

(iii) The path of D-AM can be straight or complex. In some languages, again, relevance plays an important role in the type of path constructed.

(iv) In most D-AM cases, the figure of the motion event is identified with the subject of the verb, but if the verb is transitive the figure may be identified with the object of the verb. More rarely another participant can also be the figure of the motion event.

(v) D-AM readings tend to arise in contexts where a canonical directional reading is not derivable, often outside of motion event descriptions. It follows from this that D-AM interpretations of deictic directionals are in complementary distributions with their basic directional readings.

(vi) D-AM presuppositions are context-dependent.

This section has highlighted the main semantic properties of AM marked by deictic directionals. As mentioned in section 1, this phenomenon is very similar to Associated Motion marked by an inflectional or affixal category (I-AM) found in a number of Australian and Amerindian languages. The remainder of this paper is devoted to comparing the two systems. In section 4 below, a description of I-AM is provided. In section 5 a comparison between D-AM and this category is provided.

4. Inflections of Associated Motion in Australian and Amerindian languages

Affixes marking Associated Motion are reported in a range of Australian languages from the Pama–Nyungan phylum, such as Yidiŋ (Dixon, 1977), Mparntwe Arrernte (2006), Adnyamathanha (Tunbridge, 1988); Diyari (Austin, 1989) and Kaytej (Koch, 1984). Similar categories are described in some indigenous North and South American languages, such as Atsuwegi (Talmy, 2000: 123), Olutec and Cavineňa (Guillaume, 2009). This category is not widely studied and only a few detailed descriptions exist for particular languages (cf. Koch, 1984; Tunbridge, 1988; Wilkins, 2006; Guillaume, 2009), especially outside of the Pama-Nyungan family. These rare overviews describe I-AM as occurring productively, with a wide range of verbs including motion verbs, and present paradigms which mark a number of semantic oppositions, many identical to those found with D-AM. Guillaume (2009: 198-201) provides a brief comparative survey of the features encoded by these affixes in two languages ─ Cavineňa, a Tacanan language spoken in Bolivia and Mparntwe Arrernte, spoken in central Australia ─, and concludes that such systems encode specifications of the time relations between the I-AM event and the verb's event, information about the deictic direction and shape of path of the added motion component, as well as a specification on the identity of the figure. These suffixes may also encode other kind of information relating either to the verb's event, the AM itself or its outcome.

Given the small amount of data available, a typological description of the phenomenon is not possible at this stage. However, as an illustration of the types of semantic features encoded by the affixes and for comparative purposes with D-AM, an overview of I-AM paradigms in four languages which systems have been described in more or less detail, are presented below. These languages ─ Cavineña, Mparntwe Arrernte, Adnyamathanha and Atsuwegi ─ belong to three distinct genetic groups (respectively Tacana, Pama-Nyungan and Palaihnihan) and are spoken in each of the three different geographical areas where AM affixes have been reported: South-America, Australia and North-America.

4.1. Cavineña: South America

Cavineña (Guillaume, 2009) has a complex system of I-AM suffixes, which occur frequently with all verb classes except deictic verbs. These suffixes form complex paradigms, and are each paired with a meaning which may include deictic orientation, a specific time relation with the verb's event, some aspectual specification of the motion event (iterative or punctual), some aspectual specification of the outcome of the motion event (whether it is temporary or permanent) and a specification about the identity of the figure. Some examples are provided in (36)-(38).

(36) **Cavineña** (Bolivia; Guillaume 2006)

a. tudya =ekwana ba-***li***-kware takure.
then =1PL see-**GO.TEMP**-PASS.REV chicken.
'Then we went and saw our chickens (at the back of the truck where they were gathered and dead)

b. wipichitana=tsewe =tura =Ø tya=***na***-ya jae.
curtsey=with =3SG.ERG =1SG give-**COME.TEMP**-IMPF fish
'She came to give me fish with a curtsey.'

(37) …kiketere-aje-kware maju-***diru***=ishu.
complain-**GO.ITERATIVE**-PASS.REV die-**GO.PERM**=but
'(The jaguar which I had just shot) complained repetitively while going away from me to go and die further away.'

(38) era =tu peta-***kena***-chine epuna
1SG.ERG =3SG look-**GO**-PASS.REC femme
'I looked at the woman and I left.'

The suffix *-li* in (36a) marks a prior motion event, directed away from the deictic anchor, and whose end location is temporarily visited by the figure. The suffix *-na* in (36b) encodes the same semantic traits, but provides a different deictic direction, since the motion event is directed toward the deictic anchor. They contrast with example (37) in which there are two I-AM markers: *-aje*, which encodes iterativity and simultaneity in addition to deictic direction, and *-diru*, which encodes deictic direction, precedence, but also specifies that the end location is a permanent one. The example in (38) involves the suffix *-kena*, which marks subsequence. Cavineña also has two I-AM markers, *-tsa* and *-dadi*, found on transitive verbs, which in addition to deixis and simultaneity, specify that the entity moving is the object of the verb. Two examples are provided in (39) below with the verb *ba* 'to see'. In both cases, the stimulus, realised grammatically as the object of the verb, is understood as the figure of the associated motion event.

(39) a. tume =pa =taa =tuja =tu ba-***tsa***-ya ekwita…
Then =RAP =EMPH =3SG.DAT =3SG see-**COME(O)**-IMPF person
'He saw a person coming in his direction.'

b. peadiya ekwita =tuke=Ø ba-***dadi***-wa…
one person =3SG =1SG see-**GO(O)**-PRF
'I saw a person go (with two ducks he was stealing from us).'

4.2. Mparntwe Arrernte and Adnyamathanha: Australia

Wilkins (1995; 2005) describes a similar system in Mparntwe Arrernte. Like the Cavineña affixes, the inflections occur with all classes of verbs, except verbs of deictic motion. The reasons given by Wilkins & Hill (1995) for this restriction are that the information encoded by the inflectional markers is usually already encoded by the deictic roots, and also that most of the inflectional markers are themselves derived from some of the deictic roots. The suffixes, fifteen in total, encode similar information to Cavineña: a specific time of occurrence of the motion event relative to the verb's event,

a deictic direction, and some identification of the figure. Additionally, they provide information about the particular shape of the path of motion, its orientation, and about the speed of the motion event. A few examples from Wilkins (2005) involving the verb *angke* 'to speak' are provided in (40). The examples in (41-42) show that oriented motion verbs (i.e. those that encode a path component), and also verbs describing events such as 'sit down' can occur with the markers of I-AM (contrary to what is found in D-AM languages).

(40) **Mparntwe Arrernte** (Australia; Wilkins, 2005)

a. angk-***intye***-ke
speak-**DO.COMING**-PC
'spoke while coming this way.'

b. angke-***nhe***-ke
speak-**DO.PAST**-PC
'spoke while going past.'

c. angk-***artn***.alpe-ke
speak-**Quick:DO&GO.BACK**-PC
'Quickly spoke and then went back.'

d. angke-***ty.intye***-ke
speak-**DO.on.ARRIVAL**.of-PC
'Spoke to Z as soon as Z arrived.'

(41) … ahelhe-ke anteme itne irrpe-***ty.alpe***-ke.
ground-DAT now 3PLS enter-**GO.BACK&DO**-PC
'they (the ancestral caterpillars) went back (to Emily Gap) and now entered the ground (there).' (lit. entered after having returned).

(42) a. Alhere-le re arelhe ikwere-nge an-***intye***-ke
creek.bed-LOC 3SGS woman SGDAT-ABL(WITH) sit-**DO.COMING**-PC
'He stopped and sat in the creek-bed with that woman on his way coming here.'

b. Ampe re ulyentye-le inte-***ty.antye***-ke uterne-ketye-nge
child 3SGS shade-LOC lie.down-**DO.UPWARDS**-PC sun/heat-AVER-ABL
'The child lay down from shade to shade while going upwards lest (he suffer from) the hot sun.' (i.e. the child is climbing a hill and keeps getting tired out from heat and exertion.)

Tunbridge (1988) describes similar types of affixes with a similar range of features in Adnyamathanha, a language also spoken in central Australia. Adnyamathanha displays ten affixes, each one encoding motion and a number of other semantic features found in Mparntwe Arrernte, but also in Cavineña. These semantic features include a particular temporal relation between the two events involved, and a deictic orientation for the added motion event. This is illustrated with the examples in (43) below. The affix -*vara*-, in (43a) marks a motion event which occurs prior to the verb's event and is

directed away from the speaker. It contrasts with the suffix *-mana-* in (43b), which encodes prior motion directed toward the speaker, and *-wandha-* in (43c) which encodes motion directed away from the speaker but occurring subsequently to the verb's event.

(43) **Adnyamathanha** (Tunbridge, 1988: 270-274)

a. anha-nga=idla wandi-***vara***-ndy-adna=angu
there-AT=EMPH camp-**GO AND**-PROB-3PLNOM=CONF
'The would go and camp there.'

b. mai ngarlku-***mana***-angg-athu
food eat-**COME AND**-PERF-1SG-ERG
'I came and ate the food.'

c. artu-nga veldha marli-***wandha***-anggu
woman-ERG clothes wash-**AND LEAVE**-PERF
'The woman washed the clothes and cleared off.'

Adnyamathanha AM affixes may also provide information about the aspectual contour of the event lexicalised by their host verb. Thus, the language has a number of affixes which also specify whether the event lexicalised by the verb occurs once, repetitively or continuously over the duration of the AM. The affixes in (44) all encode a motion event concomitant to that expressed by the verb. The affixes *-ndhena-* and *-nali-* incorporate ventive deixis, while *-nangga-* incorporates itive deixis. Each also aspectually specifies the events of the verb as punctual, continuous or iterative. In (44a), the event lexicalised by the verb *ngarlku* 'to eat' is understood to have happened only once during the motion event. The event described by the verb *witi* 'to pierce' in (44b) is marked as occurring continuously over the duration of the motion event. Finally, the event of sitting/ staying in (44c) occurs repetitively over the course of the motion event.

(44) a. mai ngarlku-**ndhena**-k-alpurla
food eat-**ONCE COMING**-NARR-1PL-ERG
'We stopped and ate once on the way home.'

b. yarta veni witi-**nali**-angg-alu
ground very pierce-**CONT COMING**-PERF-3SG-ERG
'It poked into the ground all the way to here.'

c. ika-**nangga**-k-adna
sit-**GO ALONG**-NARR-3PL NOM
'They stayed one night in once place, another elsewhere, and so on, as they travelled.'

Quite interestingly, Adnyamathanha affixes may include some modal information in addition to the motion event and its properties. For example, the affix *-vara-* also encodes the speaker's negative opinion of the performed event. This is shown in (45) below.

(45) a. wangngu-***vara***-angg-alu-wa
tell-**GO AND**-PERF-3SG-ERG-3SG.ACC
'He went and told her (he shouldn't have).

4.3. Atsuwegi: North-America

Talmy (2000: 123) describes affixes with AM meanings occurring in Atsuwegi, a Palaihnihan language from California. Unfortunately, he does not provide any examples or detail about the form of these affixes, and succinctly defines the range of semantic features that each encodes. Atsuwegi belongs to a language family distinct from the three languages discussed above, and is spoken in a different region. Interestingly, its AM affixes incorporate the same basic range of semantic features as those found in other languages.

According to Talmy, ten AM affixes are found in Atsuwegi. They differ, like the affixes described in previous sections, with respect to the time relation they specify between the verb's event and the added motion event. Hence, these affixes either encode motion which occurs prior to the event, e.g. 'go and V', or motion which is concomitant to the verb's event. Affixes of concomitant motion additionally contrast in the deictic orientation they mark. Thus an affix with the semantics 'go Ving along' contrasts with another one encoding 'come Ving along'. Atsuwegi affixes may provide additional information about the motion event, such as its purpose or add a comitative component to it. Talmy describes an affix whose meaning is 'Ving going to meet someone', and another translated as 'V going along with someone'.

4.4. Summary

As already stated, the aim of the above description is not to construct a typology of I-AM systems. However, the overview shows that affixes of AM encode a similar nucleus of semantic features and display some common properties. Given that these occur across languages that are genetically and geographically unrelated, they can be assumed to be basic inherent features and properties of I-AM. Their inherent features, which contrast with D-AM, are listed below:

(i) I-AM markers each conventionally encode a contrastive bundle of information about the motion component.

(ii) I-AM affixes are rigidly rather than consistently associated with a specific time relation; either precedence, concomitance or subsequence.

(iii) In I-AM, the figure of motion might coincide or not with the main verb's subject. However, the identity of the figure is fixed by distinct affixes, rather than contextually derived.

(iv) I-AM systems involve more refined semantic oppositions between markers, including aspect and manner specifications, or information about the outcome of the motion event.

(v) I-AM may occur with motion verbs, including those lexicalising an orientation or direction (except deictic ones).

5. Comparative perspective

Overall, the contrasts between D-AM and I-AM are not major. Both represent non-verbal strategies for framing a verb's event with respect to a motion co-event. D-AM can even be considered to be a subpart of I-AM: it fulfils the same role as I-AM affixes with deictic path specifications, and attributes a similar array of features to the motion co-event. This overlap is not surprising since many I-AM affixes develop from verbs which originally express a deictic direction (Tunbridge, 1988; Austin, 1989; Guillaume, 2006). Guillaume (2006), for instance, shows that the two affixes *-diru* and *-eti*, respectively encoding the AM 'go permanently' and 'come permanently' in Cavineña, are grammaticalised forms of two verbs expressing the same deictic motion: *diru* 'to go permanently' and *jeti* 'to come permanently'. For Adnyamathanha, Tunbridge (1988) proposes that the AM affix *-na-* expressing motion towards the speaker is derived from the deictic motion verb *yana* 'to come'. This affix is also found incorporated in a few other AM affixes, such as *-mana,* which encodes prior ventive motion 'come and', *-nali-* 'continuously coming' or *-ndhena-* 'once coming'.

The fact that I-AM markers are paired with fixed semantic meanings while D-AM have meanings which vary depending on the pragmatic context and semantics of the verb is nevertheless analytically significant. A simple explanation can be provided to explain the differences relating to the particular path shape contributed by the motion co-event. If I-AM markers derive from lexical items, such as verbs, an element whose meaning contribution is equivalent to 'to return' is more likely to grammaticalize into an affix encoding a return-shaped path than one whose meaning is 'to go' or 'to come'. On the other hand, for deictic directionals, which primarily express a simple deictic path, return-shaped path interpretations have to be derived pragmatically. The differences relating to time relations, identity of the figure and the verb they can modify, on the other hand, cannot be accounted for in the same way. Here, I tentatively propose that differences between I-AM and D-AM highlighted in the previous sections reflect their different stages of grammaticalization and semantic reanalysis. While I-AM forms are fully grammaticalized, D-AM seems to be at an earlier stage of the process. In other words, the most prototypical systems of I-AM and D-AM represent two different points of the same spectrum. According to this account, the pragmatic variations found in D-AM contexts are therefore due to the fact that deictic directionals used to mark the motion co-event are not yet fully reanalysed as functional markers of AM and not yet paired with fixed meanings. However, deictic directional elements occur recurrently in particular contexts, in which they are paired with consistent pragmatic inferences. These consistent inferences may be cancelled if not available in a specific context, and replaced by others more relevant. This hypothesis is compatible with the view that semantic change and the construction of new meanings for a particular linguistic unit is strongly driven by pragmatics (Traugott & Dasher, 2002).

If the grammaticalization hypothesis presented above is correct, it should be possible to find AM systems that display defining properties of both I-AM and D-AM. Two languages seem to fit this description: Yidiŋ a Pama-Nyungan language of Queensland Australia, and Lowland Chontal, a Mayan language from Mexico. Yidiŋ (Dixon, 1977) displays two 'aspectual' affixes **—** *ŋaliŋ* and *ŋadan* and their phonological allomorphs **—** indicating AM directed toward or away from the deictic centre.

(46) **Yidiɲ** (Australia; Dixon, 1977)

a. yagaldʲida: ŋayu wula***ŋaliŋ*** ŋuŋgu
'I'll go and die there in Yagadyida.'

b. ɲundu:ba ɖambu:l wuna***ŋadan*** yiŋgu
'You two come and sleep here!'

Like D-AM markers, the two affixes can trigger concomitant or precedent time relations depending on context. Examples where the two markers involve concomitant relation between the time of the motion co-event and the time of the verb's event are provided in (47) and (48).

(47) bana ŋalal ɖuŋga***ŋali:n***a / bundu
water-ABS big-ABS run-**GOING**-PURP dilly-bag-ABS

ɖuŋga***ŋada:ɲ*** / gangul:l / ginda:ɖa / buruʤu:r/
run-**COMING**-PAST grey wallaby-ABS cassowary-ABS pademelon-ABS
'[The water rose and] as a result lots of water went rushing [into the camp]; [then] the dilly-bag, the grey wallaby, the cassowary and the pademelon came running [out of the camp, fleeing from the water].'

(48) bana bayi***:ldaɲ***u
water-ABS emerge-**COMING**-PST
'[During a volcanic eruption the ground split and] water came [rushing] out.'

The two verbs involved in the previous examples, and which are associated with concomitant motion, are both motion verbs. Example (49) below shows that motion verbs do not necessarily trigger this particular relation. There the verb for 'to emerge' is modified by the 'going' affix, and the motion event is interpreted to be prior to the verb's event.

(49) ŋaŋʤi bulmba:gu galiŋa:lna / bayi:***lin***a
we-SA camp-ALL go-COMIT-PURP emerge-**GOING**-PURP
'We must take [the women] to the camp, [we must] go so that we emerge [from the bush] at the people ['s camp].'

In fact, example (49) highlights a property of these Yidiɲ affixes which is more typical of I-AM: they can express precedent or subsequent motion even with verbs already encoding motion and orientation. Another similarity these morphemes have with I-AM affixes is that they seem to be part of a system involving more complex AM contrasts. Thus, Dixon (1981) describes another AM suffix *-:n-bidʲ-n* used with verbs of posture and the verb of motion 'to fall down' to express that the event or state referred to occurs in various spatial locations. An example is provided in (50).

(50) ŋaŋɖʝ gindanuyi burgiŋ /miɲa:gu
we-SA moon-COMIT-ABS walkabout-PRES animal-DAT

wurba:ɖiŋ / miɲam ŋaŋɖʝ gadaŋ
look.for-PRES animal-ABL we-SA come-PRES

ɲinaŋadaŋ / wuna***:nbidʝ*** wurmba
sit-COMING-PRES lie.down-**:NBIDʝ**-PRES asleep-ABL

wunaŋ
lie down-PRES
'We go walkabout by moonlight, looking for animals. We come home after [hunting] animals, come to stay at home, lie down **anywhere**, lie down sleeping.'

Lowland Chontal (O'Connor, 2007) displays four I-AM suffixes, all encoding some kind of deictic information. O'Connor describes them as markers of AM (which according to the dichotomy assumed here would correspond to I-AM) but they also display some canonical properties of D-AM. Two markers encode an additional motion, deictically directed away from the deictic centre. The main distinction between the two seems to be with the point of the path of motion they deictically anchor. The first one, the andative, is translated as 'to leave here and do V', while the second one, the dislocative, is translated as 'to go there and do V'. The andative seems therefore to be bounded at the start of the path, while the dislocative seems to presuppose that the event described by the verb will occur in a location distinct from the deictic centre, but no departure.

(51) **Lowland Chontal** (Mexico; O'Connor, 2007)

a. layx'api pang-***ix***-pa Estados Unidos
my.sister live-**AND**-PFV.SG states united
'My sister went and lived in the US.'

b. Jaape ay-'ma jo k'incho-***ta*** jaape.
Where depart-IMPRF.SG or gather.firewood-**DLOC**.SG where
'He doesn't go anywhere, not to go and fetch firewood somewhere.'

Those affixes cannot occur with motion, which is a feature more characteristic of D-AM.

The language has another pair of affixes described as markers of ventive AM, the venitive and the cislocative. The distinction between the two seems to be with the verbs that they can modify. The cislocative is restricted to a small class of deictic verbs, while the venitive applies to most verbs, even motion verbs. Both markers always involve a concomitant relation between the events described. Two examples are provided below.

(52) mu-***ñay***-wa sage pa'-na-pa' sajpe.
sink-**VEN**-PROG.SG 3S come-TERM-PFV.PL 3P
'Here he comes descending, here they come.'

(53) joypa cho-***lyu***-pa la'i ma-k'e-ta sa=yma'.
already rise-**CLOC**-PFV.SG bread cook-CAUS-DLOC.SG DEM=2S.AGT
'Once the bread has risen, you go off and bake it (in the local oven).'

The examples provided by O'Connor for the venitive marker all involve motion verbs, but assuming, as she claims, that it occurs also with non-motion verbs to encode a concomitant motion co-event, it seems to belong to the I-AM category. The cislocative, however, displays the main features of deictic directionals, and does not seem to encode AM. Lowland Chontal, therefore, shows that elements involved in a similar system may be on different paths of grammaticalization into AM. Some, such as the cislocative, may even simply be deictic directionals.

6. Conclusion

This small survey has shown that in addition to their primary function as ventive and itive path markers, deictic directionals may also be used as a non-verbal strategy to encode motion co-events. The relation between deictic directionals and Associated Motion is found in a number of African languages, where it presents approximately similar properties. It is not yet clear how widespread the phenomenon is, as this is work in progress. According to research so far it seems to be quite geographically concentrated. Although contact might play an important role in its diffusion, the phenomenon is similar to the type of AM reported in Australian and American languages, and might even be considered a subtype of AM, representing a different stage of grammaticalization. The similarity between the two categories alternatively points toward a cross-linguistic tendency for deictic path encoding elements to grammaticalize into markers of AM, a tendency which may develop from their function of locating or framing events in relation to the speech participants. It also shows that AM might not be a remote phenomenon, or just an areal feature of Australian and Amerindian languages, but a cross-linguistic strategy to package multiple events on a par with multi-clausal constructions (subordination, coordination, clause-chaining), Serial Verb Constructions or verb-compounding.

References

Alamin, S., G. Schneider-Blum & G. J. Dimmendaal (2012) Finding your way in Tima in Mietzner, A. & U. Claudi (eds.) *Directionality in Grammar and Discourse: Case Studies from Africa*. Köln: Rüdiger Köppe Verlag.

Andersen, T. (1988) Consonant alternation in the verbal morphology of Päri in *Afrika und Ubersee*, 71: 63-113

Austin, P. K (1989) Verb Compounding in Central Australian Languages. *La Trobe Working Papers in Linguistics* ed. by Kate and Graham Scott Burridge, vol.2, pp. 43-71. Melbourne: La Trobe University.

Beavers, J. (2008) On the nature of goal marking and delimitation: Evidence from Japanese. *Journal of Linguistics*, 44, pp. 283-316. Cambridge: University Press.

Beavers, J., B. Levin & S.W Tham (2009) The typology of motion expression revisited. *Journal of Linguistics*, 46 (02), pp. 331 – 377. Cambridge: University Press.

Belkadi, A. (2014) Verb meaning and deictic paths in Taqbaylit Berber. In, A. Belkadi, K. Chatsiou & K. Rowan (eds.) *Proceedings of Conference on Language Documentation and Linguistic Theory 4*. School of Oriental and African Studies: London.

Belkadi, A. (In Press) Deictic directionality and Space in Berber: A typological survey of the semantics of =d and =nn in *Corpus*. Base, Corpus, Langage, UMR 6039.

Bentolila, F. (1969) Les modalités d'orientation du procès en berbère. *La Linguistique, 1*.

Bourdin, P. (2006) The marking of directional deixis in Somali How typologically idiosyncratic is it? Voeltz, F.K. Erhard (ed.) *Studies in African Linguistic Typology*. Philadelphia, USA: John Benjamins Publishing Company.

Claudi, U. (2012) Who moves, and why? Somali deictic particles in in Mietzner, A. & U. Claudi (Eds) *Directionality in Grammar and Discourse: Case Studies from Africa*. Köln: Rüdiger Köppe Verlag.

Dixon, R. M. W. (1977) *A grammar of Yidiny*. Cambridge: Cambridge University Press.

El Mountassir, A. (2000) Langage et espace. Les Particules d'orientation –d/-nn en berbère (tachelhit). In S. Chaker & A. Zaborky (eds) *Études berbères et chamito-sémitiques: mélanges offerts à Karl-G. Prasse.* Peeters publishers.

Evans, V. (2013) *Language and Time: A Cognitive Linguistics Approach*. Cambridge: Cambridge University Press.

Frajzyngier, Z. (1989) *A Grammar of Pero*. Berlin: D. Reimer.

Fortis, J.M. & B. Fagard (2010) *Space and Language*, Leipzig summer school in typology. Part V – Deixis.

Guillaume, A. (2009) Les suffixes verbaux de 'mouvement associé' en cavineña, in *Faits de Langues, Les Cahiers, N.1*. OPHRYS.

Heath, Jeffrey. 2005. *Grammar of Tamashek (Tuareg of Mali)*. Berlin: Mouton de Gruyter.

Heine, B., U. Claudi, & F.Hünnemeyer (1991) *Grammaticalization: A conceptual framework.* Chicago: University of Chicago Press.

Heine, B. & T. Kuteva (2002) *World lexicon of grammaticalisation*. Cambridge: Cambridge University Press.

Koch, H. (1984) The category of 'associated motion' in Kaytej. *Language in Central Australia* 1: 23–34.

Kossmann, M. (2011) *A Grammar of Ayer Tuareg (Niger)*. Köln: Rüdiger Köppe Verlag.

Kossmann, M. (2013) *A Grammatical Sketch of Ghadames Berber (Libya).* Köln: Rüdiger Köppe Verlag.

Lakoff, G. & M. Johnson (1981) *Metaphors we live by*. Chicago: the University of Chicago Press.

Mettouchi, A. (2011) The Grammaticalization of directional clitics in Berber, paper presented at the workshop 'Come and Go off the grammaticalization path', convened by J. van der Wal and M. Devos at the 44th Annual meeting of the Societas Linguistica Europaea.

Mietzner, A. (2012) Spatial orientation in Nilotic languages and the forces of innovation in Mietzner, A. & U. Claudi (eds.) *Directionality in Grammar and Discourse: Case Studies from Africa*. Köln: Rüdiger Köppe Verlag.

O'Connor, L. (2007) *Motion, Transfer and Transformation: The grammar of change in Lowland Chontal*. Studies in Language Companion Series, Vol. 95. John Benjamins Publishing.

Talmy, L. (2000) *Towards a Cognitive Semantics*. MIT Press: Cambridge, Mass.

Traugott, E. & R.B. Dasher (2002) *Regularity in Semantic Change*. Cambridge: University Press.

Tunbridge, D. (1988) Affixes of motion and direction in Adnyamathanha in P.K. Austin (ed.) *Complex Sentence Constructions in Australian languages*. Typological studies in language 15. John Benjamins Publishing Company.

Weber, D. J. (1989). *A Grammar of Huallaga (Huánuco) Quechua*. Berkeley: University Press.

Wilkins, D. & D. Hill (1995) When GO means COME: Questioning the Basicness of Basic Motion Verbs. *Cognitive Linguistics* 6.2/3,209-259. La Hague: Mouton de Gruyter.

Wilkins, D. (2006) Towards an Arrernte grammar of space in S.C. Levinson and D.P. Wilkins (eds.) *Grammars of space: explorations in cognitive diversity*. Cambridge: University Press.

SOAS Working Papers in Linguistics Vol. 17 (2015): 77-95

On Wilmsen on the development of postverbal negation in dialectal Arabic

Christopher Lucas
cl39@soas.ac.uk

1. Introduction

In a recent article and subsequent monograph, Wilmsen (2013a, 2014) advances a novel proposal regarding the genesis of dialectal Arabic negative constructions that include a postverbal enclitic *-š* morpheme. He argues against two commonly held positions: i) that the *-š* morpheme is grammaticalized from *šayʾ* 'thing'; and ii) that the various types of negative construction observed in dialectal Arabic are the result of a crosslinguistically common set of cyclical changes in the expression of negation, known as 'Jespersen's cycle'. Instead, Wilmsen proposes that *-š* came to function as a negator through reanalysis of the interrogative particle *ši*, which does not, according to Wilmsen, derive from *šayʾ*.

The purpose of the present work is to offer a critique of Wilmsen's proposals, and in so doing to address some broader methodological issues for work on syntactic reconstruction, in particular the role that comparative data from unrelated languages have to play in deciding between competing reconstructions.

The paper is structured as follows. Section 2 provides data on the basic verbal negative constructions that occur in dialectal Arabic. Section 3 outlines the consensus view on the evolution of Arabic negation, as well as Wilmsen's alternative hypothesis. Section 4 discusses Wilmsen's criticisms of the consensus view, while section 5 evaluates Wilmsen's alternative, concluding that the consensus view is better supported by the available evidence. Given this conclusion, section 6 seeks to clarify the actual historical relationship between *šayʾ* 'thing', negative *-š* and interrogative *ši/-š*. Section 7 concludes the paper.

2. Negation in contemporary dialectal Arabic

Dialects in which the basic means of negating a verbal predicate involves the enclitic *-š* morpheme are spoken in one large contiguous area stretching across coastal North Africa, from Morocco into the southwestern Levant, and separately also in much of Yemen, and possibly parts of Oman. In the vast majority of this area, *-š* cannot function as the sole exponent of negation, but forms part of a bipartite negation construction together with a preverbal element *mā*,[1] as illustrated in (1).

(1) **Cairene Egyptian Arabic** (Woidich 2006: 366)
ma-basʾal-ak-**š**^{i} ʿamalt^{i} walla **ma**-ʿamalt-^{i}**š**
NEG-ask.IMPF.1SG-2MSG-**NEG** do.PRF.2MSG or **NEG**-do.PRF.2MSG-**NEG**
'I'm not asking you now whether you did it or not.'

[1] The length of the vowel in this element varies from dialect to dialect. I refer to it pandialectally as *mā*. Similarly, while *šayʾ* 'thing' is realized in almost all dialects where it is retained as a nominal element as *šī* or *šē(y)*, I refer to it pandialectally as *šayʾ*.

However, in the dialects of a few much smaller, non-contiguous regions within these two large areas, it is also possible for enclitic *-š* to function as the sole exponent of negation in at least some contexts, as illustrated in (2).

(2) **Hebron Palestinian Arabic** (Seeger 1996: 4)
šū lli biddak īyā baʿṭī-k īyā bass
what REL want.2MSG 3MSG give.IMPF.1SG-2MSG 3MSG only

uskut w-tiḥkī-**š**
be.silent.IMP.MSG and-speak.IRR.2SG-**NEG**
'Whatever you want I will give you, just be quiet and don't speak.'

Purely postverbal negation of this kind has thus far been reported in six separate dialects or dialect areas: i) all of the sedentary dialects of historic Palestine, as well as those of northwestern Jordan, southwestern Syria and southern Lebanon (at least) as far as Baskinta (Abu Haidar 1979; Behnstedt 1997: 450–1; Palva 2004; Lucas 2010); ii) marginally in Cairene Egyptian (Lucas 2010); iii) the Upper Egyptian dialect described by Khalafallah (1969); iv) in Maltese, but only in prohibitives (Ambros 1998: 108); v) the Omani dialect described by Reinhardt (1894), which may however no longer exist, since *-š(i)* does not function as a negator in most of Oman today (Domenyk Eades p.c.); vi) perhaps also in the small cluster of dialects of the southern part of the Yemeni Tihama that Behnstedt (1985: 172–3) identifies as negating the existential verb *fī* with suffixed *-ši* alone, though whether this morpheme can act as the sole exponent of negation in other contexts in these dialects is not clear.

Negation of verbal predicates in all other dialects is with preverbal *mā* (or *lā*) alone, as illustrated in (3) and (4).

(3) **Damascus Levantine Arabic** (Cowell 1964: 383)
mā žarrabᵊt laʾanno **mā** kān maʿ-i waʾᵊt
NEG try.PRF.1SG because **NEG** be.PRF.3MSG with-1SG time
'I haven't tried because I haven't had time.'

(4) **Kuwaiti Gulf Arabic** (Brustad 2000: 284)
kill ʾarbaʿ w-ḫamīs […] **mā** ʾastaθgil-ha
every Wednesday and-Thursday **NEG** find.heavy.IMPF.1SG-3FSG
'Every Wednesday and Thursday – I don't find it too onerous.'

3. Approaches to the evolution of Arabic negation

No one (as far as I am aware) doubts that the oldest of these three negative constructions is the one composed of just the preverbal negator *mā* (or an alternative preverbal element *lā*). All the ancient Semitic languages, including Quranic Arabic (5), have exclusively preverbal negation.

(5) *Quran 2:9*
mā yaḫdaʿūna ʾillā ʾanfusa-hum
NEG deceive.IMPF.3MPL except self.PL.ACC-3MPL
'They only deceive themselves.'

Note also that, in the dialects where bipartite negation is usually obligatory, in the crosslinguistically conservative context of proverbs, negation with preverbal *mā* alone becomes a possibility (6).

(6) **Cairene Egyptian Arabic** (Bergman 1996: 241)
illi **ma** yḫāf min ʾaḷḷāh ḫāf minn-u
REL **NEG** fear.IMPF.3MSG from God fear.IMP.2MSG from-3MSG
'Fear him who does not fear God.'

The question, then, is how the constructions with *-š* were innovated. As noted in section 1, the general consensus is that negative *-š* represents one endpoint in the grammaticalization of *šayʾ* 'thing'. On this scenario, *šayʾ*, or some derivative thereof, is used so often as a reinforcing element in negative sentences that it is eventually bleached of all meaning and becomes, together with preverbal *mā*, simply one element in a bipartite negative construction. Numerous observers, beginning, it seems, with Gardiner (1904), have noticed the parallels between this scenario and the documented evolution of negative constructions in various Romance, Germanic and other languages. The most famous case is that of French, in which the original preverbal negator *ne* is joined by a grammaticalized form of the noun *pas* 'step' to form a bipartite negative construction. The parallels between Arabic and French become even more compelling if one assumes, like virtually all those who have commented on the matter, that the Arabic construction with *-š* alone develops out of the bipartite construction through the omission or phonetic reduction of the preverbal element *mā*. This is because a directly analogous process has occurred in colloquial French, where negation today is typically with *pas* alone.

This crosslinguistically common pattern of renewal and loss of negative elements has been known, since Dahl (1979), as 'Jespersen's cycle', naming it for the Danish linguist Otto Jespersen, who described it as follows (Jespersen 1917: 4):

> The history of negative expressions in various languages makes us witness the following curious fluctuation: the original negative adverb is first weakened, then found insufficient and therefore strengthened, generally through some additional word, and this in turn may be felt as the negative proper and may then in the course of time be subject to the same development as the original word.

A considerable body of literature has emerged in the past few decades, documenting the occurrence of changes of this sort in both European and non-European languages, and investigating such matters as the most common sources of new negative markers, the speed of the cycle and the extent to which its three basic stages tend to represent discrete phases in the history of a language, and the mechanisms which drive cyclical change of this sort (see Willis, Lucas & Breitbarth 2013 for an overview). As such, we have a wealth of comparative data to draw on when we wish to reconstruct the progression of apparently

similar developments in languages, such as dialectal Arabic, for which pre-modern documentary evidence is relatively scarce.

With this in mind, we turn to Wilmsen's proposals concerning the Arabic developments. Wilmsen's contention is that the immediate source of negative *-š* was the particle *ši* (or a reduced version *-š*) that can in certain contemporary dialects optionally be added to affirmative interrogative clauses to attenuate the force of the question, in a similar fashion to English *by any chance*. The unreduced version of this particle, typically coming in clause-final position, is a prominent feature of contemporary Levantine varieties, exemplified by Damascus Arabic in (7). It also occurs in this position in the Libyan dialect described by Owens (1984), as in (8).

(7) **Damascus Syrian Arabic** (Cowell 1964: 378)
ṭʾūmt-i ʾəžet mən ʿand ᵊl-kawwa **ši**
suit.PL-1SG come.PRF.3FSG from at the-ironer *ši*
'Have my suits come back from the cleaners?'

(8) **Eastern Libyan Arabic** (Owens 1984: 102)
il-ḥawli simīn **šī**
the-sheep fat *ši*
'Is the sheep fat?'

The same Libyan variety also exhibits the reduced version of this particle *-š*, which, like the homophonous negative morpheme, is a clitic on the predicate:

(9) **Eastern Libyan Arabic** (Owens 1984: 102)
šiftū-**š** miḥammad
see.PRF.2MPL-*š* Muhammad
'Have you seen Muhammad?'

Despite the homophony, this reduced version of the particle, whose present-day distribution appears largely restricted to the dialects of eastern North Africa (Tunisia, Libya, Malta, Egypt), is clearly not the same item as negative *-š*. This is can be seen from the example in (10) where it is clear from the context that the *-š* morpheme cannot be expressing negation.

(10) **Cairene Egyptian Arabic** (Woidich 2006: 358)
bēt abū-ya fēn walla akun-**š**^{i} ɣliṭṭ^{i} fi š-šāriʿ
house father-1SG where or be.IMPF.1SG-**š** err.PRF.1SG in the-street
'Where's my father's house? Or have I got the wrong street?'

While Wilmsen is very clear on his view of the etymology of negative *-š* ("[r]eanalysis of the polar interrogative *ši* is what produced the negator *š*, 2014: 209), it can be hard to pin down precisely how he sees this reanalysis having taken place. Consider first of all his claim (2013a: 26–7) that "Arabic dialects negating without a pre-posed *mā* with forms like *fīš* ['there is not'] and *biddīš* ['I do not want'] and *tistaḥīš* ['aren't you embarrassed'] have not lost the pre-positioned negator in a third stage of Jespersen's cycle, for, such forms have been in place as long as an enclitic *-š* has been used in interrogation and

negation" (translations added – CL).[2] This must surely be taken as amounting to the claim that *š(i)* was reanalysed as a negator in sentences not containing *mā*, i.e. affirmative (interrogative) sentences. Although he does not state this clearly, it seems that Wilmsen has in mind specifically rhetorical questions analogous to (11) as the kind of context in which a speaker could intend an affirmative polar question, while the addressee, reanalysing, parses the utterance as negative.

(11) **Andalusi Arabic** (al-ʾAhwānī 1962: 314, cited by Wilmsen 2013a: 14)
ʾsm-u ʿlyy w-hu ʾ**š** yɣty riğlayy
name-3MS on.1S and-he **-*š*** cover.IMPF.3MS leg.PL.1S
Wilmsen's (2013a: 17) suggested interpretation:
'His name is on me, but does it cover my legs?'

Potential innovative interpretation:
'His name is on me, but it does not cover my legs.'

On this scenario, the chronological sequence of possible negative constructions would therefore have been PREVERBAL>POSTVERBAL>BIPARTITE, rather than PREVERBAL>BIPARTITE>POSTVERBAL, as assumed in the traditional Jespersen-type scenario outlined above.

Various statements in Wilmsen (2014) point in a different direction, however. For example, he says (2014: 115): "it remains only to hypothesize that the negative quality of *-š* came about as a reanalysis of *the negative response* to polar questions formed with any part of speech, requiring an answer 'yes' or 'no', it being negated with *mā* with an associated cliticized *ši*, such that the *-š* began to lose its interrogative quality" (emphasis added). This confusing claim requires some picking apart. Here it seems Wilmsen is suggesting that it was not polar questions themselves (containing *ši*) that were reanalysed as negative. Instead, he is suggesting that the interlocutor's response to such questions is the locus of reanalysis: specifically, responses that are negated with *mā*, but which also contain *ši*. It is apparently this hypothesized development that the illustrations reproduced in (12) are supposed to represent.

(12) a. From Wilmsen (2014: 209):
Negator: 'is?' > 'not is'
(h(uw)(iy)a)ši? → mā (h(uw)(iy)a)ši!
'Is he/she/it?' → 'He/she/it is not'

b. From Wilmsen (2013b: 33):
huwā-š? → Lā. mā huwā-š
he-it.is ? no not he-is
'Is he?' → 'No. He isn't'

An important question to ask regarding this scenario is: why would these negative responses, which are of course declarative, not interrogative, contain interrogative *ši*?

[2] This passage is repeated with added hedges and a tweaked final clause in Wilmsen (2014: 117): "Arabic varieties negating without a pre-posed *mā* with forms like *fīš* and *biddīš* and *tistaḥīš* have not necessarily lost the pre-positioned negator in a third stage of Jespersen's Cycle, for such forms may well have been in the language throughout the entire time that the process of reanalysing interrogatives as negators has been taking place."

There seems to be no reason for thinking they should, and Wilmsen does not provide one. As such, it is unlikely that this particular hypothesis is on the right track.

A third hypothesis hinted at by Wilmsen is that it is *ši* in the function of an indefinite determiner 'some, any' that was reanalysed as a negator. This is more plausible, and was, in fact, proposed by Davies (1981) (though see Lucas 2010: 183, fn.17 for problems with this hypothesis). Wilmsen (2014: 57) suggests that "*ši* begin[s] to acquire negative connotations when occurring as part of a longer construction, *often likely remaining an interrogative*, with the pseudo-verb governing a true object: **mā ʿandak ši* + [indefinite noun] 'you don't/don't you have any [entity]' and **mā fī ši* + [indefinite noun] 'there is/is there not some [entity]'" (emphasis added). While this scenario is certainly plausible, it is hard to square with Wilmsen's basic contentions that a) the immediate ancestor of negative *-š* is specifically interrogative *ši*, and b) purely postverbal negation did not emerge through omission of preverbal *mā* in a third stage of Jespersen's cycle. Concerning (a), in the above quote Wilmsen himself indicates that *ši* is not necessarily interrogative on this scenario, and he offers no evidence to support his claim that the construction in question "often likely remain[ed] an interrogative". Concerning (b), this scenario in fact entails a classic three-stage Jespersen-style development in the expression of negation: PREVERBAL>BIPARTITE>POSTVERBAL. The only important respect in which the scenario diverges from the mainstream view of the evolution of Arabic negation concerns Wilmsen's (2014) wider proposals about the ultimate origin of the various types of *š(i)* morphemes. As noted above, Wilmsen rejects the usual derivation from *šayʾ* 'thing', and instead proposes that grammatical *š(i)* derives ultimately from the Proto-Semitic 3rd person pronouns formed with s^1 (*[s]), which, following traditional practice, Wilmsen transcribes as *š*. As explained in detail by Al-Jallad (forthcoming), this etymology is untenable, since Arabic *š* corresponds to Proto-Semitic s^2 (*[ɬ]), not s^1. For more on the generally accepted etymology *-š* < *šayʾ*, see section 5.

There is thus no single, coherent proposal that Wilmsen offers for precisely how the development he envisages – from interrogative *ši* to negative *-š* – might have proceeded. Despite this, we can nevertheless allow that the development itself is plausible in principle. The same can perhaps also be said for Wilmsen's proposed dating and siting of the change. Lucas & Lash (2010) argue that negation with *-š* most likely arose in Egypt some time between the 8th and 11th centuries CE, and then diffused through neighbouring dialects westwards across North Africa and eastwards into the southwestern Levant (positing a separate, undated development for the presence of *-š* negation in Yemen). By contrast, Wilmsen (2014: 212) suggests that negation with *š(i)* is "likely to have been present [throughout the western Arabian Peninsula and Fertile Crescent] before the 7th century AD, perhaps for as long as a millennium and a half, but certainly present around the time that Arabic-speaking Muslims arrived". The fact that negation with *-š* in the present day is absent from the vast majority of the Fertile Crescent Wilmsen (2014: 146) puts down to "later waves of bedouinization", suggesting (2014: 212) that "Arabic speakers whose dialects retained the reflexes of *ši* used in negation could have lost that feature in less than a generation" once the Islamic conquests caused them to come into contact with Peninsular Arabs whose dialects lacked this feature.

If we accept that Wilmsen's proposals for the etymology and dating of negative *-š* are plausible in principle, the question then is whether they are preferable to the existing

reconstruction, and, more broadly, what kind of criteria we should use to distinguish between competing hypotheses of this sort. We deal with this question in the next two sections: section 4 addresses Wilmsen's criticisms of the traditional view, while section 5 outlines some of the problems with Wilmsen's hypothesis.

4. Wilmsen's criticisms of the Jespersen-type reconstruction

It seems that Wilmsen has three main objections to the Jespersen-type reconstruction of the history of Arabic negation. First, that the mechanisms by which *šayʾ* '(any)thing' could come to be grammaticalized as a negator have never been made clear. Second, that this scenario depends on erroneous assumptions about the relationship between Classical Arabic and the contemporary dialects. Third, that comparisons between the Arabic developments and the history of negation in French are unhelpful. Let us consider each of these points in turn.

Wilmsen (2013a: 10, 2014: 45) claims that the assumptions about the grammaticalization of *šayʾ* '(any)thing' "are never supported by linguistically defensible mechanisms". Given this claim, it is surprising that Wilmsen chooses to critique Esseesy's (2010: 65–6) very brief presentation of this reconstruction, when all concerned are in agreement that this is "highly schematic" (Wilmsen 2013a: 9, 2014: 44) and in no way constitutes the focus of that work or of Esseesy (2009), where the same reconstruction is also briefly mentioned. By contrast, I discussed various possibilities for these developments in some detail in Lucas (2007: 416–23), a work that Wilmsen cites several times in other contexts. There I argued that one plausible scenario for the grammaticalization of *šayʾ* as a negator is via an intermediate stage as an adverb with a negative-polarity-item (NPI) distribution, that is, like English *at all*, restricted to negative, conditional and interrogative contexts, and most frequent in negative contexts. This is, of course, not so dissimilar to Wilmsen's proposal for the immediate origin of negative *-š* (see section 6 for further discussion). However, in Lucas (2007) I go on to argue that a preferable reconstruction is one which sees *šayʾ* (or a reduced form *ši*) in the function of an indefinite pronoun 'anything' as the item that was reanalysed as a negator,[3] in view of pseudo-verb structures such as (13), in which *-š* appears to function both as negator and indefinite pronoun.

(13) **Cairene Egyptian Arabic** (Woidich 1968: 40)

wi	ma-mʿanā-**š**	yikammil	taman
and	have.1PL-**NEG+anything**	complete.IMPF.IRR.3MSG	price

id-dawa
the-medicine
'and we don't have anything that could make up the cost of the medicine'

Whether this reconstruction should be taken as definitive is naturally open to debate, but it is odd that Wilmsen chooses not to engage with it at all, since it directly addresses

[3] Note that *šayʾ* can safely be assumed to have already had an indefinite-pronoun function in pre-diaspora vernacular Arabic, at least in the context of negation, since there are a number of examples of this kind in the Quran, e.g.:

(i)

wa-lā	tušrikū	bi-hi	šayʾan	
and-NEG	associate.JUSS.2MPL	with-3MSG	thing.ACC.INDEF	

'and do not associate anything with Him'
(*Quran* 4:36)

criticisms he makes (2013a: 10, 2014: 45) of Esseesy's presentation of the Jespersen-type approach and the valency-changing consequences of *šayʾ* 'anything' being reanalysed as a non-argument negator. Regarding this issue Wilmsen makes the following objection: if we suppose that *šayʾ* 'anything' is reanalysed as a non-argument negator when functioning as the object of some verb, then this reanalysis must also change the verb from transitive to intransitive in that context; but in this case, Wilmsen (2013a: 10, 2014: 45) observes, there is then "no motivation" for later also negating unambiguously transitive verbs with *š(ayʾ)*. Here it seems that Wilmsen is failing to distinguish between reanalysis and 'extension' (Harris & Campbell 1995) or 'actualization': "the gradual mapping out [in a wider range of grammatical contexts – CL] of the consequences" of a reanalysis (Timberlake 1977: 141). It is a basic observation of syntactic change that reanalyses tend to happen first in a particular 'bridging context' (Clark 1975, Heine 2002), and that the new structures often take many generations to spread to a wider set of contexts (see, e.g., Kroch 1989 on the gradual spread of *do*-support in Early Modern English, and Lucas 2010 for some suggestions for what drives extension/actualization in general).

While we do not have any documentary evidence for this kind of gradual spread with negative *š(ayʾ)* in Arabic, directly equivalent developments can be observed in other languages which have undergone, or are in the process of undergoing, Jespersen's cycle. For example, in the Romance language Venetan there is an apparently recently grammaticalized negative strengthener *gnente* 'at all' that co-occurs with the preverbal negator *no* (Poletto 2008: 72–74). At the same time, an identical form (*gnente*) continues its original function as a negative indefinite pronoun occurring in a wide range of contexts. Adverbial *gnente*, on the other hand, occurs principally with a restricted class of intransitive verbs, as in (14), and is only compatible with a direct object when this is part of an explicit contrast, as in (15).

(14) **Venetan** (Poletto 2008: 72)
No-l dorme **gnente**
NEG-3MSG sleep.PRES.3SG *gnente*
'He doesn't sleep at all.'

(15) **Venetan** (Poletto 2008: 73)
a. *No-l leze **gnente** i libri
NEG-3MSG read.PRES.3SG *gnente* the books
Intended meaning: 'He doesn't read books at all.'

b. No-l leze **gnente** libri, solo giornai
NEG-3MSG read.PRES.3SG *gnente* books only newspapers
'He doesn't read books at all, only newspapers.'

By contrast, in related varieties such as Piedmontese, where the cognate element *nen* is a fully grammaticalized negator that can function as the sole expression of negation in a clause, there is no longer any restriction according to verb type (Zanuttini 1997: 67).

This comparative evidence adds weight to what should anyway be the default assumption, given the well-established concept of actualization: if it was indeed *šayʾ* as an indefinite pronoun 'anything' that was first reanalysed as a negator, then we should expect that its

unrestricted present-day distribution would only have come about gradually, and that it was restricted initially just to intransitive verbs. So it is wrong to suggest that there are no well-understood mechanisms by which *šayʾ* '(any)thing' could have come to function as an unrestricted negator.

Wilmsen's second objection to Jespsersen-type reconstructions of the Arabic facts is that "they seem to have been adopted whole from medieval linguistic traditions about the Arabic dialects arising as corruptions of [Classical Arabic]" (Wilmsen 2013a: 10, 2014: 45). It is hard to see much justification for this claim at all, beyond the fact that Esseesy (2010: 65) includes a nominative-inflected form of *šayʾ – šayʾun* – as part of his schematization of the initial stage of Jespersen's cycle in Arabic. Wilmsen (2013: 8, 2014: 50) also cites Obler's (1990: 136) perfectly reasonable statement that "Classical Arabic, whatever its status, can in any case be taken as representative of a form of Arabic earlier than that of the modern dialects", wrongly suggesting that taking such a view entails the belief that "the Arabic dialects arose after the Arab diaspora". Believing, as most linguists working on Arabic today do, that the Arabic dialects cannot be said to descend from the Arabic variety of the Quran and pre-Islamic poetry should not blind us to the truism that early vernacular Arabic will have been in many respects more similar to Classical Arabic than the present-day dialects are. And this is especially likely to be true of features such as exclusively preverbal negation, concerning which Classical Arabic, other ancient Semitic languages, and contemporary Bedouin dialects of the Arabian Peninsula are all in agreement. In the absence of compelling evidence to the contrary, therefore, it is sensible to treat the negation system of Classical Arabic as being basically similar to that of pre-Islamic vernacular Arabic. It should go without saying that this does not entail the most likely incorrect view that pre-Islamic vernacular Arabic was identical to Classical Arabic.

Finally, Wilmsen is scathing about the comparisons between the various Arabic negative constructions and those of French, which he sees as being "dutifully mentioned by all" (2013a: 26, 2014: 117) who discuss the Arabic developments. His position seems to be that, at least in this case, comparative data from apparently similar developments in unrelated languages are actively misleading. This is an important issue, and it merits some discussion.

On the one hand, there is no doubt that linguists need to guard against bringing analytical biases derived from well-studied languages to bear on less well-studied languages (cf., e.g., Van Valin's 2010: 704 criticism of Anglo-centric syntactic frameworks in which the notion of 'subject' is taken to be universal). Haspelmath (2007: 125) is surely right that "[i]nstead of fitting observed phenomena into the mould of currently popular categories, the linguist's job should be to describe the phenomena in as much detail as possible, using as few presuppositions as possible". At the same time, no one can reasonably deny that syntactic reconstructions should, in general, be informed by familiarity with changes from the same domain in other languages. Deutscher (2006: 464) notes that, as a result of intensive work on grammaticalization in hundreds of languages over the past few decades:

> A picture of overwhelming unidirectionality has emerged, where language after language goes down similar paths of change, from concrete lexical elements to abstract grammatical markers, and often from the same lexical sources to the same grammatical elements. It is obvious why such insights are important for reconstruction, whether in Semitic or in any other language family. Just as phonological reconstruction requires an awareness of what sound changes are likely (e.g. *p* > *f* is very likely, *f* > *p* exceedingly unlikely), so does the unidirectionality of many 'grammaticalization' changes provide a framework for reconstructing morphology and syntax.

The question, then, is how to negotiate the potentially conflicting requirements of looking at each change in each language with fresh eyes on the one hand, and ensuring on the other that our reconstructions are informed by an understanding of the types of changes that comparative work should lead us to expect. My own answer to this question would be that all possible language-internal data must be considered when developing reconstructive hypotheses – spotting parallels with the histories of other languages is useless if it is based on a superficial inspection of the language under investigation – but whenever there is a choice to be made between two possible hypotheses or interpretations of data, a choice which cannot be decided purely by internal considerations, we ought to favour the hypothesis or interpretation which conforms most closely to patterns of change that have been observed in other languages.

5. Problems with Wilmsen's proposals

While the final stage of Wilmsen's proposed etymology of negative *-š* – that it derives from interrogative *ši* – seems plausible, it is worth noting that there do not appear to be any documented cases of a negator deriving from an interrogative particle (see, e.g., Devos & van der Auwera 2013: 232–3).[4] Given the preceding discussion, however, a more important point is that Wilmsen's proposals are not favoured by the Arabic-internal facts either. Two pieces of evidence are especially relevant here: i) the existence of dialects in which the postverbal negator has an allomorph that is unequivocally a reflex of *šayʾ*; ii) the (restricted) geographical distribution of both the bipartite (*mā…-š*) and purely postverbal (*-š*) constructions in present-day dialects. We deal with both of these points in turn in the following.

As noted in section 3, Wilmsen's proposal that the Proto-Semitic *s*[1] pronouns are the ultimate source of negative *-š* displays a disregard for, or confusion about, the details of sound correspondences in Semitic (Al-Jallad forthcoming). A similar attitude can be discerned in Wilmsen's (2014: 58, fn. 20) suggestion that we need "not concern ourselves with the vowel differences between *šayʾ*, *šay*, *šēʾ*, *šī*, and *ši*, or with the presence of the glottal stop in two of them". It should go without saying that these differences represent crucial pieces of evidence for reconstructing the chronological relationships between these various reflexes of what everyone agrees is ultimately the same basic item. For example, it is only by ignoring this phonological evidence that Wilmsen can suggest that *šayʾ* 'thing' in fact derives from the indefinite determiner use of *ši*, rather than *vice versa*. The fact is, historical monophthongization and loss of final glottal stop are well-attested features of most Arabic dialects, whereas the reverse processes are virtually unattested.

[4] Though Dryer (2009) does tentatively propose that the origin of many of the clause-final negators of central African languages could be reanalysis of clause-final question particles.

As such, the existence of dialects in which there is a non-monophthongized alternative form of the postverbal negator strongly points towards such forms being original. For example, Heath (2002: 212) records *ma…-šay* as a rare variant of *ma…-š* in two Moroccan dialects: Zagora in the south, about 400km east of Agadir, and Taounate in the north, about 70km northeast of Fes. Heath (2002: 212) also notes that *šay* "occurs more widely in M[uslim] dialects [of Morocco – CL] as an emphatic version of *ma…-š*" (see also Caubet 1993: 68). Similarly, Khalafallah (1969: 100–2) reports *šey* as being the basic negative morpheme in the Qinā-governorate dialect of Egyptian Arabic that he describes. To judge from his examples, *-šey* has an allomorph *-ši* in this variety, *-šey* occurring utterance-finally, *-ši* elsewhere. It is hard to think of a plausible interpretation of this data other than that these non-monophthongized forms are relics, and the more widespread *-š(i)* forms are derived from them, having been subject to the kind of phonetic reduction typically associated with grammaticalization. There is thus no reason to doubt that the ultimate origin of *-š* negation is a form *šayʾ*.

Turning to the geographical evidence, the problem here is that neither the bipartite nor the purely postverbal constructions have the kind of distribution in present-day dialects that we would expect if Wilmsen were correct that the purely postverbal construction is at least as old as the bipartite construction and that either or both were "present [throughout the western Arabian Peninsula and Fertile Crescent] before the 7th century AD, perhaps for as long as a millennium and a half".

As noted in section 2, the bipartite *mā…-š* construction is the default for negating verbal predicates in the sedentary dialects spoken all across coastal North Africa (including Maltese) and into the Levant, and also in much of Yemen. By contrast, negation with *-š* alone is attested only sporadically within this large region, having been reported in six separate areas: Palestine and surrounding parts of the Levant, marginally in Cairo (and presumably also much of the Delta), Upper Egypt, Malta, part of the southern Yemeni Tihama, and the late nineteenth-century Omani dialect of the Ḥajar mountains and Zanzibar (Reinhardt 1894). If, as Wilmsen (2013a: 26–7) claims, "Arabic dialects negating without a pre-posed *mā* […] have not lost the pre-positioned negator […], for, such forms have been in place as long as an enclitic *-š* has been used in interrogation and negation", then why is this purely postverbal negative construction so rarely attested as a possibility today? Wilmsen would presumably have to answer that it has been lost in all the other dialects which retain the bipartite construction. But how might this have happened? While the presence of purely postverbal negation (alongside bipartite negation) throughout Palestine and neighbouring areas of the Levant is plausibly explained as resulting from innovation of this construction (via omission of *mā*) in one sub-dialect and its subsequent diffusion throughout the rest of that region, it is not plausible to argue that an innovation resulting in the *absence* of this construction (with no new construction gained) could diffuse in the same way. Since there is also no possibility of arguing that all the dialects without the purely postverbal construction form a genetic subgroup, Wilmsen's account requires him to posit a huge number of independent parallel losses of purely postverbal negation in those dialects. While it might at first glance seem unparsimonious to posit six independent innovations of *mā*-dropping

in the dialects listed above, it is clearly extremely unparsimonious to posit countless independent parallel losses of the purely postverbal construction.[5]

Then there is the fact both the bipartite and the purely postverbal constructions are absent in the present-day dialects of the whole of Iraq and most of Syria, despite this being the region in which Wilmsen believes the construction was once widespread. As noted above, Wilmsen (2014) believes that negation with *-š(i)* was lost in these regions after the Islamic conquests as a result of contact with supposedly prestigious Peninsular dialects negating with preverbal *mā* alone. While here we at least have a mechanism that is plausible in principle, it remains the case that positing the innovation and subsequent loss of a construction in a set of dialects is less parsimonious than assuming the dialects in question never had the construction at all. The less parsimonious hypothesis should be rejected unless we have strong evidence to support it. Wilmsen offers two types of evidence on this score, but neither stand up to scrutiny.

The first line of argument that Wilmsen pursues focuses on the presence in Maltese, Cypriot Maronite Arabic and Central Asian Arabic of items derived from *qaṭṭ* '(n)ever, (not) at all' plus a suffixed *-š*. These are Maltese *qattx* 'ever',[6] Cypriot *kitš* 'nothing' and Central Asian *qaṭṭiš* 'nothing'. Wilmsen (2014: 149) views the *-š* in these items as the familiar negative morpheme, seeing in it a "remnant of an earlier negation pattern, complete with suffixed *-š*", which has otherwise been lost in Cypriot and Central Asian. In the case of these latter two dialects this analysis is untenable, given the existence of dialects in which the equivalent expression is transparently composed of two morphemes **qaṭṭ* 'no' and **šayʾ* 'thing', for example *qəṭṭ štya* 'nothing' in the Kinderib dialect of southeastern Turkey (Jastrow 2005: 115; cf. Souag 2009: 62; cf. also *qəṭṭ-aḥḥad* 'no one' in the same dialect). Maltese *qattx* should be seen as a separate development of the same original items, though the suffix cannot be derived from negative *-š* here either, since *qattx*, meaning 'ever', is non-negative and does not occur in negative sentences (Borg 2004: 389; the meaning 'never' in negative sentences is rendered by *qatt* without suffixed *-x*. See section 6 for more on non-negative *-x* in Maltese). There is thus no good linguistic evidence that negation with *-š* was formerly present in dialects of the Fertile Crescent that lack it today.

Wilmsen's second line of argument relates to the present-day distribution within Syria of dialects which at least optionally feature the bipartite (*mā…-š*) construction. Referring to Figure 1 (reproduced from Behnstedt & Woidich 2005: 101), Wilmsen (2014: 213) suggests that "the dialect islands in the Syrian steppe, where negation with *-š* is optional, look to be precisely what would be expected in a geography where a major dialect change had taken place [i.e. widespread loss of negation with *-š* – CL]: remote island holdouts that the change had passed over". This line of argument backfires, however, if we consider the dialect geography of the Syrian steppe in more detail. The information given in Figure 1 is a simplified version of that provided by Behnstedt (1997: 450–1) in his *Sprachatlas*

[5] Also relevant here is the fact that, at least in Cairene and Palestinian, there are clear restrictions on the grammatical contexts in which purely postverbal negation can appear, and these are very different in each of these two dialects – further evidence that the presence of this construction in both should be seen as the result of two independent developments. This issue is discussed in detail in Lucas (2010). Regarding the process of data collection for Palestinian in that work, Wilmsen (2014: 102) suggests that "its technique and sample size limit its reliability". Readers of Lucas (2010) may judge that issue for themselves, but note that the findings presented there are fully consistent with those of other scholars who have described the distribution of negation in the dialects of Palestine and neighbouring areas of the Levant (e.g. Obler 1975: 101, 105; Shahin 2000: 37; Seeger 2013: 148–9, p.c.; Herin 2014).

[6] Maltese <x> represents IPA /ʃ/, otherwise transcribed here as <š>.

von Syrien. In the latter work we see that, while the bipartite construction is at least optional in the oasis towns of il-Qaritēn, Palmyra and Soukhne, in many of the smaller villages surrounding these towns it is absent. It has long been known that innovations will often 'jump' from one urban settlement to the next faster than they spread through the villages in between (e.g. Trudgill 1974). The reasons for this can be complex, but a key factor is the fact that population size and density is typically at least as important a determinant of contact between settlements as mere proximity. Against this background, it seems unlikely that these oasis towns should have been conservative with regard to developments in the expression of negation while the more remote surrounding villages have been innovative. Instead, the distribution of *-š* negation in Syria is fully consistent with its being an innovation that has spread from the southwest, and not an ancient feature whose loss has gradually diffused from the east.[7]

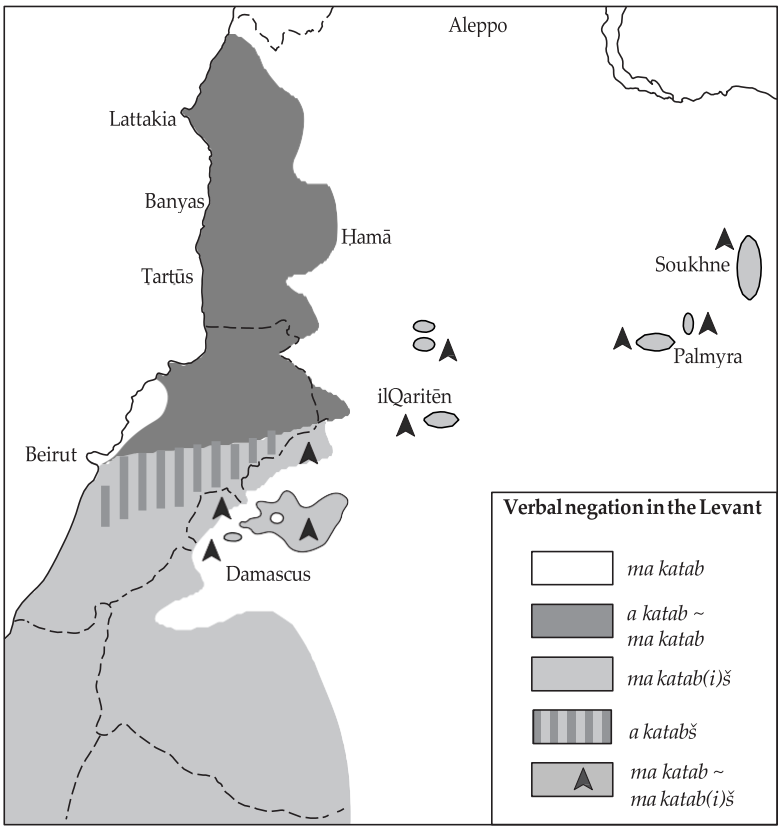


Figure 1. Distribution of negative constructions in the northwestern Levant
Source: Behnstedt & Woidich 2005: 101 (via Wilmsen 2014: 153)

[7] Mention should be made here of the non-contiguous distribution of the construction in the dialects north of Damascus. That Damascus itself and the area immediately surrounding it have resisted adoption of this external feature can be attributed to the considerable local prestige of this dialect (the same argument applying to the dialect of Beirut). The thin band of *š*-less dialects running parallel to the border with Lebanon, on the other hand, corresponds closely to the area of the Qalamoun mountains – a region where linguistic conservatism is to be expected.

In summary, the Arabic-internal linguistic and geographical facts do not support Wilmsen's novel reconstruction of negative *-š*. As such, the conventional Jespersen-type reconstruction remains the most plausible account that has yet been offered of the emergence of this morpheme. In the following section I offer some more detailed suggestions for what this account entails, and how we should understand the diachronic relationship between negative *-š* and interrogative *ši*/*-š*.

6. How are *šayʾ* '(any)thing', negative *-š* and interrogative *ši*/*-š* related?

In order to understand more clearly how *šayʾ* '(any)thing', negative *-š* and interrogative *ši*/*-š* are related, we first need to clarify the true status of the non-negative *-š* morpheme, illustrated in (9) and (10), that Wilmsen sees as being restricted to interrogative contexts. There is no good reason to doubt that this morpheme, like the homophonous negative morpheme, derives from *šayʾ*, and is a reduced form of the usually clause-final particle *šī*/*ši*, illustrated in (7)–(8). The fact that the unreduced version of this morpheme may, in at least some dialects, also occur directly after the predicate, as in (16), presents an obvious route via which it could have become a verbal enclitic.

(16) **Damascus Syrian Arabic** (Cowell 1964: 378)
ʿam-təʾṣod ***ši*** ʾənn-i kazzāb
PROG-intend.IMPF.2MSG *ši* COMP-1SG liar
'Are you suggesting I'm a liar?'

Grammars of dialects that feature either the free or clitic form of this item (or both) all seem to report them as being restricted to polar questions (either main-clause or embedded), with some authors (e.g. Woidich 2006: 306) adding that the effect is to make the question more tentative than it would otherwise be, as with English *at all*, *in any way*, *by any chance*. While it is perfectly plausible that a specialized question particle could grammaticalize from an indefinite pronoun such as *šayʾ*, evidence from the grammaticalization of indefinite pronouns as adverbs in other languages (e.g. Welsh; Willis 2013: 245), leads us to expect that there will have been an intermediate stage in which a generalizing particle of this kind had an NPI distribution, occurring also in negative and conditional contexts. There is, in fact, some evidence that this was the case. We saw above that in a number of Moroccan dialects emphatic negation may be expressed with *ma…-šay* rather than the more normal *ma…-š*. Similarly, Watson (1993: 261) says of Ṣanʿānī Yemeni Arabic that "where *šī* is used in place of *-š*, it often emphasises the negative element", giving the following example:

(17) **Ṣanʿānī Yemeni Arabic** (Watson 1993: 261)
baʿdma tzawwajt mā ḫarajt **šī**
after marry.PRF.1SG NEG leave.PRF.1SG *šī*
'After I got married, I didn't go out at all.'

Deciding whether *šī*/*šay* in constructions such as this should synchronically be considered part of a bipartite negative construction or whether it is instead merely a non-negative NPI adverb like English *at all* or *one bit* is a complex and controversial matter,[8] and one that need not concern us here. More instructive is that, at least in Maltese, non-negative -

[8] As such, Wilmsen (2013: 12–3, 2014: 68–9) is too hasty in declaring the three written examples he is able to find of Andalusi negative constructions containing *šī* to be "unequivocal" evidence of bipartite negation in that variety.

š (orthographically *-x*) occurs not only in root interrogative clauses as in (18) and embedded interrogative clauses introduced by the conditional particle *jekk* 'if, whether' as in (19), but also occasionally in true conditional clauses, as in (20).

(18) **Maltese** (*MLRS v.2.0 religion_lh294)*

Intom-**x**	taraw	il-postijiet	fejn	għandhom	jitwaħħlu
2PL-***š***	see.IMPF.2PL	the-places	where	should.3PL	attached.IMPF.3PL

l-lampi?
the-lamps
'Do you see the places where the lamps should be attached?'

(19) **Maltese** (*MLRS v.2.0 parl10202)*

Xtaqt	nistaqsi	jekk	sibtu-**x**	li	kien
want.PRF.1SG	ask.IMPF.1SG	if	find.PRF.2PL-***š***	COMP	be.PRF.3MSG

hemm	xi	titjib	jew	jekk	l-affarijiet	baqgħu-x
there	any	improvement	or	if	the-matters	remain.PRF.3PL-***š***

kif	kienu	jew	jekk	marru-**x**	lura
how	be.PRF.3PL	or	if	go.PRF.3PL-***š***	back

'I wanted to ask if you've found […] that there has been any improvement or if matters have remained as they were or gone backwards.'

(20) **Maltese** (*MLRS v.2.0 press_orizzont64068)*

Jekk	trid-**x**	żwieġ	mingħajr	skadenza	isma'
if	want.IMPF.2SG-***š***	marriage	without	deadline	listen.IMP.SG

x'jgħid-lek	l-arċipriet	u	ħalli-k
what-say.IMPF.3MSG-to.2SG	the-archpriest	and	keep.IMP.2SG-2SG

mill-ħmerijiet.
from.the-nonsense
'If you want a marriage without a deadline, listen to what the archpriest tells you and keep away from mischief.'

As argued in Lucas (2010), the neatest explanation for these facts is that *šayʾ* first grammaticalized as a generalizing adverb with a weak NPI-distribution, similar to English *at all*. In many dialects, such as Eastern Libyan (8)–(9) or Damascene (7), (16), the distribution of this adverb (realized as *šī*, *-ši*, or *-š*) has narrowed just to polar questions. But in Maltese it seems to retain something like its original distribution, albeit it is far more frequent in embedded questions than in conditional or root interrogative clauses.

Regarding the question of exactly which item was reanalysed as the second element of a bipartite negative construction when this innovation first occurred in Arabic, there are, as noted in section 4, two obvious alternatives: either *šayʾ* in its original indefinite pronoun function ('anything') or the grammaticalized generalizing adverb *šayʾ* ('at all') just discussed. The latter should probably be seen as the default hypothesis, but see Lucas

(2007: 418–22) for some arguments in favour of the former. A comparative illustration of Wilmsen's and the Jespersen-type reconstructions of these items is given in Figure 2.

7. Conclusion

For a variety of reasons, Wilmsen's (2013a, 2014) alternative reconstruction of the development of negative *-š* in Arabic dialects cannot be sustained. Therefore the conventional reconstruction, according to which Arabic has undergone Jespersen's cycle, remains the most plausible account to date. It is important to note that the evidence against Wilmsen's reconstruction that we have reviewed here is at least as much Arabic-internal as it is comparative, despite Wilmsen's implication that the conventional reconstruction is led astray by superficial comparisons with the history of French. Nevertheless, in examining Wilmsen's proposals and his criticisms of the conventional reconstruction, we have had to confront an important but complex methodological problem for syntactic reconstruction in general: how to balance the need for unbiased appraisal of the relevant language-internal facts with the obvious requirement that reconstructions should be informed by an understanding of historical processes that are known to be crosslinguistically common. The answer that I have suggested here is that while linguistic facts internal to the reconstruction in question will always trump any external considerations, whenever we are faced with a choice between two or more reconstructions or analyses that cannot be resolved solely by internal linguistic data, the rational move is to favour the one that most resembles diachronic developments that are known to be common in the histories of other languages.

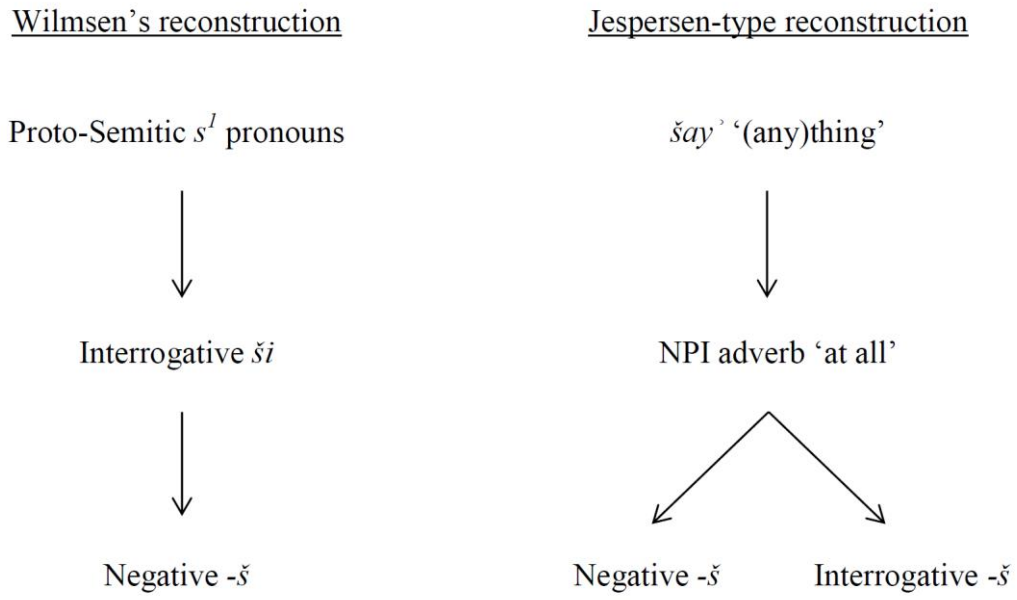


Figure 2. Alternative reconstructions of negative and interrogative *-š* in Arabic

Primary sources

MLRS v.2.0 = The Maltese Language Resource Server Corpus, version 2.0 BETA, Institute of Linguistics, University of Malta

Quran = The Holy Quran

References

Abu Haidar, F. (1979) *A study of the spoken Arabic of Baskinta*. Leiden: Brill.

al-ʾAhwānī, ʿA. (1962) ʾamθāl al-ʿāmma fī al-andalus. In ʿAbd al-Raḥman Badawī (ed.) *ʾilā ṭāha ḥusayn fī ʿīd mīlādihi as-sabʿīn: dirāsāt muhdāt min ʾaṣdiqāʾihi wa talāmīðih (= Mélanges Taha Husain offerts par ses amis et ses disciples à l'occasion de son 70ième anniversaire)*, pp. 235–367. Cairo: Dār al-Maʿārif.

Al-Jallad, A. (forthcoming) What's a caron between friends? (Review Article of Wilmsen, D. (2014) *Arabic indefinites, interrogatives, and negators: A linguistic history of western dialects*. Oxford: Oxford University Press.). *Journal of Linguistics* 51,

Ambros, A. (1998) *Bonġornu, kif int: Einführung in die maltesischen Sprache*. Wiesbaden: Reichert.

Behnstedt, P. (1985) *Die nordjemenitischen Dialekte, Teil 1: Atlas*. Wiesbaden: Dr. Ludwig Reichert.

Behnstedt, P. (1997) *Sprachatlas von Syrien*, vol. 1: *Kartenband*. Wiesbaden: Harrassowitz.

Behnstedt, P & M. Woidich. (2005) *Arabische Dialektgeographie: Eine Einführung*. Leiden: Brill.

Bergman, E. M. (1996) *ma-tirʿaf xeeri:* Verbal negation in Egyptian and Moroccan Arabic proverbs, in Eid M. & D. Parkinson (eds.) *Perspectives on Arabic Linguistics IX*, 223–45. Amsterdam: John Benjamins.

Borg, A. (2004) *A comparative glossary of Cypriot Maronite Arabic (Arabic–English)*. Leiden: Brill.

Brustad, K. (2000) *The syntax of spoken Arabic: A comparative study of Moroccan, Egyptian, Syrian and Kuwaiti dialects*. Washington DC: Georgetown University Press.

Caubet, D. (1993) *L'arabe marocain*. Paris: Peeters.

Clark, H. (1975) Bridging. In R. C. Schank & B. L. Nash-Webber (eds.) *Theoretical issues in natural language processing*, pp. 169–74. Cambridge, Mass.: Yale University Mathematical Social Sciences Board.

Corriente, F. (1977) *A Grammatical sketch of the Spanish Arabic dialect bundle*. Madrid: Instituto Hispano-Árabe de Cultura.

Cowell, M. W. (1964) *A reference grammar of Syrian Arabic*. Washington DC: Georgetown University Press.

Dahl, Ö. (1979) Typology of sentence negation. *Linguistics* 17, 79–106.

Davies, H. T. (1981) *Seventeenth-century Egyptian Arabic: A profile of the colloquial material in Yūsuf al-Širbīnī's Hazz al-Quḫūf fī Šarḥ Qasīd Abī Sādūf*. dissertation.

Deutscher, G. (2006) Review of Rubin (2005).

Esseesy, M. (2009) Reanalysis. In K. Versteegh, M. Eid, M. Woidich & A. Zaborski (eds.) *Encyclopedia of Arabic language and linguistics, Volume IV*, pp. 37–43. Leiden: Brill.

Esseesy, M. (2010) *Grammaticalization of Arabic prepositions and subordinators: A corpus-based study*. Leiden: Brill.

Gardiner, A. H. (1904) The word *iwn3*. *Zeitschrift für Ägyptische Sprache und Altertumskunde* 41, 130–5.

Harris, A. C. & L. Campbell. (1995) *Historical syntax in Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.

Haspelmath, M. (2007) Pre-established categories don't exist: consequences for language description and typology. *Linguistic Typology* 11, 119–32.

Heath, J. (2002) *Jewish and Muslim dialects of Moroccan Arabic*. London and New York: Routledge.
Heine, B. (2002) On the role of context in grammaticalization. In I. Wischer & G. Diewald (eds.) *New reflections on grammaticalization*, pp. 83–101. Amsterdam: Benjamins.
Herin, B. (2014) Salt, Dialect of. In L. Edzard & R. de Jong (eds.) *Encyclopedia of Arabic language and linguistics online*. Brill Online.
Jastrow, O. (2005) *Glossar zu Kinderib (anatolisches Arabisch)*. Wiesbaden: Harrassowitz.
Jespersen, O. (1917) *Negation in English and other languages* (Historisk-filologiske Meddelelser 1). Copenhagen: A. F. Høst.
Khalafallah, A. (1969) *A desriptive grammar of Saɛi:di colloquial Arabic*. The Hague: Mouton.
Kroch, A. S. (1989) Reflexes of grammar in patterns of language change. *Language Variation and Change* 1, 199–244.
Lucas, C. (2007) Jespersen's Cycle in Arabic and Berber. *Transactions of the Philological Society* 105, pp. 398–431.
Lucas, C. (2010) Negative *-š* in Palestinian (and Cairene) Arabic: Present and possible past. *Brill's Annual of Afroasiatic Languages and Linguistics* 2, 165–201.
Lucas, C. & E. Lash. (2010) Contact as catalyst: The case for Coptic influence in the development of Arabic negation. *Journal of Linguistics* 46, 379–413.
Obler, L. (1990) Reflexes of Classical Arabic *šay'un* 'thing' in the modern dialects. In J. Bellamy (ed.) *Studies in Near Eastern culture and history in memory of Ernest T. Abdel-Massih*, pp. 132–52. Ann Arbor: University of Michigan.
Owens, J. (1984) *A short reference grammar of Eastern Libyan Arabic*. Wiesbaden: Harrassowitz.
Palva, H. (2004) Negations in the dialect es-Salṭ, Jordan. In M. Haak, R. de Jong & K. Versteegh (eds.) *Approaches to Arabic dialects: A collection of crticles presented to Manfred Woidich on the occasion of his sixtieth birthday*, pp. 221–36. Leiden: Brill.
Poletto, C. (2008) On negative doubling. *Quaderni di Lavoro ASIt* 8, 57–84.
Reinhardt, C. (1894) *Ein arabischer Dialekt gesprochen in 'Oman und Zanzibar: Nach praktischen Gesichtspunkten für das Seminar für orientalische Sprachen in Berlin*. Berlin: Spemann.
Rubin, A. (2005) *Studies in Semitic grammaticaliztion*. Winona Lake, Indiana: Eisenbrauns.
Seeger, U. (1996) *Der arabische Dialekt von il-Xalīl (Hebron)*. Magister dissertation, Universität Heidelberg.
Souag, L. (2009) Siwa and its significance for Arabic dialectology. *Zeitschrift für Arabische Linguistik* 51, 51–75.
Timberlake, A. (1977) Reanalysis and actualization in syntactic change. In C. N. Li (ed.) *Mechanisms of syntactic change*, pp. 141–77. Austin: University of Texas Press.
Trudgill, P. (1974) Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society* 3, 215–46.
Van Valin, R. D. (2010) Role and reference grammar as a framework for linguistic analysis. In B. Heine & H. Narrog (eds.) *The Oxford handbook of linguistic analysis*, pp. 703–38. Oxford: Oxford University Press.
Watson, J. C. E. (1993) *A syntax of San'ani Arabic*. Wiesbaden: Harrassowitz.

Willis, D. (2013) Brythonic Celtic. In D. Willis, C. Lucas & A. Breitbarth (eds.) *The history of negation in the languages of Europe and the Mediterranean, Volume I: Case studies*, pp. 239–98. Oxford: Oxford University Press.

Willis, D, C. Lucas & A. Breitbarth. (2013) Comparing diachronies of negation. In D. Willis, C. Lucas & A. Breitbarth (eds.) *The history of negation in the languages of Europe and the Mediterranean, Volume I: Case studies*, pp. 1–50. Oxford: Oxford University Press.

Wilmsen, D. (2013a) The interrogative origin of the Arabic negator *-š*: Evidence from copular interrogation in Andalusi Arabic, Maltese, and modern spoken Egyptian and Moroccan Arabic. *Zeitschrift für Arabische Linguistik* 58, 5–31.

Wilmsen, D. (2013b) Negation with post-positive *-š* in Arabic dialects: An original feature not the end of a cycle. Paper presented at the 10th conference of the Association Internationale de Dialectologie Arabe, Doha, Qatar.

Wilmsen, D. (2014) *Arabic indefinites, interrogatives and negators: A linguistic history of western dialects*. Oxford: Oxford University Press.

Woidich, M. (1968) *Negation und negative Sätze im Ägyptisch-Arabischen*. PhD dissertation, Ludwig-Maximilians-Universität, Munich.

Woidich, M. (2006) *Das Kairenisch-Arabische: Eine Grammatik*. Wiesbaden: Harrassowitz Verlag.

Zanuttini, R. (1997) *Negation and clausal structure: A comparative study of Romance languages*. New York: Oxford University Press.

Historical Linguistics and Sociolinguistics

SOAS Working Papers in Linguistics Vol. 17 (2015): 99-111

Lingua Franca – a not so simple pidgin

Joanna Nolan
129216@soas.ac.uk

1. Introduction

The existence and documentation of Lingua Franca, allegedly the earliest European pidgin, is subject to much debate. Linguists from Schuchardt (1909) onwards offer contradictory theories regarding its place and date of origin and its evolution, with later research (Minervini 1996, Selbach 2008) questioning whether the evidence suggests that the contact language spoken across the Mediterranean and in North Africa was instead an L2 approximation of Italian, or one of its dialects. I would posit that Lingua Franca was an established pidgin, widely used in the Barbary regencies in North Africa from the 16th-19th centuries, and that the particular population and linguistic makeup of the region was key to the pidgin's spread and duration.

Lingua Franca was an almost exclusively oral language which originated presumably in late medieval times for speakers of, amongst others, Italian and its dialects (Venetian and Genoese, in particular), Spanish, Provençal, Portuguese, Arabic, Turkish and Berber. Given its oral (and, now, dead) status, the evidence for Lingua Franca's existence is, by definition, circumscribed and limited. The corpus is minimal. As Whinnom (1977) states: "our specimens are especially sparse and tantalizing, spread across the centuries from 1300 to 1850" (Whinnom, 1977: 5). Often, excerpts are incomplete in what they reveal lexically and / or grammatically. The father of creolistics and an early scholar of Lingua Franca, Hugo Schuchardt, compared the pidgin to the great 'Seeschlange' (sea snake), a mythical creature, never seen but partially and fleetingly and yet much-discussed and opined on.

> Mit besonderer Leichtfertigkeit und bei jeder Gelegenheit spricht man naemlich seit 50 oder viellicht seit 100 Jahren von der 'lingua franca'; es ist fast die Geschichte von der grossen Seeschlange, die keiner gesehen hat
>
> 'For the last 50 or maybe 100 years people have been talking about 'lingua franca' without much thought and on every conceivable occasion; it is almost like the story of the great sea snake who no one has ever seen.' (Schuchardt, 1883:282, trans. Lutz Marten).

Rachel Selbach, a creolist, updates the analogy, comparing Lingua Franca with the Loch Ness Monster (Selbach, 2008: 2). It is a valid comparison. Both are well-known, regularly discussed, subject to speculation and even conspiracy theories, but with little basis or evidence, for the speakers' opinions. Cremona gave his sustained research project into Lingua Franca, some of which was never published, the working title, "Sherlock". Expert creolists like Schuchardt (1909, trans. 1980), Keith Whinnom (1977) and Robert Hall Jr. (1966) disagree vehemently with, and detract from, one another's theories, while rarely substantiating their strongly-held opinions.

A hallmark of pidgins is their simplicity. Yet Lingua Franca has complex, contradictory and contentious roots. Its very name provokes debate, with theories that the term comes from the Arabic, the Byzantine, or is of Romance origin (Kahane & Kahane, 1976). Commentators from the 16th – 19th centuries offer similar but different formulae for its constituent languages including Italian (or some of its various dialects - Venetian, Tuscan, Genoese, Sardinian and Sicilian -as Italian was not until the late 19th century an established entity), Spanish, French, Provençal, Turkish, Arabic and Portuguese. Given that it was allegedly an exclusively oral language, the source languages attributed to Lingua Franca often depend on the speaker's own linguistic background, as well as where the language was overheard or reported to have been spoken. For example, Cervantes, a slave in Algiers for several years, confirms the linguistic mix as:

> lengua que en toda la Barberia y en aun Constantinople se halla entra cautivos y moros, que ni e morisco ni castellano ni de otra nación alguna, sino una mexcla de todos las lenguas, con la qual todos nos entendemos…
>
> 'the language which throughout Barbary and even in Constantinople is used between Arabs and captives, which is neither Morisco[1] nor Catalan, nor any other nation's language, but rather a mix of all languages and which everyone can understand…'
> (Cervantes, Miguel, 1605. *Don Quixote de la Mancha*, Part 1 Ch. 41; my translation)

Its spread, both geographical and temporal, was remarkable: throughout the eastern Mediterranean, into the Levant and along the Barbary Coast of North Africa (in the ports of Tripoli, Tunis and Algiers as well as possibly into Morocco). The further West one travelled, the more Spanish the pidgin became, according to one of its earliest reporters, Diego del Haedo, a Spanish priest who spent several years in Algiers (Haedo, 1612). Further East, Lingua Franca's more Italian character is reflected in its alternative nomenclature in the Levant, for example Levantine Italian and Levant Venetian (Rousseau, ed. Cranston, 1991; Byron, ed. Murray, 1922; Baglioni, 2013). This is likely a continuum or development of the earlier jargon (based on Venetian), dating back to the 15th century, spoken on the Dalmatian coast, in Greece and on islands like Chios and Crete. Confusingly, it was also known as *franco*.

The predominance of Italian inevitably poses the question of whether Lingua Franca was actually just an L2 version of Italian – foreigner talk or an interlanguage (Selinker 1972), though the fact that many of the commentators referring to it are themselves Italian would seem to bely this. Nevertheless Italian was the principal language of diplomacy, trade and the nautical domain. In the case of the former two, much of the documentation which dates from the 16th- 18th centuries is in Italian, even in French and English chanceries (Cremona, 2002: 26-27). Some documents from these sources seem to confirm that Lingua Franca was used in writing. The "Italian" used appears to offer lexical and grammatical evidence of Lingua Franca rather than of a more standard language.

---

[1] Moriscos were Muslims expelled from Spain in the late 16th century who spoke a language called *aljamiado*, Spanish but with a strong Arabic influence

2. What's in a name?

According to Kahane and Kahane in their article, *Lingua Franca: the story of a term* (1976), the original and eponymous Lingua Franca was a trading language (Kahane & Kahane, 1976: 26), between Europeans and Arabs across the Mediterranean. This is a theory that has been proposed and upheld by many researchers, despite the fact that there appears to be scant if any evidence to substantiate it. There are no trading documents in Lingua Franca, no accounts of trading conversations overheard in Lingua Franca, and nor does the later compilation of the pidgin's lexicon include trade terminology, with barely a mention of monetary terms, or merchandise. Despite this, I believe the pidgin plausibly evolved on the Barbary Coast from the nautical jargon spoken by sailors, merchants and significantly corsairs. This nautical jargon would have been Romance-based, and Italian terms were dominant in the maritime domain (Kahane & Tietze 1954: viii).

The etymology of the term itself, *lingua franca*, is also much debated. Kahane and Kahane (1976) detail the numerous possible derivations for the pidgin's name. The Französisches Etymologisches Wörterbuch (FEW) interprets *franca* as free (FEW, 2002). As such, lingua franca was a free language which could be spoken (and understood – to some degree) anywhere, and escaped the linguistic regulation to which other languages were subject. This would account, in part, for Lingua Franca's apparent lack of rules. The expansion to mean any language of commerce came into being because the term was translated as 'free language' just as *porto franco* was translated as 'free port'. Schuchardt (1909, trans. 1980) cites an article discussing Lingua Franca and its successor, *Sabir*, by MacCarthy and Varnier (1852) who claim that Lingua Franca received its name because of the freedom from taxes it 'enjoyed' in all ports. Other linguists interpret *Franca* as *Franc*, meaning French. Robert Hall (1966) claims that the term derives from the era of the Crusades. The French played a significant role in the religious conflict and their language became dominant in the Levant and into North Africa. Hall maintains that much of the commerce across the Mediterranean in the Middle Ages came from South Eastern France, stretching from the port of Marseilles as far North as Genoa. Provençal, according to Hall (1966) (and much disputed by many others), was a key linguistic constituent in the original Lingua Franca (Hall, 1966: 3).

Another etymological hypothesis is that the name Lingua Franca is rooted in the East, and the Byzantine tradition. The word, *phrangika*, from the Greek, was a term used in the Byzantine region to denote the West, Occidentals and their languages. The early language of communication between the Byzantine Empire and Rome was Latin, which the former termed *Latinum* or *Francum*. The vernacular, which emerged and spread in commerce and diplomacy, from the early thirteenth century, was given the same name (Kahane & Kahane, 1976: 26). Western Europe was, for much of the Middle Ages, in thrall to the Venetocracy which endured until the eighteenth century. *Phrangika* came to mean Venetian as much as Italian, or indeed, any Western language (Kahane &Kahane, 1976: 31). An alternative derivation for Lingua Franca, espoused by Schuchardt (1909, trans. 1980), is from the Arabic, *lisān al-farangˇ*. *Al-farangˇī* initially referred to Latin and then, to describe the trading language employed largely by Jews across the Mediterranean. It later came to encompass the languages of all Europeans, but particularly Italians (Kahane & Kahane, 1976: 40). By the eighteenth century, *Franco* and *Lingua Franca* were both terms used in the West to describe the vernacular spoken

across the Levant and into North West Africa. Italian was at the base of the Levantine dialect while what was spoken in North Africa had Spanish roots. Algiers had long been the crucible of Mediterranean piracy. There the two linguistic varieties merged. An influx of Arabs expelled from Spain, European renegade-corsairs and their fellow European captives (over a quarter of a million European slaves held there in the 17th-19th centuries (Davis 2004)) led to the increase in domains and usage of Lingua Franca.

3. Polyglot Barbary

North Africa had long-established contact with predominantly both Spain and Italy. The Muslim invasion and occupation of Sicily from the 9th to the 11th century and the brief Norman occupation of the African coast in the mid 12th century led inevitably to mutual linguistic exchange between Italian and Arabic, as did ongoing trade across the Mediterranean (Rossi 1928: 143). Prior to the end of the sixteenth century, and the establishment of the Turkish Regencies of Algeria, Tunisia and Tripoli, there were already multiple Romance language communities across North Africa. Spanish *Presidios* – or fortified bases – existed in Morocco, and in Algeria (both in Oran and in Mers-el-Kebir) and were such linguistic enclaves. French was spoken in the *Bastion de France* in Algeria, Genovese (or Ligurian) on the island of Tabarka, Italian and Spanish in the Tunis port of La Goulette or La Goletta, and Spanish, Italian and French alternated in Tripoli depending on the controlling power – Spain, Italy or the Knights of Malta (Cremona, 2001: 290). According to Gosse (2012), the expulsion of thousands of Moors from the Iberian peninsula in 1492 by the Spanish king, Ferdinand II, had a profound effect on the North African countries to which they returned. The countries were "barely able to support a few poor tradesmen, farmers and merchants, were thrown several hundred thousand proud, civilized and warlike people with no available employment, a large measure of ambition and a passionate itch for revenge" (Gosse 2012: 11). Motivated by revenge and a desire for compensation for their lost property in Spain, the Moors declared war on Spain, and, by association, western (Christian) Europe. Gosse highlights the critical assets exploited by these Moors, "they knew the language, they were familiar with the Spanish trading habits, and they had unlimited information in the persons of their compatriots left behind in Spain" (Gosse 2012: 12). Thus, a substantial portion of the Barbary population was already familiar with a Romance language, if not a pidginized version of it. A truce between François I of France and Ottoman Emperor Suleiman in the early 16th Century (Cremona, 2001: 290) facilitated and promoted trade between Europe and North Africa and Levant, with commercial centres set up in Sidone, Aleppo and Tripoli (in Syria) as well as Algiers and Tunis. Barbary also became a centre for European renegades (or *renegadoes* as they were known in Lingua Franca), who quickly acquired power and status given the economic benefit their pirate commerce brought to the Regencies. Plantet offers the stark truth that the state of the finances of Algiers depended exclusively on the proceeds of piracy (Plantet 1889: xx). Haedo's early 16th century account confirms the trickle-down effect of corsair activity on Algiers' economy:

> "All of Algiers is happy then, because some merchants buy many of the slaves and goods that the corsairs bring with them, and other merchants sell clothing and provisions to those who come home from the sea, because many of them buy new clothes. And all is eating and drinking and triumphing" (Garcès 2012: 158).

The vast majority of these renegades came from Italy, particularly Venice and Genoa, Spain and the islands of the Mediterranean. Plausibly, communication among these and fellow renegades from other countries and the residents of Barbary would have been through a pidginized Romance, *Lingua* Franca. The renegades' status meant that inevitably their principal language of communication would extend into many domains of Barbary life.

4. Italian as a lingua franca

Italian was the principal diplomatic language between Europeans ad Turks across the Ottoman Empire for much of the 16th, the whole of the 17th and first half of the 18th century. Migliorini in his history of the Italian language cites the 1582 work of Girolamo Muzio, entitled *Battaglia in difesa dell'italica lingua* (Battle in defence of the Italian language):

> Andate alla Corte del Signor de' Turchi & ritrovate chi sappia latino: ritrovatane appresso il Re di Tunisi, nel regno del Garbo, di Algier, & in altri luoghi; la nostra lingua ritrovarete voi per tutto.
>
> 'Go to the Court of the Sultan of the Turks, and you find there people who speak Latin, you'll find it around the King of Tunis, in the Kingdoms of Gerba and Algiers, and in other places; our language you will find everywhere' (Migliorini, 1960: 380-1; my translation).

Further evidence of the diplomatic use of Italian is provided by two treaties between France and Ottoman territories – Tunis (1666) and Constantinople (1672), both written in Italian (Cremona, 1996, 85-97). Lewis (1999) relates how correspondence was conducted between Queen Elizabeth I and the Ottoman Sultans, Suleiman, Mehmed II and Selim. The Sultan was referred to as the *Gran Signor*, an Italian title. His correspondence was in Turkish, and "a contemporary translation was provided in Italian which the English could understand; the reply was drafted in English, sent in Italian and presumably translated into Turkish". (Lewis 1999: 14)

French Chanceries opened in many North African ports and cities of the Levant. They facilitated trade opportunities for French merchants and seafarers and offer a wealth of documentation including legal contracts, transactions, registrations of debt, claims, captures at sea and wills. The French Chancery in Tunis held documents dating from 1582 relating to French but also many merchants and seamen of other European nationality, and Tunisian merchants whether Turkish, Arab or Jew, testifying to the commercial and corsair character of the city. Cremona (2002) calculates that of the approximately 15,000 documents preserved in the registers, two thirds are written in Italian. As a general rule, documents are only written in French if all those named in the document were French. If any participant were of a different nationality – including

Tunisian, Arab, Turkish, Jewish, Italian, Greek, Spanish, English, Flemish etc, the document is written in Italian. The proportion shifts to predominantly French only at the end of the seventeenth century (Cremona 2002: 27).

Cremona provides yet more evidence of the centrality of Italian, and its reach. Two English business partners in dispute over the wording of a contract registered their protests to one another through the French Chancery in 1628. These were all recorded in Italian. Finally, exasperated and keen to set sail having learned of Algerian corsairs off the coast of Tunis, one of the men, John Barker, the Captain of the Golden Cockerel, implores the French Consul to engage a reliable translator to translate the original English of the agreement into Italian, "accioché ognuno le possa intendere e fare vedere il dretto delle nostre differenze a chi le tiene" (so that everyone can understand and show who is right in this dispute) (Cremona 1997: 59-61; my translation).

The registers from the English consulate in Tunis date from 1675. They are in the vast majority written in Italian, as are those of the Tripoli Consulate. The latter's title is even recorded in Italian as is the 1679 appointment of a new Consul, Thomas Baker.

> Nel nome del omnipotente e Signor Iddio, Libro de Reggsitri della Cancellaria dell'Illlmo Sig Thomaso Bacher, per l'Invit[tissmo Rè della Granda Bertagna [*sic*]; et Difensor della Fede, Carlo 2do, Console nella città et regno in Tripoli di Barbaria, l'anno 1679 adi 17 aprille, giorno del suo possesso in detta carica.
>
> 'In the name of the all-powerful Lord God, the register of the Chancery of the worthy Mr. Thomas Baker, by his Majesty the King of Great Britain and Defender of the Faith, Charles 2nd, the Consul of the city and Kingdom of Tripoli of Barbary, on the 17th of April in the year 1679, the day he came into possession of this role' (NA, 161/20, f°184; my translation).

Writing in 1675, the anonymous author of *Histoire Chronologique du Royaume de Tripoly* (manuscript in 2 volumes (nn. 12219-12220) held in the National Library of Paris), stated that,

> à Tripoly et dans les autres villes de Barberie ils sont en habit court et ils se servent presque toujours de la langue italiaenne pour ce qu'elle est assez connue en Barberie
>
> 'in Tripoli and other Barbary cities they are in court dress and they almost always use Italian as it is well known throughout Barbary'
> (Anonymous, 1675: manuscript 12219, p.175).

Almost a century later (1757), Knecht, the English Consul, who wrote a guidebook for his successors to the complicated maze of diplomatic relations in Tripoli, enumerated among the responsibilities a 'hasnadawr Grande' ed un hasndawr Piccolo' – a greater treasurer and a lesser treasurer (Pennell, 1982: 97; my translation). The combination of Arabic and Italian to describe the role is revealing. Regarding the celebration of Muslim festivals such as Bayram and Ramadan, the Consul was expected to attend with

the Pasha and greet him, kissing his hand and wishing him, "Buona Festa, vostra Eccellenza" (Pennell, 1982: 104). Knecht suggests to later Consuls that the issuing of passports to British subjects "be made out in English and Italian" (Pennell, 1982: 102) and that the designation for a ship's licence in port was "pattenta netta (or pattenta brutta)" – a clean (or not clean) licence (Pennell, 1982: 103; my translation).

Italian was then the *de facto lingua franca* among merchants and seafarers around the Mediterranean, as well as the diplomatic language between its European, Ottoman and Arab coastlines. Braudel suggests that as early as the 16th century Italian was the language of commerce throughout the Mediterranean (Braudel 1987: i., 121). The language of consular documents appears to be predominantly (official) Tuscan Italian but given the plethora of non-native writers of the documents, gallicisms and other Italian dialectal elements are evident. In his comprehensive analysis, of language, especially Italian, used in the Chanceries (predominantly the French Chancery) of Tunis in the 16-19 centuries, Baglioni (2010, continuing the work of Cremona,) concludes that the Italian used by non-native speakers could be examples of Lingua Franca. He does, however, qualify this observation by stating that the variations of Italian found in the documents might be instead individual versions of 'standard' Italian rather than mere alternatives of a separate pidgin (Baglioni 2010:268).

The reach of written Italian intimates a similarly widespread use of spoken Italian. Although those working in the consulates and chanceries would have been educated and literate, it is not improbable that Italian was also spoken by many of the inhabitants of the coastlines of the Mediterranean. This Italian would have necessarily been subject to rather more variation, dialectal differences and would have been more a foreigner talk than Italian *per se*. Could this, in fact, be the Lingua Franca, as described by many a traveller, priest and diplomat? No more than a version of foreigner talk – 'corrupted Italian. I believe it is possible that the pre-pidgin or jargon form of Lingua Franca may well have resembled or even been one and the same as the 'street' or rather sea Italian identified here. This hypothesis must take into account the fact that Italian as a language did not exist until the late 19th century. Conversely, what appears to be the variation of Lingua Franca spoken in the Levant and the Eastern Mediterranean – Greece and modern-day Turkey – is often termed Levant Italian, or Levantine Venetian.

5. Barbary Lingua Franca

Haedo was a Spanish priest who travelled through North Africa and spent time in Algiers, and wrote *Topographia e historia general de Argel*, written at the end of the 15th century, and published in 1612. Haedo's observations and citations of Lingua Franca are some of the earliest of the Barbary States pidgin. Describing the linguistic situation in Algiers, Haedo (1612) refers to the two principal languages, Arabic and Turkish, before discussing at length the third. Haedo does not differentiate between the influences of European languages on Lingua Franca in Algiers, other than perhaps his ordering of the source languages, respectively Italian, Spanish and Portuguese. This may be because, at the time of the late 16th century, Haedo had not been further East in Barbary (to Tunis or Tripoli) so he was unaware of geographical variation.

> La tercera lengua que en Argel se usa, es la que los moros y turcos llaman franca, o hablar franco, llamando ansi a la lengua y modo do hablar christiano… porque mediante este modo de hablar que esta entre ellos en uso, se entienden con los christianos, siendo todo el, una mezcla de varias lengua christianas, y de bocablos , que por la mayor parte son Italianos, y Espanoles, y algunos Portugueses…
>
> 'The third language spoken in Algiers is what the Moors and Turks call *Franca*, or *hablar franco*, calling thus the language and way of speaking of the Christians….because with this language they can communicate with the Christians, since it is a mix of various Christian languages and vocabulary, being mostly Italian and Spanish and Portuguese' (Haedo Ch.XXIX, ff.23 verso – 24 recto; my translation).

Almost contemporary with Haedo was the French redemptionist priest, Pierre Dan, who travelled to Algiers in the 1630s in a bid to free French slaves captured at sea by Barbary pirates. His description of Lingua Franca, like that of Haedo's lacks any sense of variation within the pidgin, and suggests a jargon-like quality to it: *un barragoüin facile et plaisant, composé de Français , d'Italien & d'Espagnol* (a simple and jolly gibberish made up of French, Italian and Spanish (Dan 1637:93; my translation). La Condamine was a French traveller whose account of his 1731 journey to Algiers is recorded in an article by Marcel Emerit, published in the *Revue Africaine* 98 (1954: 354-381). La Condamine states that Lingua Franca, which he refers to as *la mauresque*, is the national language of Algiers. He notes that although the Turks use Turkish amongst themselves, all contact with Europeans is conducted in Lingua Franca. La Condamine attests that it is spoken throughout the Levant and the ports of the Mediterranean. He identifies regional variation, suggesting that what is spoken in Algiers is much more influenced by Spanish than the Lingua Franca found in Tripoli, characterised as a mix of Provencal, Greek vernacular, Latin and "Italien corrompu" (corrupted Italian) (Emerit 1954: 375). Blaquiere, author of *Letters from the Mediterranean*, stated that Ahmed, ambassador for the Bey of Tripoli, Qaramanli in Spain, spoke Italian '*particularly well'* (Blaquiere, 1813:94). A few years later, in 1818, Lyon recorded that *"a bad Italian is generally spoken by the Inhabitants of the town*; *so that Christians have not much difficulty in transacting business*" (Lyon, 1821: 13). Despite the aforementioned regional variation of Lingua Franca within Barbary, it seems inevitable to acknowledge that Italian was a key lexifier of the pidgin. Given the preeminence of Italian in commercial, diplomatic and legal spheres, texts require careful analysis to confirm that the Lingua Franca - a discrete pidgin rather than poor imitations of Italian – existed. Yet the apparent ubiquity of this independent means of communication, mentioned by a plethora of travellers, diplomats, priests, imprisoned slaves – men (and women) at every level of society, serves to confirm its very existence.

6. Levantine Lingua Franca

Lingua Franca is often asserted to have spread across Barbary and the Levant, with very little substantiation in the case of the latter. Nevertheless there are a number of authors who refer to varieties of Lingua Franca found along the Eastern Mediterranean coastlines and into today's Middle East. According to the French philosopher Rousseau (1712-1754, ed. Cranston 1991), a mastery of Italian allowed him to understand the

vernacular spoken by many inhabitants of the Levant. While staying in Neuchatel as a young man, Rousseau happened to meet a Greek monk who claimed to be the Archimandrite of Jerusalem, and who was travelling around Europe to raise money for holy sites. The latter spoke no French but "Rousseau's knowledge of Italian enabled him to understand the *lingua franca*, or bastard Levantine Venetian, which the traveller spoke". (Cranston, 1991: 96). Lord Byron, the English poet, on a grand tour of Europe in the first decade of the nineteenth century also referred to the Levantine pidgin. Writing to Frances Hodgson, the future author of 'The Secret Garden", Byron expounds on his linguistic competence: "my current tongue is Levant Italian, which I gabble perforce. My late dragoman spoke bad Latin, but having dismissed him, I am left to my own resources, which consist in tolerably fluent Lingua Franca" (Byron, ed. Murray,1922: 29).

According to Philip Mansel (2010), the Abbé Prévost, a French Abbot, traveller and author of 'Manon Lescaut', described Lingua Franca in 1775, as 'in use among seamen in the Mediterranean and merchants who go to trade in the Levant and which is understood by people of all nations' (Mansel, 2010: 14). In his "histoire générale des voyages…" Prévost describes the companions of the Capitaine du port a Mocha, Yemen who come to meet the French mission arriving there in 1708 as

> un interprête Banian, qui parlait la langue Portuguaise, et qui était vêtu de blanc, avec une belle ceinture brodée et une écharpe de soie sur son épaule, et d'un Hollandais du Comptoir, vêtu a la Turque, qui parlait la langue Franque
>
> 'an Indian interpreter who spoke Portuguese, dressed in white with a beautiful brocade belt and a silk scarf around his shoulders, and a Dutchman from the trading post, in Turkish dress, who spoke lingua franca' (Prévost, 1752: 292; my translation).

Antoine Galland, cited by Mansel, is best known for his translation from the Arabic of "The Thousand and One Nights". He worked for several French ambassadors to the Ottoman Empire, spending several years in Smyrna and Constantinople at the close of the seventeenth century (Raynard, 2012: 97-102).

Galland describes the variation of Lingua Franca spoken by Jews:

> Ils se servent de ce qu'ils ont apporté de l'Espagne, laquelle approche plus du portuguais que de l'espagnol, que chacun entend assez bien parce qu'ils y mettent des mots italiens, et ainsi ils n'ont point de peine à se faire entendre aux marchands.
>
> 'They use the language they have brought from Spain, which is more like Portuguese than Spanish. Anyone can understand it pretty well because they substitute Italian words, and thus have no trouble being understood by traders' (Galland (ed.) 2010; 150; my translation).

Levant Lingua Franca – or its pseudonyms Levant Italian or Levantine Venetian – clearly shares lexifiers and features with its Barbary relation. Its function as a means of communication between Arabic speakers and Europeans for matters of commerce predominantly. Galland's above assertion (2010) suggests that variation is a key element of the pidgin, which enables rather than hinders understanding, and does not seem to diminish the pidgin's viability.

7. The authoritative source?

The predominant source for Lingua Franca lexicon is the anonymously written *Dictionnaire* (Anonymous,1830), written as a manual for French forces as they sought to colonise Algiers. Although the existence of such a dictionary does much to dispel detractors of the pidgin, it is not as authoritative as it might be. Little is known about the *Dictionnaire*. The lack of a named author has led to speculation as to the identity and nationality of its compiler, or perhaps compilers. Cifoletti (2004) assumes the author(s) to be French but with an impressive knowledge of Italian. The orthography appears to be French: round back vowels /u/ are often rendered as <ou>, as in the cases of *locou* 'crazy', *mouchou* 'very'or *festouk* 'pistachio', and the Italian geminate 'zz' is rendered as 'tz' in *matza* 'kill'. The *Dictionnaire* features both a wordlist and a set of *dialogues*, a collection of phrases and dialogues intended to help the soldiers communicate with the residents of Algiers. It comprises just over 2,100 words, the vast majority of which derive from Italian. Other languages significantly represented include Spanish, French / Provençal, Portuguese and Arabic . The only other lexical inventory within the Lingua Franca corpus is a much earlier document, compiled by the Spanish priest, Serrano in Algiers in 1670. His Lingua Franca wordlist, *Nombres de las cosas en Argel*, 'Names of things in Algiers' is Arabic-, rather than Romance-dominated, and seems more likely to be a variant form of Arabic (Serrano, 1670). He also includes numbers that appear to be Arabic in Roman orthography. Numbers hardly feature elsewhere in the corpus, other than in the *Dictionnaire* where they resemble Italian numbers only in French orthography: *ouno* 'one', *doué* 'two', *tré* 'three', *quatro* 'four' (Anonymous, 1830).

The *Dictionnaire* cannot be considered a wholly reliable, or at least, definitive source of Lingua Franca lexicon. While it clearly sets out to be comprehensive, often providing multiple alternatives for each meaning, drawn from various Romance languages the *Dictionnaire* does not account for the full extent of the pidgin's variation. Such variation would seem inevitable, both diachronic and geographic, given the plethora of source languages, the several centuries of recorded existence in the Barbary States and the extent of its geographical spread, would seem inevitable. In terms of place, the location where Lingua Franca was recorded influences the form used. The constituent populations of each city-state often seem to correspond with the linguistic variation of the Lingua Franca spoken there. However, such factors did not imply a consistent choice of lexicon. Lingua Franca's most consistent feature is its inconsistency.

8. Conclusion

Lingua Franca's status as a pidgin, rather than a jargon or mere 'foreigner talk' seems indisputable, despite – or perhaps because of – its variation over time and space. A pidgin spoken for nearly 300 years, and perhaps written too, across the Mediterranean, into the Levant and predominantly along the coastline of North Africa by multiple linguistic communities, people of diverse social standing, religion, education and

occupation, inevitably manifests lexifying, grammatical and even nomenclature differences. The divergent sources offered for the original term, Lingua Franca, only reinforce this. Its existence on the Barbary Coast would seem inarguable given the sheer number of commentators bearing witness to its centrality in commerce and daily life, and the plethora of linguistic examples (as well as the publishing of the *Dictionnaire* (Anonymous 1830). Indeed, it seems reasonable to conclude that this embodiment of Lingua Franca represents the most established and comprehensive variety of the pidgin.

The key to the establishment of Lingua Franca in the North African regencies and its sustained usage appears to lie in the linguistic makeup of the urban populations. The combination of Arabs with an understanding of Romance languages, European (majority Italian) renegades who quickly acquired wealth and status, and thereby penetrated the Arab power structures, and the significant European slave population made a Romance-based contact language indispensable. Earlier references to *franco* may well describe an earlier stage of the pidgin, or a jargon, a less formed or reduced version of the Barbary Lingua Franca. Whether Lingua Franca was ever more than an oral pidgin remains to be seen. Further analysis of Chancery texts may yet prove that it was written down, and provide more information of linguistic features. Cremona's 'Sherlock' working title remains a guiding principle.

References

Baglioni, D. (2010) *L'Italiano delle cancellerie Tunisine (1590-1703)*. Roma: Accademia Nazionale dei Lincei.

Baglioni D. (2013) Italian Loanwords in Maltese and the Vocabulary of "Levant Italian", in: Borg, A, Caruana, S and Vella, A (eds) P*erspectives on Maltese Linguistics.* Berlin: Walter de Gruyter, pp. 45-58.

Bergna, P. (1925) *Tripoli dal 1510 al 1850*. Tripoli: Nuove arti grafiche.

Blaquiere, E. (1813) *Letters from the Mediterranean*. London.

Braudel, F. (1987) *La Méditerranée et le monde méditerranéen à l'époque de Philippe II*, 8th edition. Paris: Colin.

Cervantes, M. (1605) *El ingenioso hildalgo Don Quixote de la Mancha*. Francisco de Robles.

Cifoletti, G. (2004) *La lingua franca barbaresca.* Roma: Il Calamo.

Cohen, M. (1912) *Le parler arabe des Juifs d'Alger*. Paris: Champion

Cranston, M. (1991) *Jean-Jacques*: *the early life and work of Jean-Jacques Rousseau*, *1712-1754*. Chicago: University of Chicago Press.

Cremona, J. (1997) Acciocch'ognuno le possa intendere: The use of Italian as a Lingua Franca on the Barbary Coast of the seventeenth century. Evidence from the English. *Journal of Anglo-Italian Studies* 5, pp. 52-69.

Cremona, J. (2001) Geografia linguistica e 'lingua franca' del Mediteranneo, in Carlo Napoli e il Mediterraneo, in: Galasso, G and Musi, A (eds) *Atti del convegno internazionale svoltosi dall'11 al 13 gennaio, 2001*. Naples: Società napoletana di storia patria.

Cremona, J. (2002) Italian-based Lingua Francas around the Mediterranean, in: Lepschy, A and Tosi, A (eds) *Multilingualism in Italy: Past and Present*. Oxford: Legenda.

Dan, P. (1637) *Historie de la Barbarie, et de ses corsaires divisée en six liures*. Paris: P. Rocolet.

Davis, R. (2004) *Christian Slaves, Muslim Masters: White Slavery in the Mediterranean, the Barbary Coast and Italy, 1500-1800*. London: Palgrave Macmillan.

Dictionnaire de la Langue Franque ou petit Mauresque, suivi de quelques dialogues familiers, et d'un vocabulaire de mots arabes les plus usuels; a l'usage des français en Afrique. (1830) Marseille: Feissat et Demonchy.

Emerit, M. (1954) *Le voyage de la Condamine à Alger,* 1731. Algiers: Société Historique Algérienne.

Ferrari, G. (1912) *La spedizione della Marina sarda a Tripoli nel 1825*. Rome: Officina Poligrafica.

Galland, A. (2000) *Le voyage à Smyrne*, Paris: Chandeigne.

Haedo, D. (Livres 1612 (republished 2004)) *Topographie et histoire générale d'Alger.* Algiers: Éditions Grand-Alger.

Hall, R. Jr. (1966) *Pidgin and Creole Languages.* Ithaca: Cornell University Press.

Histoire chronologique du Royaume de Tripoly. (1675).

Kahane, H. and Kahane, R. (1976) Lingua Franca: the story of a term. *Romance Philology,* vol. XXX. No. 1.

Kahane, R. and Tietze, A. (1954) *The Lingua Franca in the Levant: Turkish nautical terms of Italian and Greek origin*. Illinois: University of Chicago.

Lewis, B. (1999) From Babel to dragoman: the tortuous history of the interpreter in the Middle East. *Times Literary Supplement*, 23 April, pp. 12-14.

Lyon, G. (1821) *A narrative of travels in Northern Africa in 1818, 19, 20.* London.

MacCarthy and Varnier (1852) La Langue Sabir. *L'Algérien, journal des intérêts de l'Algérie*. 11 May.

Mansel, P. (2010) *Levant: splendor and catastrophe on the Mediterranean*. London: John Murray.

Miglorini, B. (1960) *Storia della lingua Italiana.* Florence: Sansoni.

Minervini, L. (1996) La lingua franca mediterranea. Mistilinguismo, plurilinguismo, pidginizzazione sulle coste del Mediterraneo tra tardo medioevo e prima età moderna. *Medioevo Romanzo*, vol. 20, pp. 231-301.

Murray, J. (ed.) (1922) *Lord Byron's Correspondence*, vol. 1. Cambridge: Cambridge University Press.

National Archive (NA), Kew: *Foreign Office* 161/20 *(Tripoli di Barbaria*).

Pennell, R. (1982) Tripoli in the mid eighteenth century. *Revue d'histoire maghrébine*, vol. 9 (issue 25-6), pp. 91-121.

Prévost, A. (1752) *Histoire générale des voyages, ou Nouvelle collection de toutes les relations de voyages par mer et par terre*, vol. 10. Didot.

Raynard, S. (2012) *The teller's tale: lives of the classic fairy tale writers*. New York: Suny Press.

Rossi, E. (1928) La Lingua Franca in Barberia. *Riviste delle colonie italiane*, numero speciale, Rome, pp. 143-151.

Schuchardt, H. (1883) Review article of: Coelho, los dialectos romanicos ou neolatinos na Africa, Asia e America. In: *Literaturblatt für Germanische und Romanische Philologie, 7*

Schuchardt, H. (1909 (translated 1980)) *Selected essays on pidgin and creole languages* (trans. Glenn Gilbert). Cambridge: Cambridge University Press.

Selbach, R. (2008) Lingua Franca: the linguistic Loch Ness? Logos Series *Studies in Eurolinguistics*, vol.4.

Selinker, L. (1972) Interlanguage. *International Review of Applied Linguistics*, 10, pp. 209-241.

Serrano, B. (1670) *Nombres de las cosas en Argel.*

Tournefort, J. (1718) *Relation d'un voyage du Levant: fait par ordre du Roi*. Anison et Posuel.

Von Wartburg, W. (2002) *Französisches Etymologisches Wörterbuch* (FEW), Tübingen: Niemeyer

Whinnom, K. (1977) The Context and Origins of Lingua Franca, in Meisel, J.M. (ed.) *Langues en contact - Pidgins - Creoles.* Tübingen: Niemeyer.

SOAS Working Papers in Linguistics Vol. 17 (2015): 113-145

Accommodation and resistance in the implementation of a minority language: A survey of headteacher attitudes across primary schools in Cornwall

Sarah Croome
293283@soas.ac.uk

1. Introduction

In 2002 the United Kingdom government specified the Cornish language under Part II of the Council of Europe's Charter for Regional or Minority Languages. A Strategy for the Cornish Language (CCC, 2004) was subsequently developed by local councillors and Cornish language organizations, aimed at significantly reviving Cornish, a language which fell into functional disuse during the 18th and 19th centuries, over a twenty-five year period. Maga, established in 2005, is the organization responsible for implementing the Strategy, of which education is a vital element.

What does Charter recognition mean? Its stated purpose is to protect and promote European regional or minority languages for the contribution they make to Europe's cultural diversity and historical traditions, and because many are in danger of extinction. The UK government is required to base its policies, legislation and practice on the objectives and principles detailed in Part II, Article 7, which include:

> *'f. the provision of appropriate forms and means for the teaching and study of Cornish at all appropriate stages'*

Fishman (2001a: 478) accepts there is no conclusive evidence against which to evaluate the effectiveness of language policy at international or state level, but warns that 'a lack of priorities and linkages seems to characterize the entire legalistic approach' and Romaine (2002: 200) singles out the European Charter as an example of such 'weak linkages', asserting that:

> '*The effectiveness of any initiatives on the supranational level can always be undermined by individual states unless there is some way of guaranteeing the implementation of language-related measures on a supranational level'*

UK implementation of Cornish in the education system typifies this position. Unlike Wales, Scotland or the Isle of Man, Cornwall relies on Westminster legislature, therefore responsibility for teaching lies with the local authority. The strategy for Cornish implementation is overseen by Maga whose *Vision 1* is 'the opportunity for all who wish to learn Cornish, at all levels of education', their stated target being 'An accessible education programme from pre-school to higher and adult education' (CCC, 2004). In reality, Maga are underfunded[1] and understaffed, with only two full-time

[1] Total income/expenditure £212,00 (2010/2011) (Maga Business Plan, 2008-2011)

Project Officers and two part-time Education Officers.[2] Outside this they are 'wholly reliant on volunteer assistance' (J. Lowe,[3] pers. comm., 29th November 2011) which presents problems of availability and suitability. Education is severely affected by this lack of capacity, therefore much work involves 'taster sessions' rather than being 'an integral part of the curriculum', a minimum requirement in the Charter (Article 8.1.*b*.iii).

Although Maga distribute learning materials to schools and offer support with Cornish implementation, decisions regarding use and uptake are entirely at the discretion of schools themselves. In practice this means that neither declarations of policy from Strasbourg, nor local visions, can be effective without commitment from individuals. Due to the mechanisms by which Cornish is currently delivered, headteachers have become the final link in a fragile chain and their attitudes are critical to Cornish implementation in schools. There appears to be no research, however, which examines headteacher attitudes in Cornwall in this specific and relevant context. There is a need to clarify the factors underlying uptake levels, and to explore salient links between headteacher background, attitudes to Cornish and the decision to accommodate or resist the implementation of a language not used as a vernacular for many generations. Previous attitudinal research in Cornwall has been differently focussed in terms of sample and objectives (Camps, 2008; Carkeek, 2009; Dunmore, 2011; Hirner, 1999; MacKinnon, 2000; Willett, 2008; Wimmer, 2010). Comparatively little exists in the sphere of education and pedagogy and this study aimed to elicit responses within particular parameters to gain an understanding of low uptake of Cornish by schools, and to inform future approaches. Better understanding of decision-maker attitudes could facilitate more effective targeting of limited resources.

Objectives:
1. Identify drivers and barriers to Cornish implementation in primary schools
2. Evaluate critically any apparently significant correlations between headteacher background and attitudes towards Cornish
3. Provide information relevant for future approaches

2. Policy, Attitudes and Implementation

This study is not a discussion of whether a presence in the education system is sufficient to revive Cornish or any other language. Experience shows it is not (Fishman, 2001; Hinton, 2001; Hornberger, 2008; Romaine, 2006; Sallabank, 2010), however, as a revived language with no intergenerational transmission[4], Cornish is necessarily reliant on teaching and learning. With this in mind, the following areas were reviewed in order to situate the research:

- The impact of European Charter ratification on Cornish
- Weak linkages between official policy and implementation
- Teacher attitudes, underlying factors and their impact on pedagogic practice
- Previous attitudinal research in Cornwall

[2] The Isle of Man has two Education Officers and several peripatetic and embedded Manx teachers for a population of 83,327. The population of Cornwall is 535,300. (UK Census 2011).
[3] Maga Development Officer
[4] MacKinnon (2005) and Renkó-Michelsén (2013) document several families using Cornish in the home.

Charter impact on the position of Cornish and on government language policy requires initial scrutiny. Prior to Scottish, Welsh and Northern Irish devolution in 1999, the Westminster Parliament had absolute legislative authority in all UK matters including those impacting on minority languages. London-based governmental power remains significant for Cornwall which has no devolved assembly, despite campaigns by a robust nationalist movement[5]. In addition, Cornish is among the more threatened and marginalized of Europe's minority languages[6]. Assessments of the Charter's wider effectiveness and the interpretation and implementation of its requirements by the UK vary, but clear areas of concern pervade the literature. Dunbar (2003) argues that 'the process and fact of ratification have thus far had a relatively limited impact on the languages covered by the Charter'. Assessing UK application of the Charter as 'minimalist' (2003: 41), and its attitude to language activists and NGO's as 'one of indifference', he nevertheless views ratification as a step forward in that Cornish has been officially recognized as a living language meriting 'positive measures of support', a notion endorsed by Gorter (2001: 220) who cites a 'positive *symbolic* effect' in his discussion of the Netherlands. Dunbar also argues that government institutions will be 'forced to explain and account for their policies and practices' at institutional and implementation levels (Dunbar 2003: 57). Evaluation of monitoring reports, recommendations and actual results, however, casts doubt on this.

While Part III of the Charter documents precise obligations, Part II, under which Cornish is specified, contains more general undertakings. Woehrling (2005: 253) characterizes these as a mode of 'evaluating policy' rather than 'verifying actual facts', describing as the Committee's approach as 'compliance management'. He also emphasizes that Committee language is rarely condemnatory, merely indicating 'shortcomings', but concludes that a growing body of reports will facilitate the refinement of good practice in implementation. Nevertheless, Grin (2003: 196) highlights the 'weak rationale' with which implementation itself is approached, noting the emphasis on formal provisions at the expense of actual outcomes (Grin 2003: 197), making it possible for reports to omit any mention of language vitality, focussing on 'supply' rather than results. Romaine (2002: 204) argues that 'without…a variety of other enabling factors, policy statements which merely permit, encourage or recommend the use of a language…cannot be very effective', highlighting the divergence between stated policy and actual practice. She cites Gardner-Chloros (1997; 217) in asserting it is 'pointless to think that "grand declarations of policy…would be effective if they are not tied to a…legal instrument with effective machinery for reinforcement" ', but as she points out (Romaine 2002: 200), only the institutions of states themselves can regulate language policies and the EU avoids interference with such policies. Ultimately, technical compliances and the mere existence of policies or goals cannot ensure actual change which will contribute to effective language revitalization.

This is amply demonstrated by the example of Cornish. While the Charter sets out a clear requirement for Cornish to be provided in primary education *at the absolute minimum* as an 'integral part of the curriculum', Article 8 detailing arrangements for concrete action and Article 7.1.*f* involving an obligation to achieve the desired result (Woehrling 2005: 117), Cornish is still taught as a minimal, extra-curricular subject if at

[5] *Mebyon Kernow* has 4 local council seats but no MP's.
[6] There are approximately 557 fluent speakers (CCC, 2011).

all. Woehrling (2005: 150) interprets the Charter as requiring dedicated language teaching of 'a certain number of hours per week' and is clear that if the measure is not merely to be a pointless formality there should be a minimum level of intensiveness, suggesting at least three hours per week, however, Maga Education Officer, Mike Tresidder, (pers. comm., 7th August 2012) points out that Cornish is not prescribed in the curriculum; 'We end up…working on the margins, on 'out of hours' and voluntary initiatives in schools.' Teaching materials have been developed and distributed to primary schools by Maga and this is noted as an advance in the Committee's third monitoring cycle (Article 93, p.16), as is the inclusion of Cornish in the Languages Ladder[7] and the much-publicized 'Sense of Place' project in which 70 primary schools 'got an introduction to the Cornish language' (COE 3rd monitoring cycle report: Article 92, p.16). In reality, 'Sense of Place' fulfils a national curriculum requirement regarding regional distinctiveness and is not centred on language learning *per se*, while inclusion in the Languages Ladder or its successor is merely potential and arguably dependent on the very attitudes evaluated in this study. Likelihood of inclusion will reflect general frequency of uptake in schools, which remains voluntary and which Maga themselves describe as low. Tresidder (pers. comm., 7th August 2012) further points out that Maga are in the ironic position of having to persuade schools to adopt Cornish as a MFL, placing it in competition with, say, French or Spanish, and making it difficult to persuade schools of its utility. The Committee has frequently indicated that the provision of a minority language is 'not a decision that should be left to the teachers' (Woehrling 2005: 148), yet Cornish implementation relies entirely on the inclination of headteachers and staff.

Despite this scenario, Committee reports become less robust throughout the three Cornish monitoring cycles, implying that as new initiatives are documented, an appearance of compliance is constructed which may not reflect the true situation and there seems to be a lack of an evaluation culture where policies and initiatives are addressed in terms of whether they are expected to be *effective* in terms of specific criteria. This is the locus of Romaine's 'weak linkage' (2002: 200) and what Grin (2003: 84) refers to as 'the chink in the armour' at policy outcome level. Fishman (1991: 178) observed a similar failure of delivery for Frisian; there have been few changes (Gorter in Fishman 2001: 226 ff.) and language policy plans have indeed been 'public posturing and the adoption of well-meaning and good-sounding resolutions; implementation has turned out to be difficult'. In the case of Cornish, the production and dissemination of teaching materials does not in itself mean they will be used, or that this measure will be effective in increasing knowledge and use of Cornish. Success remains dependent on accommodation by existing school staff, particularly in light of Maga's own 'lack of capacity' (J. Lowe, pers. comm., 18th December 2011) which emphasizes the critical role for headteachers at implementation level and the importance of their language attitudes.

There is considerable research in the area of teacher attitudes, much of which highlights classroom situations apparently inapplicable to Cornish in the UK. Nevertheless, there are recurrent themes which are relevant. Wider research (Cantoni, 1997; Lee & Oxelson, 2010) frequently reveals negative teacher attitudes towards minority

[7] A former recognition scheme; part of a UK National Languages Strategy

languages. Studies also suggest there are likely to be correlations between particular variables in the background of headteachers and their attitudes towards minority languages (Byrnes, Kiger & Manning, 1997) and that these factors impact on pedagogic practices. Students spend a significant part of their lives in contact with teachers who can play an important role in shaping their attitudes towards their heritage languages (Corson, 2001; Nieto, 2002; Macias, 2004). Research by Clark (1988), Flores (2001), Nespor (1987) and Pajares (1992) has shown that teacher attitudes have a significant effect on both student attitudes and their own pedagogic practices. Furthermore, Wong-Fillmore (2000) has shown that a sociocultural environment which does not appear to value a minority language is likely to produce students who ultimately reject that language. Research by Franquiz & de la Luz Reyes (1998) refutes a common misconception among teachers that they must be proficient in a minority language in order to support it and the study demonstrated that positive effects were found when teachers showed an interest in the language and treated it as a resource. Similarly, Lee & Oxelson (2010) argue that positive attitudes to a heritage language and a 'willingness to value it publicly in the school space can reinforce students' desire to maintain their heritage language'. In addition, the research demonstrates that educational policies make it difficult for teachers to address student needs not directly related to standardized testing, a pressure cited by many teachers who answered my questionnaire. Strong attitudes were expressed in Lee & Oxelson's study by teachers who did not see a role for themselves in promoting heritage language maintenance, feeling it was not their responsibility. The researchers also identified a persistently negative general view of bilingualism among educators, despite many studies (Cummins, 1981; Hakuta, 1986; Krashen, 1998) which demonstrate its positive effects on the cognitive abilities of children and tendency to 'promote academic achievement' (Lee & Oxelson, 2010: 468). It is clear that teacher attitudes are of critical importance in cultivating and supporting positive attitudes to a language.

3. Methodology and data collection

Maga felt the lack of intergenerational transmission and opportunity to learn Cornish at a young age presented a strong case for targeting resources towards primary schools, but had sometimes experienced a disappointing response to their approaches. Anecdotal evidence suggests:

- Teachers avoid Cornish through lack of confidence
- Teachers struggle with a context for Cornish if they are unaware of Cornish history and culture
- Cornish is relegated to the status of 'heritage' subject and lacks 'currency'
- Teacher training institutions ignore or struggle to make a formal response to Cornish language and culture
- The attitudes of individual headteachers are highly influential

Maga wished to understand how they might improve uptake, whether teachers would prefer a different kind of support and whether they see a future for Cornish at all. They viewed an understanding of teacher attitudes to such issues as critical for planning.

The data collection method was influenced by a wish to compare resultant data, where appropriate, with that from earlier surveys and to obtain sufficient data to enable

generalization based on apparent relationships between variables. These factors, together with the need to obtain and analyze information from a large number of participants across the region, suggested a questionnaire would be the most effective and appropriate data collection method.

The questionnaire (see Appendix) consisted of 20 attitudinal statements requiring respondents to rank order of agreement on a five-point Likert scale, and a combination of 14 open and closed questions intended to elicit demographic, social and pedagogic information and allow additional remarks. Questionnaire development was informed by anecdotal evidence from Maga regarding possible reasons for poor uptake. Correlations evident in surveys of other groups within Cornwall (Carkeek, 2009; Dunmore, 2011; Willett, 2008; Wimmer, 2010) motivated the inclusion of particular questions and Likert items. Questionnaires were posted to 241of 242[8] state primary schools in Cornwall and it was stressed that the questionnaire should be completed by the headteacher or staff member responsible for languages.

3.1. Attitude

To meet the objectives of this study it was relevant to distinguish between three key attitudinal dimensions (Baker, 1992; 13); cognitive, affective and behavioural. The significance of this model is that the three dimensions do not always correspond, for example, a teacher might express a favourable attitude towards the value of preserving Cornish, but hold negative feelings towards its inclusion in the curriculum. It cannot be assumed that favourable attitudes will translate into active support for the language and the distinctions between these levels of language attitude are especially relevant for this research. At a general level teachers may value Cornish, perhaps as an identity marker, but broad principles may be negated or moderated by responses to particular manifestations of language maintenance which impact on individuals, for example, compulsion in education. It is at this level, where thoughts and feelings may or may not translate to behaviour, that barriers to implementation are likely to be located. Questionnaire content was therefore developed to reflect this.

3.2. Limitations

The closed nature of possible responses on a Likert scale may exert a negative impact on validity, particularly in the context of 'affective' variables, because choice of answer is relative to the respondent's own abstract perception of strength of choice. In addition, respondents may avoid using extreme response categories (*central tendency bias*) or attempt to position themselves in a favourable light (*social desirability bias*). It should be stressed that, inevitably, in a survey involving voluntary responses, a bias towards willingness to participate is also present; respondents cannot be described as truly random because they are necessarily self-selecting. Inferences about the population must be made with caution and it cannot be assumed that the respondents are representative of all primary school headteachers across Cornwall as there may have been an exaggeratedly high number of responses from Cornish supporters, although responses were received from 29.874% of the population[9] and describe a variety of views. Nevertheless, it is a valid question whether there is some systematic, relevant

[8] One is a respite facility for children with learning disabilities. When contacted, staff stated that the survey parameters were not appropriate to the school.

[9] Headteachers in the Cornish primary sector.

way in which non-respondents differ from respondents. One hundred non-respondents were followed up by telephone prior to the deadline[10], a process which produced many negative responses[11] towards Cornish and the survey, associated with lack of time, the low priority of Cornish and the dominance of other school activities.

3.3. Quantitative Data Analysis

Data were numerically coded for each variable and entered in SPSS (Pallant, 2007; Greasley, 2008) for statistical analysis. Responses to single Likert items are normally treated as ordinal data as the researcher cannot assume respondents perceive the difference between adjacent levels as equidistant. The data were therefore analyzed using the non-parametric significance test Spearman's rank correlation co-efficient or ρ , to assess the strength of relationships between variables. Of the 241 questionnaires distributed, 72 (29.875%) were returned. In a population of 241, this is shown on statistical reliability calculators to give an accuracy of 95% +/- 10%.[12] Nevertheless, this sample size, which appears robust and capable of providing data from which it is possible to generalize, can produce small returns in subdivisions when variables are crosstabulated, with some subdivisions showing no cases at all. This also made chi-square analysis problematic as it was not possible to maintain sufficient value in cells. Every Cornish primary school, however, was approached at the outset, so it was not possible to increase sampling to achieve a higher response. The significance of ρ is strongly influenced by sample size and moderate correlations may not reach statistical significance at $p<.05$ level.[13] Spearman rho testing in SPSS did not flag a significant strength of relationship between any variable rankings, therefore the dataset is discussed as observed and represented using tables and bar charts.

4. Findings, analysis and discussion

The (abridged) data presented here originally formed the basis of an MA dissertation submitted at SOAS in 2012 and it is not possible to explore relationships between all variables in a paper of this length, therefore only those likely to be the most salient, informed by correlations observed in previous studies, are examined. Demographic and pedagogic questions are referred to as Q, attitudinal statements as AS.

4.1. Language attitudes

Univariate analysis revealed an overwhelmingly positive affective attitude with 90.2% of respondents agreeing that Cornish should be preserved, and 86.1% perceiving the language as a symbol of Cornish identity. In addition, 87.5% disagreed that there is no point keeping Cornish alive and 66.7% disagreed it is irrelevant to modern life. These results contrast sharply with the fact that no respondent spoke Cornish and only 56.9% believed schools should provide Cornish. In addition, opinion was evenly divided on whether there was space in the curriculum for the language and whether money could be better spent elsewhere. Significantly, despite positive levels of affective attitude, 83.4% agreed there were more useful languages to learn than Cornish, therefore it is somewhat surprising that 72% agreed they would like to implement Cornish in their schools. Clearly, a large number of people declare support but have no interest in learning

[10] Pre-numbered returns were recorded on a coded list.
[11] Calls were answered by receptionists, rather than teachers.
[12] Source: www.greatbrook.com/survey_statistical_confidence.htm
[13] p = probability. It is assumed that for 5% of the time, the null hypothesis is rejected when true.

Cornish themselves, despite apparently viewing preservation as important for Cornish identity. Language is an alluring identity marker, but Cornish has not been in vernacular use for 200 years, so there may also be an element of received rhetoric underlying affective attitudes, together with a restricted idea of what language maintenance might involve. Nevertheless, symbolic attachment appears to be tempered by attitudes concerning utility and resources. The possible influence of demographic and social variables on attitudes to Cornish and to implementation will be analyzed in the context of this.

Q1 What is your position in the school?
The results confirmed the responses were relevant to the research objectives.

Table 1. Position in School

Headteacher:	70
MFL Teacher	1
School Manager	1
Total	72

Q2 School Size
School size was coded as 'large' (251+), 'medium' (100-250) and small (<100).

Table 2. School Size

	Frequency	Percent
Large	11	15.3
Medium	29	40.3
Small	32	44.4

School size was crosstabulated with AS18 'I would like to incorporate some Cornish into the curriculum'. As teachers had frequently referred to lack of time and the pressure of other school activities, it seemed that small schools might find it easier to flex the curriculum to accommodate Cornish. In fact, the proportion of large schools who expressed overall agreement (81.80%) was significantly higher than that of medium or small schools. The proportion expressing neutrality was also much lower (9.1%) than that of medium (27.6%) and small (25.0%), which may indicate larger schools are more likely to have a languages teacher willing to consider implementation.

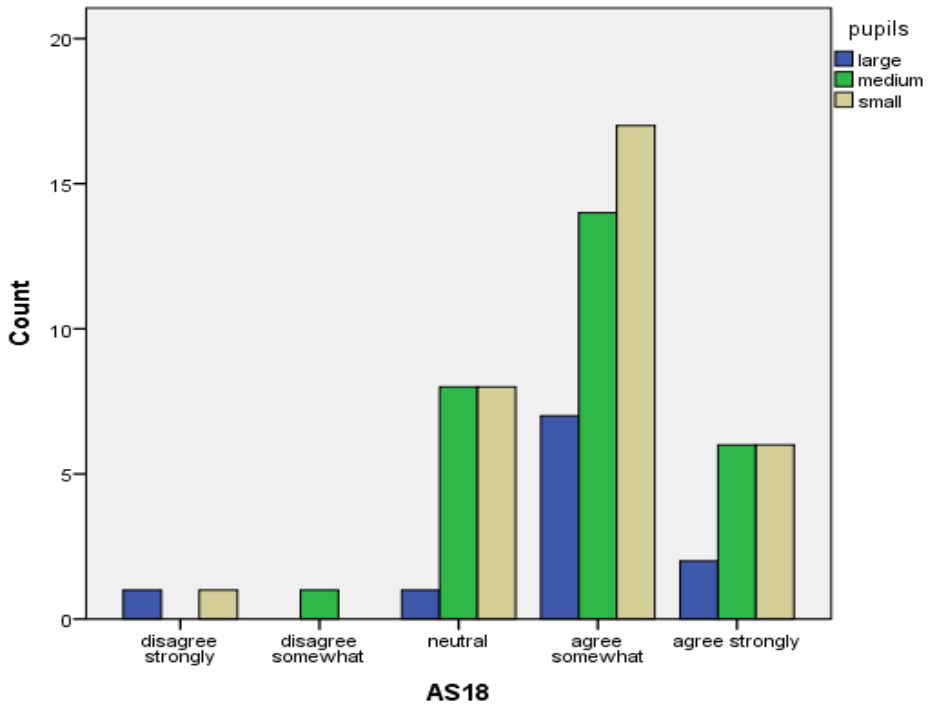


Figure 1. School Size X AS18 'I would like to incorporate some Cornish in to the curriculum'

72.2% of respondents agreed they would like to incorporate Cornish, with 23.6% remaining neutral and only 7% (3) disagreeing.

Table 3. AS18 'I would like to incorporate some Cornish in the curriculum'

	Frequency	Percent
Disagree strongly	2	2.8
Disagree somewhat	1	1.4
Neutral	17	23.6
Agree somewhat	38	52.8
Agree strongly	14	19.4
Total	72	100.0

These figures seem to indicate a more positive general attitude towards implementation than current uptake suggests, although the possibility of positive sample bias or lack of willingness to make an overt statement rejecting Cornish should be considered. In addition, statements indicating intended behaviour do not necessarily translate into

action. Nevertheless, when AS18 was crosstabulated with AS19 'I would welcome some support for this', a greater number wished for support than wished to incorporate the language, indicating some neutral respondents were open to the idea of assistance.

Q3 Gender

Table 4. Responses by Gender

	Frequency	Percent	Percent in total population
Male	18	25.0	40.0
Female	54	75.0	60.0

Frequencies were compared to the ratio of male to female headteachers across all Cornish primary schools.[14] Unsurprisingly in a female-dominated profession, female headteachers outnumber males (Thornton & Bricheno, 2006), but as Table 4 shows, a disproportionate number of respondents were female. This does not necessarily mean male heads are less supportive of Cornish. Studies show women respond at higher rates than men to surveys (Underwood et al., 2000) and that men respond better to internet surveys than to paper surveys (Kehoe & Pitkow 1996; Smith & Leigh 1997) so it may be that men were less receptive than women to the *questionnaire* rather than to Cornish. Cross tabulation of gender with AS18 supports this, with 83% of men expressing a wish to incorporate Cornish, contrasting with 68% of women, and a higher proportion of women (27.8%) were neutral to the idea than men (11.1%). These figures are somewhat unfortunate given the relative dearth of male headteachers.

[14] Source: www.cornwall.gov.uk

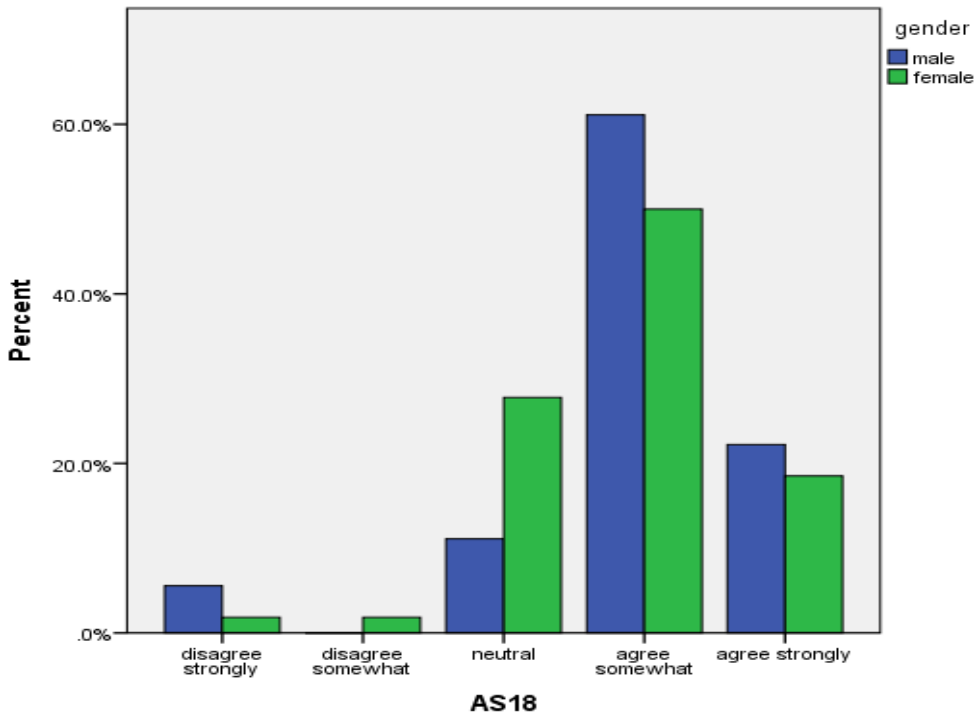


Figure 2. Gender X AS18 'I would like to incorporate some Cornish into the curriculum'

Carkeek (2009: 198), in a survey of the general adult population in Cornwall, found no gender-based difference in attitude to curriculum inclusion, but women were found to be more supportive generally of the revival than men. Extreme views were held more by men, mirroring her observations for boys.

Gender was cross tabulated with 'affective' AS2 'Cornish should be maintained because it is a symbol of Cornish identity', but contrary to observations in young people (Carkeek, 2009: 139), where boys demonstrated more extreme views than girls, there was little gender-based statistical difference in overall attitude, but a higher proportion of women agreed strongly, contrasting with Carkeek's findings in the general adult population.

Table 5. Gender X AS2 'Cornish should be maintained because it is a symbol of Cornish identity'

Percent within gender

Gender	Disagree Strongly	Neutral	Agree Somewhat	Agree strongly
Male	0.0	11.1	61.1	27.8
Female	1.9	13.0	46.3	38.9
Total	1.4	12.5	50.0	36.1

Gender was then cross tabulated with 'cognitive' AS11 'Knowing Cornish is an intellectual advantage'. Similar percentages of men and women showed agreement, but a higher percentage of men (55.6) than women (33.3) were neutral. Disagreement showed the most notable divergence between the sexes; with only 6.7% of men showing overall disagreement, in contrast with 40.8% of women.

Q4 Age

Age was coded in four categories:

Table 6. Respondents by Age

Age	Frequency	Percent
Under 25	1	1.4
26-40	14	19.4
41-55	46	63.9
56+	11	15.3
Total	72	100.00

The highest proportion of respondents (63.9%) appeared in the 41-55 category, as expected for senior professionals. (The under 25, a MFL teacher, is an outlier). When age was crosstabulated with AS18, the 56+ category showed the highest proportions of strong and overall agreement, in line with Carkeek (2009: 202), but in contrast with the results from the youngest adults in that study the 26-40 category fell only marginally short of this. Those between 41-55 showed the least agreement, and high neutrality, although only 3 individuals in the study expressed disagreement. All were in this age group, however, it contains 63.9% of all respondents and Carkeek (2009: 201) observed that the mid-aged category in the general adult population were the *most* positive to curriculum inclusion. This may, however, reflect the fact that they themselves would not, in the main, be carrying it out. Teachers in this age group (neither newly qualified nor approaching retirement) may well have made a more cautious response to the work involved.

Table7. Age X AS18 'I would like to incorporate some Cornish into the curriculum'

Count/Percent within age

Age	Disagree Strongly	Disagree Somewhat	Neutral	Agree Somewhat	Agree Strongly
26-40	0 0.0	0 0.0	3 21.4	9 64.3	2 14.3
41-55	2 4.3	1 2.2	11 23.9	23 50.0	9 19.6
56+	0 0.0	0 0.0	2 18.2	6 54.5	3 27.3

Paradoxically, when age was crosstabulated with 'affective' AS1 'The Cornish language should be preserved', overall agreement was *highest* in the 41-55 category. More remarkably, when age was crosstabulated with AS6 'Cornish is irrelevant to modern life in Cornwall', the highest proportion of overall agreement (21.7%) was shown by this very group. It was also the only group in which *strong* agreement on irrelevance appeared.

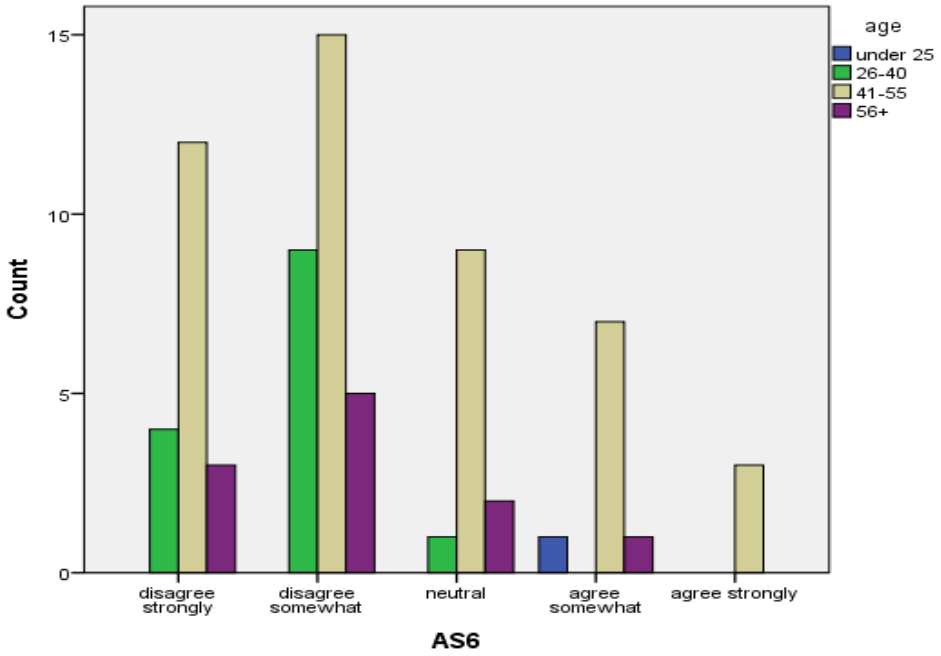


Figure 3. Age X AS6 'Cornish is irrelevant to modern life in Cornwall'
(Under 25 = a single outlier)

Strong agreement with preservation remained highest in the 56+ group, who correspondingly showed the highest proportion of strong agreement with 'affective' AS2 'Cornish should be maintained because it is a symbol of Cornish identity'. In fact, 86.1% of all respondents showed some agreement with this statement, but the lowest proportion of agreement and highest proportion of neutrality were shown by the 26-40 category, suggesting younger heads may place less importance on Cornish as a marker of identity than those aged 56+. It was also considered that younger heads may be more

likely to have moved to Cornwall from elsewhere and feel less affinity with the language as an identity marker, but in fact, crosstabulation of age with 'Cornish' showed that 26-40 held the highest percentage of Cornish self-identifiers (within age) of any category.

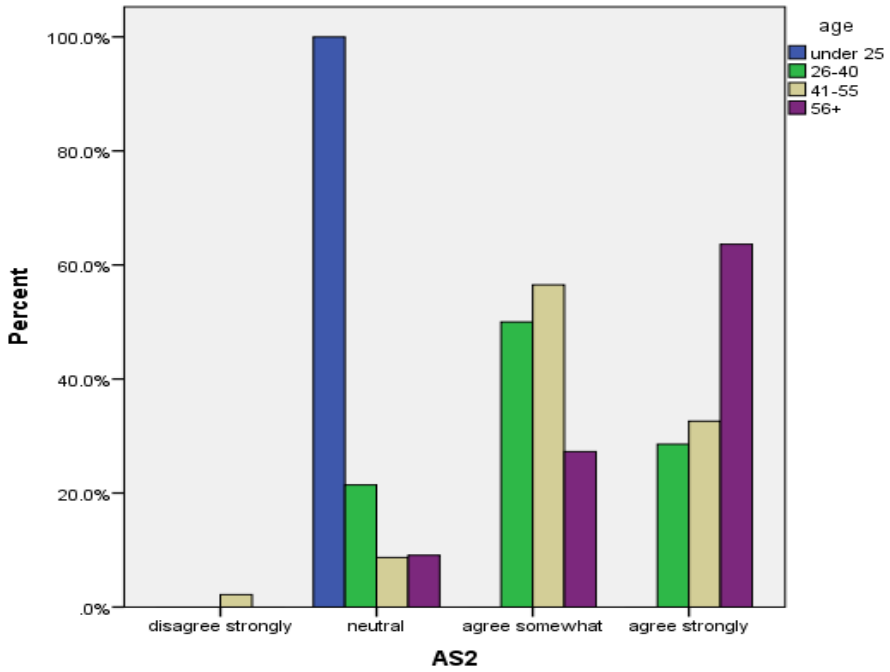


Figure 4. Age X AS2 'Cornish should be maintained because it is a symbol of Cornish identity' (Under 25 = a single outlier)

More respondents between 26-40 felt the language should be preserved (85.7%) than expressed a wish to incorporate it themselves (78.6%), indicating again that favourable attitudes do not automatically translate into active support, however, this category showed the highest proportion of *disagreement* with AS6 regarding the irrelevance of Cornish in modern life (92.9%). This is particularly difficult to interpret as this group was also *least* likely to view it as symbolic of Cornish identity or to agree it is a language worth learning.

Q5 Self -identification as Cornish

Various studies (Carkeek, 2009; Dunmore, 2011; Willett, 2008) have evaluated the criteria deemed essential for 'Cornishness' by residents of Cornwall. Q5 was not intended to reflect these, rather, to record whether or not the respondent, by their own criteria, *felt* themselves to be 'Cornish'. Respondents include people born in Cornwall, lifelong residents and some who have lived there a comparatively short time but who nevertheless self-describe as Cornish. It is self-identification itself which is relevant here, rather than the underlying criteria.

Table 8. Self-identification as Cornish

Cornish	Frequency	Percent
YES	19	26.4
NO	53	73.6
Total	72	100.0

Self-identification as Cornish indicated a correlation with AS18 'I would like to incorporate some Cornish into the curriculum'. No respondent who self-identified as Cornish disagreed with the statement and respondents who did disagree (3) self-identified as non-Cornish. Overall agreement that they would like to incorporate Cornish into the curriculum was expressed by 84.3% of Cornish self-identifiers, contrasting with 68% of those who self-identified as non-Cornish. Neutrality too, was indicated by only 15.8% of Cornish respondents, contrasting with 26.4% of non-Cornish respondents.

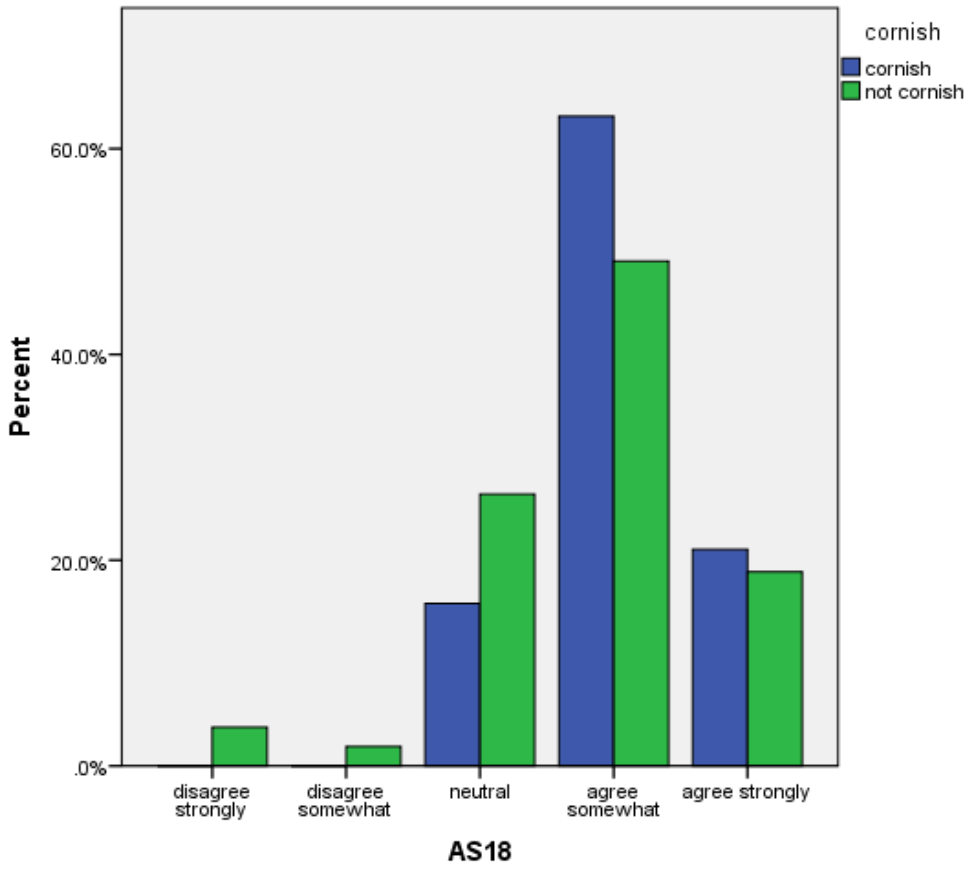


Figure 5. Self-identification as Cornish X AS18 'I would like to incorporate some Cornish into the curriculum'

Crosstabulation of self-identification as Cornish with AS12 'Schools should provide the opportunity for pupils to learn Cornish' also showed divergence, with only 5.3% of Cornish respondents disagreeing, contrasting with 22.7% of non-Cornish respondents.

Non-Cornish respondents also showed greater neutrality. Overall agreement was shown by 73.7% of Cornish respondents in contrast with only 50.9% of non-Cornish respondents.

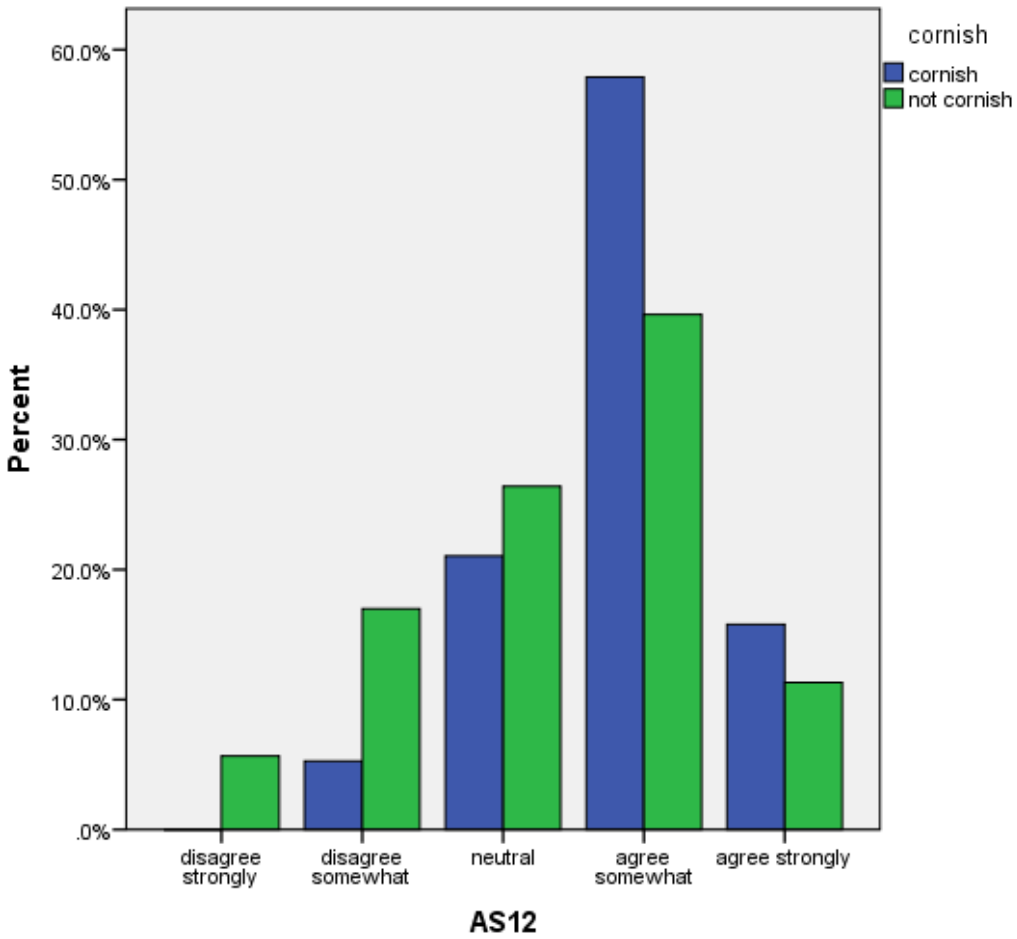


Figure 6. Self-identification as Cornish X AS12 'Schools should provide the opportunity for pupils to learn Cornish'

A higher proportion of Cornish respondents also registered strong agreement but despite this, when 'Cornishness' was crosstabulated with AS13 'Cornish should be compulsory in schools', strong disagreement was shown by similarly high proportions of Cornish and non-Cornish respondents (42.1% and 43.4%) and overall disagreement with compulsion was markedly *higher* for *Cornish* respondents (84.2% compared with 66%). A higher proportion of non-Cornish respondents remained neutral and the 3 individuals showing in the agreement categories were all *non*-Cornish, one registering strong agreement. The variable 'ethnicity' was similarly crosstabulated, having been coded to include a 'Celtic' category which included Cornish, Welsh, Scots and Irish, on the premise that respondents with such a heritage might demonstrate positive attitudes towards Cornish. Frequency of these inclusions were low, and the responses followed the same pattern as the 'Cornish' variable, with the exception of a Welsh respondent who agreed somewhat with compulsion, perhaps motivated by the positive impact of bold measures on Welsh revitalization.

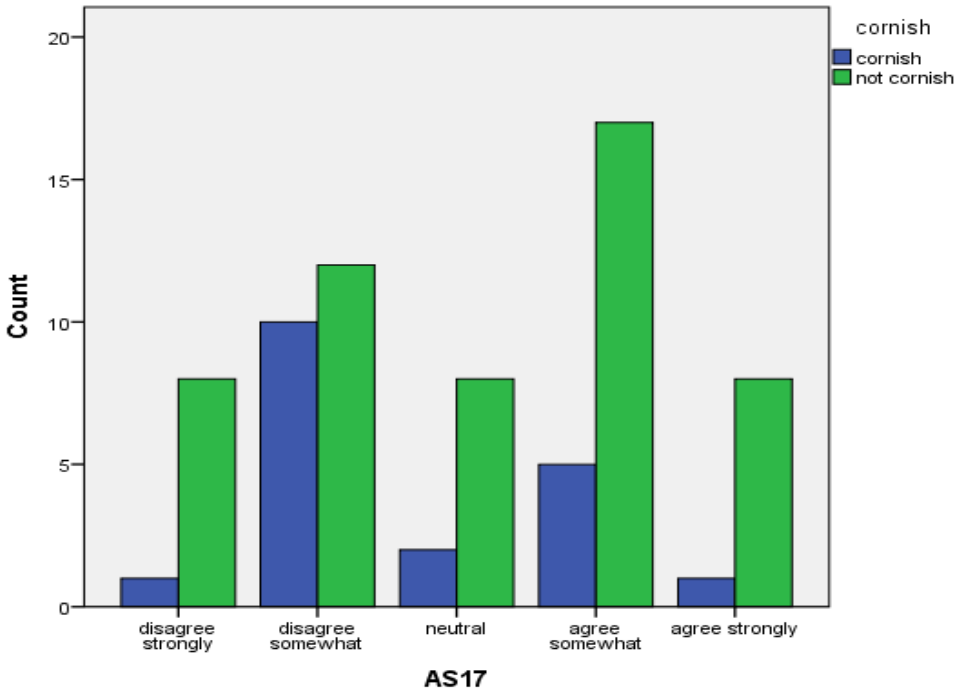


Figure 7. Self-identifying as Cornish X AS17 'There is no capacity in the curriculum for Cornish'.

The variable 'Cornish' was also crosstabulated with 'cognitive' AS17 'There is no capacity in the curriculum for Cornish', again showing marked divergence between Cornish and non-Cornish responses with 57.9% of Cornish respondents disagreeing, contrasting with only 39.6% of non-Cornish respondents. A higher proportion of non-Cornish respondents agreed with the statement, (32.1% compared with 26.3%) and this was especially marked for *strong* agreement where only one Cornish individual appeared. When crosstabulated with AS14 'There are more useful languages to learn than Cornish, overall agreement, and strong agreement in particular, was predictably highest among non-Cornish respondents, however, there was little overall disagreement from the sample, with similar low levels of neutrality and slight disagreement, and a single Cornish respondent showing strong disagreement. Even headteachers who support Cornish took a pragmatic view of its place in the scheme of things.

The notion of 'Cornishness', by whatever measure, has been identified by Willet (2008: 200) as a 'social fact' of 'strong significance' for a considerable majority of Cornwall's population and Dunmore (2011: 74) assesses the Cornish language as 'widely regarded as integral to Cornish identity'. Carkeek (2009: 224) found identification as 'Cornish' to be positively correlated with attitude and the reverse to be true for those identifying as 'English'[15] and certainly, 'Cornishness', by whatever definition, demonstrated a marked relationship with positive language attitudes in this study.

[15] Accompanying response data indicated that no lifelong resident had self-identified as being other than Cornish.

Q6 How long have you lived in Cornwall?

Table 9. Years resident in Cornwall

Years	Frequency	Percent
5 or less	5	6.9
6-10	11	15.3
11-20	21	29.2
20+	17	23.6
Always	15	20.8
N/A	3	4.2
Total	72	100.0

Just over one quarter of respondents self-identified as Cornish, however, the percentage of all respondents resident in Cornwall for a significant length of time was far greater. Only 6.9% (5) had been resident for five years or less, and only 15% (11) for 6-10 years. Almost 30% had been resident for 11-20 years and the remainder, just under 45%, for 20+ years (including lifelong residents). Three individuals did not live in Cornwall, commuting from Devon.

While Carkeek (2009: 206) found, surprisingly, that length of residency exerted little influence on the attitudes of adults in the general population, crosstabulation of Q6 data with 'affective' AS3 'Cornish is a language worth learning' produced a counterintuitive pattern whereby those who had lived in Cornwall all their lives showed the highest levels of strong agreement *and overall disagreement.* Overall agreement was highest among those in the 20+ years category, who also showed no disagreement. Agreement levels in other categories also showed counterintuitive patterns, with agreement falling as the number of years increased, before peaking at 70% for residents of 20+ years, then falling sharply *overall* for lifelong residents despite high *strong agreement.* Neutrality was markedly high for this variable, possibly indicating reluctance to make an overtly negative statement about the value of Cornish, while qualified agreement was lowest in the lifelong category.

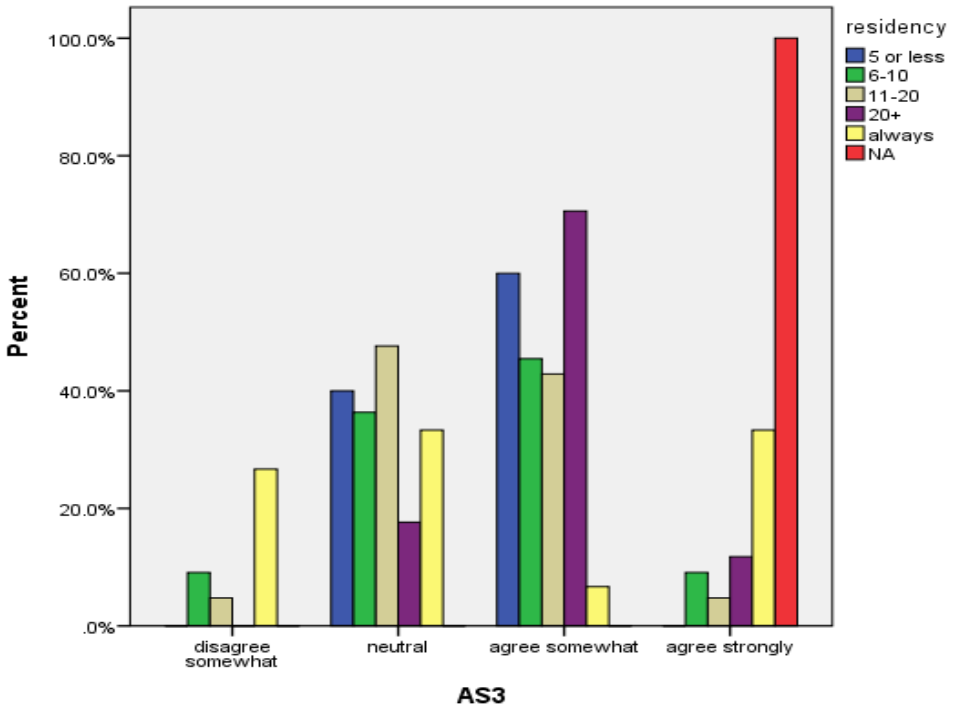


Figure 8. Years resident X AS3 'Cornish is a language worth learning'

Perhaps respondents in the lifelong category felt less inhibited about expressing a negative opinion, while high agreement in the 5 or less category might reflect the enthusiasm of the incomer in embracing all things Cornish. Indeed, Willet (2008: 197) observes 'sense of belonging' and 'love of place' as significant factors which 'bind the Cornish together', regardless of ethnicity.

Nevertheless, when the variable 'years resident' was crosstabulated with AS12 regarding provision in schools, lifelong respondents expressed the lowest overall disagreement (6.7%) and the second highest proportion of overall agreement (53.3%) and strong agreement (13.3%). Again, agreement was highest in the 20+ years category at 76.4%. Oddly, the 5 or less category who attached high value to learning Cornish in AS3 (60.0%) did not show equal enthusiasm for provision in schools, with agreement at only 40.0%. This represents a small number of individuals, but again demonstrates value judgments failing to translate into active expressions of supportive behaviour.

Q6 was then crosstabulated with AS17 'There is no capacity in the curriculum for Cornish', producing a slightly surprising result in that the 20+ years category showed the highest proportion of agreement (having also shown the highest proportion of agreement with the notion of provision in schools!) The highest levels of disagreement with AS17 (both at 60%) were shown by the lifelong residents (unsurprising as they are 'Cornish') and the 5 or less category. The latter is more difficult to explain. This group is composed of 2 respondents in the 26-40 age group and 3 in the 41-55 age group so age does not appear to be a factor. It was considered that headteachers new to a region and to a school may be more receptive to the idea of curriculum adaptation, or make a particular effort to embrace regional activities, however, crosstabulation with AS18 shows this is unlikely. The 5 or less category showed the lowest proportion of

agreement and the highest proportion of disagreement with the idea of implementing Cornish themselves, followed by the 6-10 years category. Yet again, the 20+ category showed the highest agreement, and no disagreement, but were closely followed by the 11-20 years category and the lifelong category. There was a high overall wish to incorporate Cornish, rising with number of years resident, falling slightly for the lifelong category, however, the sample may contain an exaggeratedly high number of Cornish supporters due to self-selection. Nevertheless, the actual number of heads expressing a wish to incorporate Cornish (52) must indicate a positive foundation for learning and teaching, together with the 17 individuals who remained neutral, some of whom said they would welcome support from Maga.

Q7 School Location

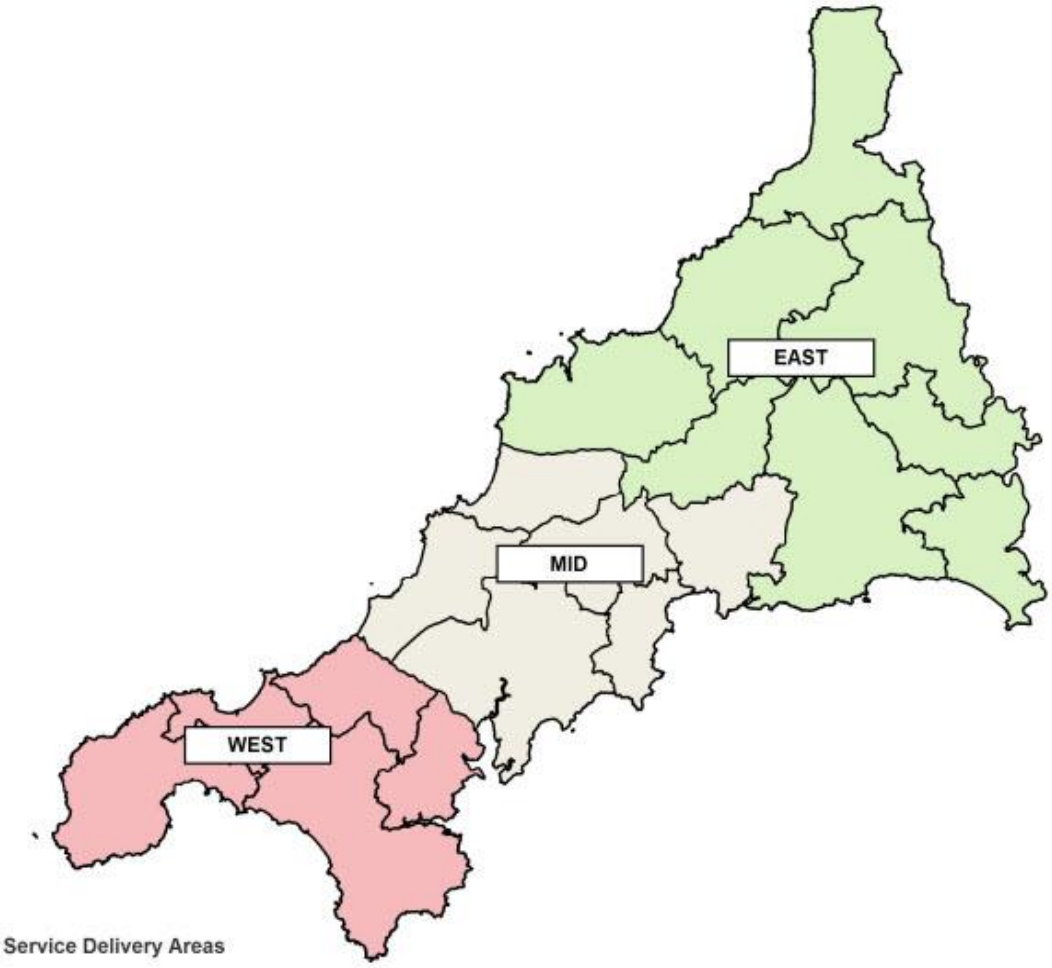


Figure 9. Map of Cornwall showing coding divisions West, Mid and East.

Table 10. Response by District

	Frequency	Percent	Percent in district
East	36	50.0	35.64
Mid	27	37.5	44.26
West	9	12.5	11.39

Historically, Cornish declined from east to west, with West Cornwall retaining vestiges of the language until the 18th century. It has been characterized as possessing a stronger sense of Cornish identity than East Cornwall (Aldous and Williams, 2001; Stoyle, 2002) and was therefore expected to produce high response levels and favourable attitudes towards Cornish. As Table 18 shows, the reverse is true, with only 12.5% (9) of questionnaires being returned by schools in West Cornwall, (11.39% of schools in that district). Willet (2008: 199) also observed counterintuitive results, with Caradon (East) scoring highly in her district rankings for 'Cornishness', and while Carkeek (2009) observed no significant *overall* correlations between location and attitude, respondents in Mid Cornwall appeared significantly more favourable to the idea of curriculum inclusion. Willet (2008: 199) suggests support in the East might reflect resistance to cultural homogenization from neighbouring areas such as Plymouth, and the data from Mid Cornwall may reflect the impact of tourism and second home ownership on fashionable coastal margins, which has forced the indigenous population into the central uplands.

Cross tabulation, however, produced results more in line with conventional thinking, with the West showing the highest proportion of agreement with AS18 (77.8%) and no disagreement. The East showed greater agreement (72.2%) than Mid Cornwall (70.4) (although numbers are small), and while the highest proportion of survey responses did indeed come from Mid Cornwall, correlations with positive attitudes were strongest overall in the West, reflecting traditional notions regarding strength of identity.

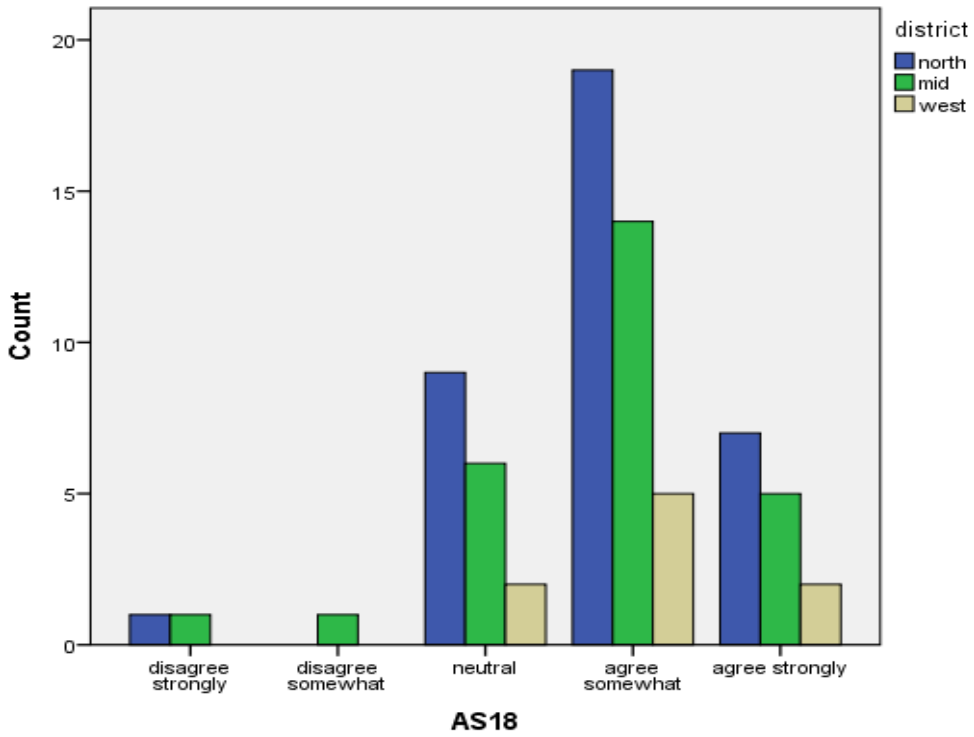


Figure 10. District X AS18 'I would like to incorporate some Cornish into the curriculum'

Location was also cross tabulated with Q10 and Q11 regarding receipt of teaching materials from Maga. Respondents from West Cornwall were clearest about whether they had received materials, showing 0.0% 'not sure' responses, however, they also showed the highest proportion of non-receipts (33.3%). The East showed the highest proportion of receipts (69.4%) and the lowest proportion of non-receipts (11.1%), however, this category also showed disproportionately high 'not sure' responses (19.4%). Overall, 68.1% of heads (49) reported receiving materials, and of these 89.8% said they had used them, although it may be that some respondents felt uncomfortable about saying they had received materials and not used them. Nevertheless, 10.2% of those who had received materials stated openly that they had not used them, citing lack of curriculum time, lack of staff interest, pressure to achieve in English and Maths and the greater utility of other European languages. Several respondents felt the materials were 'no use without additional support', indicating teaching assistance is vital, for initial implementation at least.

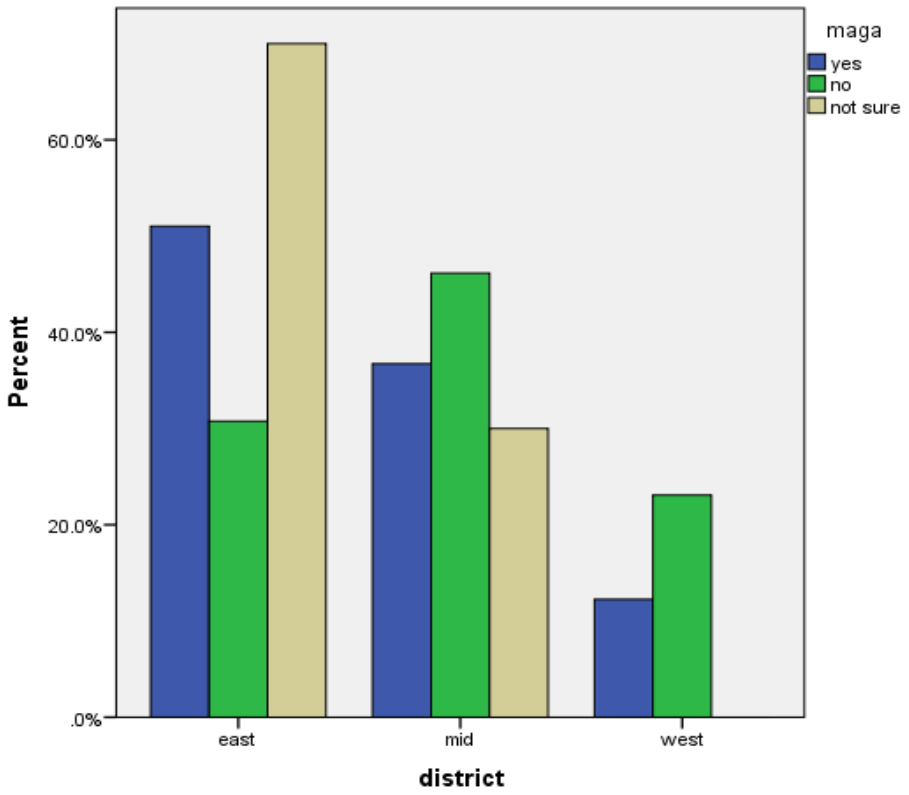


Figure 11. District X Receipt of Maga materials

Q9 What languages do you speak?

Q9 was included in an attempt to ascertain whether the linguistic background of headteachers impacts on their attitudes towards Cornish and its implementation. No respondent was proficient in Cornish although four said they spoke a few words. It was thought teachers proficient in languages other than English might display more positive attitudes towards Cornish and its inclusion in the curriculum than those with little or no language proficiency, due to an appreciation of the benefits of bilingualism or because they might feel less daunted by the idea of implementing an unfamiliar language. The variable was coded as below:

Table 11. Q9 Linguistic Proficiency

	Frequency	Percent
Bilingual/fluent	13	18.1
Basic	43	59.7
Few words	13	18.1
None	3	4.2
Total	72	100.0

In fact, cross tabulation with both AS18 and AS12 ‘Schools should provide the opportunity for pupils to learn Cornish’ produced unexpected results. While it is true

that the number of categories generated here produced small numbers from which to generalize, both cross tabulations showed agreement and disagreement in inverse proportion to language proficiency. While more bilingual/fluent heads agreed than disagreed that schools should provide Cornish, this group also showed the highest proportion of overall disagreement with AS12. Non-linguists were the only group showing no disagreement. Bilinguals/fluent linguists and those with a basic grasp of a language other than English also showed less overall agreement with the idea of school provision than those with only a few words and non-linguists. Cross tabulation with AS18 produced similar inverse results, with the only disagreement and the lowest proportion of agreement recorded by linguists, while non-linguists and those who spoke only a few words showed the highest proportion of agreement and no disagreement.

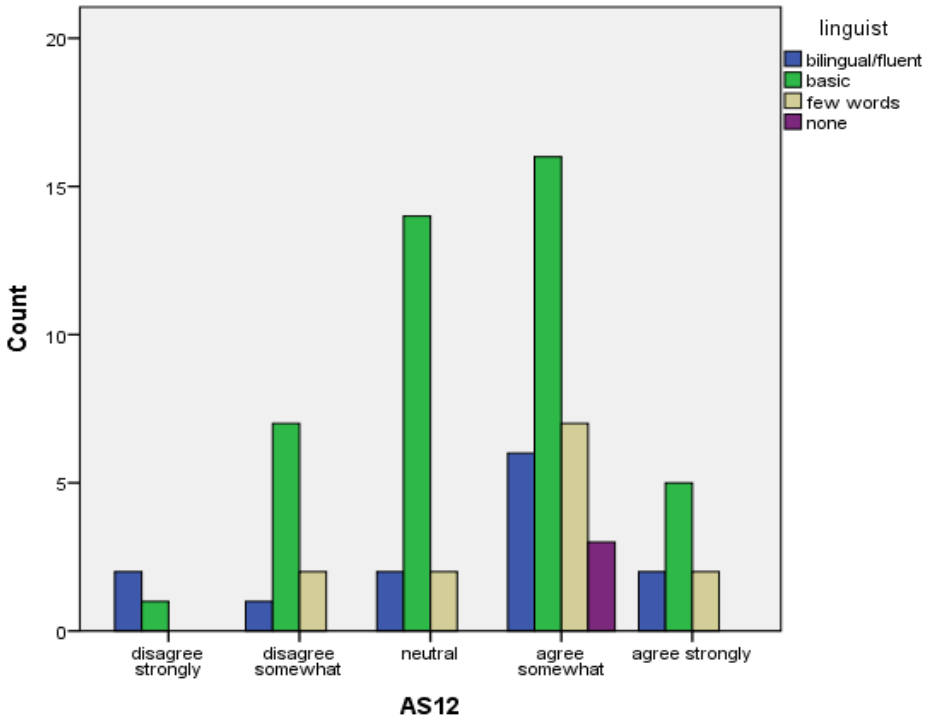


Figure 12. Linguistic proficiency X AS12 'Schools should provide the opportunity for pupils to learn Cornish' (NB 'none' = 3 respondents)

The reasons for this result are difficult to discern and no other Cornish studies have examined the impact of wider linguistic ability on attitudes to Cornish, therefore it is not possible to compare data, but in light of research elsewhere (Lee & Oxelson, 2010) this result seems entirely counterintuitive. Perhaps linguists have a more realistic view of how much effort is required to reach proficiency in a language and feel proficiency should be the ultimate goal, or believe such effort should be put into learning languages other than Cornish which are overtly functional. When linguistic proficiency was crosstabulated with AS14 'There are more useful languages to learn than Cornish' this appeared to be borne out. Despite broad overall agreement, the highest proportion of strong agreement was shown in the bilingual/fluent category and the lowest by non-linguists.

Time and Capacity

Where respondents explained why their school did not provide Cornish (Q12), this related to perceived lack of curriculum space, lack of time caused by pressure to achieve in key subjects linked to league tables, and lack of staff interest/capability.

Table 12. AS17 'There is no room in the curriculum for Cornish'

	Frequency	Percent
Disagree strongly	10	13.9
Disagree somewhat	22	30.6
Neutral	10	13.9
Agree somewhat	22	30.6
Agree strongly	8	11.1
Total	72	100.0

Overall opinion regarding lack of curriculum space was, in fact, more evenly divided than expected, given current levels of inclusion, with 44.5% of the sample (32) disagreeing, 41.7% (30) agreeing and 13.9% (10) remaining neutral, indicating a significant number of headteachers feel the curriculum could be interpreted to accommodate Cornish. Surprisingly for a non-affective statement, AS20 'Learning to deliver Cornish would assist the professional development of staff' prompted polarized responses, showing unusually high levels of strong agreement and strong disagreement in addition to a significant level of neutrality, suggesting respondents interpreted the statement in different ways, or that they were unable to make a judgment on whether this ability would enhance professional development. Some may have focussed on the language element itself rather than the wider benefits which can accrue from undertaking a new skill. AS20 was crosstabulated with Q8 'Where did you train as a teacher?' the variable having been coded as 'South West Britain' (including colleges as far north as Gloucestershire) and 'elsewhere'. Although the high level of neutrality persisted, disagreement, including strong disagreement, was proportionately higher (42.2%) in the 'elsewhere' category than in the SWB category (18.5%). Similarly, agreement was higher in the SWB category (29.6%) than in the 'elsewhere' category (10.56%) and strong agreement was significantly higher (22.2% contrasting with 8.9%).

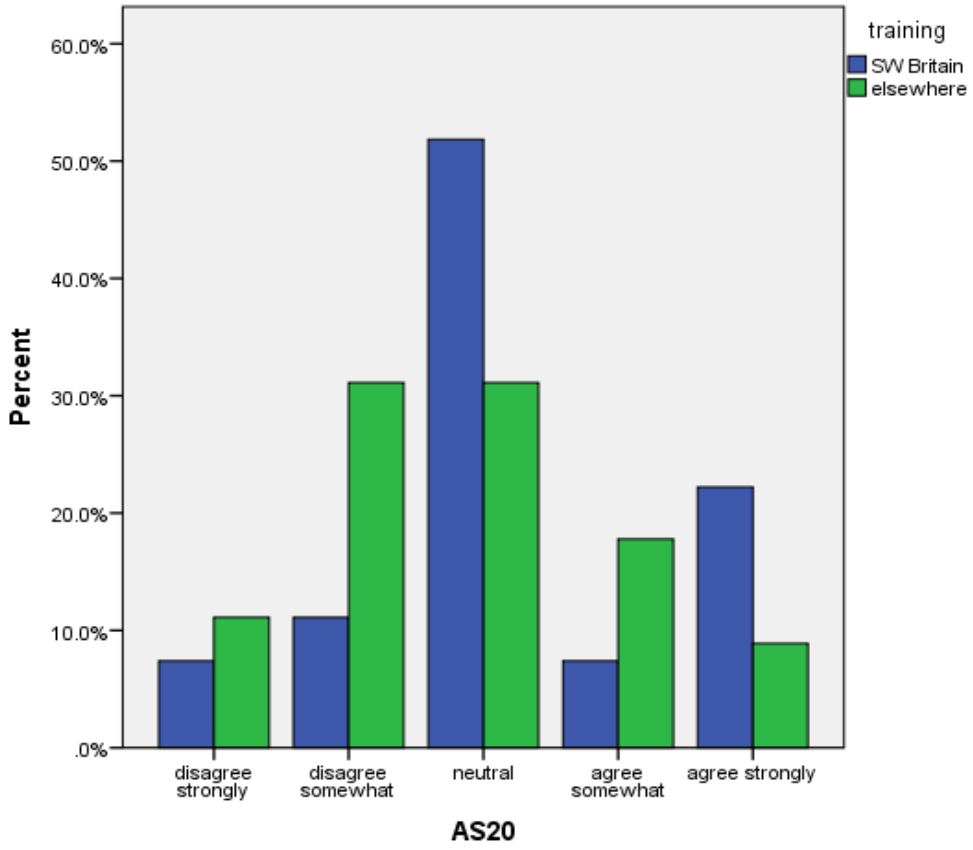


Figure 13. Training location X AS20 'Learning to deliver Cornish would assist the professional development of staff'

School centred training is now available in Cornwall, but was not available to current headteachers, therefore no respondent had trained in Cornwall itself. It was considered that the SWB category might contain a high proportion of teachers originating from Cornwall who might demonstrate positive attitudes to Cornish. This seemed more likely than training in SWB having had an effect on teacher attitudes, however, on examination, the SWB category was *not* overrepresented by Cornish self-identifiers. Overall, 62.5% (45) of respondents had trained elsewhere, only 37.5% (27) having trained in SWB. Of those who had trained in SWB, 44.4% (12) were Cornish, contrasting with 55.6% (15) self-identifying as non-Cornish, so it is possible there is a relationship between training in the South West and positive pedagogic attitudes to relating to Cornish. Nevertheless, overall frequencies show that 52.7% of headteachers who responded were not Cornish and did not train in SWB, which the statistical relationships observed in this study suggest has significance for Cornish implementation.

5. Conclusion

Analysis of headteacher attitudes to Cornish and its implementation in schools sought to identify variables which influence levels of support and to differentiate between the impact of affective, cognitive and behavioural attitudes. While affective attitudes to Cornish appear hugely positive, the majority supporting preservation and characterizing Cornish as a symbol of Cornish identity, cognitive and behavioural attitudes are

associated with more qualified support. Affective attitudes, modified by practical considerations and beliefs, fail to translate into behavioural practices - a significant barrier to successful implementation. Results appear to show that particular variables are influential, not only on affective attitudes, but on cognitive and behavioural attitudes. Although positive affective attitudes were high in the West, survey responses were markedly low from this area, similarly, a high proportion of women completed the survey, but men demonstrated greater enthusiasm for implementation. Linguistic background produced the most counterintuitive results as it had been surmised that linguists would be receptive to Cornish, but they appeared least supportive, citing the greater utility of modern languages. Younger heads appeared to place less importance on Cornish as an identity marker and were more likely to feel money could be better spent elsewhere, which may represent a generational shift detrimental to the language, as it was also widely agreed that the language is a powerful symbol of identity but has little instrumental use. One headteacher commented that Cornish learning could acquire more currency if provision were linked to wider schemes which endorse achievement, but unless pupils are encouraged to choose Cornish, potential links cannot be effective.

The most significant explanatory variable of positive attitudes, on all levels, was self-perceived ethnic identity. Self-identification as Cornish appeared to be the most influential overall factor in all aspects of support for Cornish, including curriculum inclusion and belief in adequate curriculum capacity, a phenomenon which has important implications for official policy and is worthy of further investigation. If teachers who self-define as Cornish have more favourable attitudes to the language, they are more likely to implement it in their schools. Carkeek (2009: 167) demonstrated that those who have taken lessons are more likely to self-define as Cornish and that young self-definers are less likely to move away from Cornwall. Language learning clearly impacts on self-identification; if Cornish is *not* provided then Cornish self-definers will presumably be less numerous, which has implications for regional distinctiveness and the local economy, again an area for future research. Teachers who trained in South West Britain and those who had lived in the region for a significant length of time also demonstrated more supportive attitudes towards Cornish. Now that school-centred training is available in Cornwall itself, an opportunity exists to build early relationships with those apparently most receptive to the notion of implementation. The widespread dissemination of teaching materials without initial support was criticized by several teachers and it may be that the cultivation of relationships with fewer, genuinely engaged schools who would capitalize on intensive help would be more effective in increasing learner numbers. Indeed, ten schools receiving strong Cornish provision would have greater impact than fifty schools receiving materials which remain unused. Schools likely to implement Cornish could be targeted over the long term, not merely for taster sessions, and a headteacher self-defining as Cornish appears to be the best predictor of likely support.

References

Aldous, P. & Williams, M. (2001) A Question of Ethnic Identity. In: P. Payton (ed.), *Cornish Studies* 9, Redruth: Institute of Cornish Studies.

Alim, S. H. (2005) Critical language awareness in the United States: Revisiting issues and revising pedagogies in a resegregated society. *Educational Researcher, 34*, pp. 24-31.

Baker, C. (1992) *Attitudes and Language.* Clevedon: Multilingual Matters.

Baugh, J. (2007) Plantation English in America: Nonstandard varieties and the quest for educational equity. *Research in the teaching of English,* 41 (4), pp. 465-472

Byrnes, D. A., Kiger, G. & Manning, M. L. (1997) Teachers' attitudes about language diversity. *Teacher and Teacher Education,* 13, 637-644.

Camps, C. (2008) *Attitudes to language and culture in the teaching profession in the South West of Britain.* Truro: DESKY Research.

Cantoni, G. (1997) Keeping Minority Languages Alive: The School's Responsibility. In: J. Reyner (ed.), *Teaching Indigenous Languages,* Flagstaff: Northern Arizona University, pp. 1-9.

Carkeek, A.(2009) *Cornish Language Revival: Attitudes, Behaviour and the Maintenance of an Ethnic Identity.* Unpublished thesis (PhD), University of East Anglia.

Clark, C. M. (1988) Asking the right questions about teacher preparation: Contributions of research on teaching thinking. *Educational Researcher,* 17 (2), pp. 5-12.

Corson, D. (2001) *Language diversity and education.* New Jersey: Erlbaum.

Council of Europe, (2011) *Evaluation Reports of the Committee of Experts: Documents of the Monitoring Process,* 1st, 2nd and 3rd monitoring cycles, 2002-2010. Strasbourg: Council of Europe Publishing.

Council of Europe, (1993) *The European Charter for Regional or Minority Languages and Explanatory Report*. Strasbourg: Council of Europe Publishing.

Cummins, J. (1981) The role of primary language development in promoting educational success for minority language students. In: California State Department of Education (ed.) *Schooling and language minority students: A theoretical framework,* pp. 3-49. Los Angeles: Evaluation, Dissemination and Assessment Center, California State University.

Dunbar, R. (2003) *The ratification by the United Kingdom of the European Charter for regional or minority languages.* Mercator Working Papers. Barcelona: Ciemen.

Dunmore, S. (2011) Xians-via-Yish? Language attitudes and cultural identities on Britain's Celtic periphery. In: P. Payton (ed.), *Cornish Studies,* 19 (3), pp. 60-83.

Fishman, J. A. (1991) *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages.* Clevedon: Multilingual Matters.

Fishman, J. A. (2001) *Can Threatened Languages be Saved? Reversing Language Shift, Revisited - A 21st Century Perspective.* Multilingual Matters: Clevedon: Multilingual Matters.

Flores, B. (2001) Bilingual education teachers' beliefs and their relation to self-reported practices. *Bilingual Research Journal,* 25, pp. 275-299.

Franquiz, M. E. & de la Luz Reyes, M. (1998) Creating inclusive learning communities through English language arts: From chanclas to canicas. *Language Arts,* 75, pp. 211-220.

Gardner-Chloros, P. (1997) Vernacular literacy in new minority settings in Europe. In: Tabouret-Keller et al. (eds.), *Oxford Studies in Anthropological Linguistics,* 13, pp. 189-221.

Garmon, M. A. (2004) Changing preservice teachers' attitudes/beliefs about diversity: what are the critical factors? *Journal of Teacher Education,* 55 (3), pp. 201-213.

Görter, D. (2001) A Frisian Update of Reversing Language Shift. In: J. Fishman (ed.), *Can Threatened Languages Be Saved? Reversing Language Shift, Revisited: A 21st Century Perspective.* Clevedon: Multilingual Matters, pp. 215-233

Greasley, P. (2008) *Quantitative Data Analysis using SPSS. An Introduction for Health and Social Science.* Maidenhead: Open University Press.

Grin, F. (2003) *Language Policy Evaluation and the European Charter for Regional or Minority Languages.* Basingstoke: Palgrave MacMillan

Haig, Y. & Oliver, R. (2003a) Is it a case of mind over matter? Influences on teachers' judgments of student speech. *Australian Review of Applied Linguistics,* 26, pp. 55-70.

Haig, Y. & Oliver, R. (2003b) Language variation and education: Teacher's perceptions. *Language Education,* 17 (4), pp. 266-280.

Hakuta, K. (1986) *Mirror of language: The debate on bilingualism.* New York: Basic Books.

Harrison, N. (2004) Self-recognition and well-being: Speaking Aboriginal English in healthy classrooms. *The Australian Journal of Indigenous Education* 33, pp. 7-13.

Hinton, L. (2001) Sleeping Languages: Can They Be Awakened? In: L. Hinton & K. Hale (eds.), *The Green Book of Language Revival in Practice.* San Diego: Academic Press.

Hirner, U. (1997) *Language Death and Revival: A Sociolinguistic Investigation of Cornish and Welsh.* Unpublished thesis (MA), University of Graz.

Hornberger, N. (1996) Language and Education. In: S.L. McKay & N. Hornberger (eds.), *Sociolinguistics and Language Teaching,* pp. 449-73, New York: Cambridge University Press.

Hornberger, N. (2008) Introduction. In: N. Hornberger (ed.), *Can Schools Save Indigenous Languages? Policy and Practice on Four Continents,* Basingstoke: Palgrave MacMillan, pp. 1-11.

Jones, R. G. (2003) The "Bad" American Dialects. In: *Second Language Learning and Teaching,* 3.

Kehoe, C. M. & Pitkow, J.(1996) Emerging Trends in the WWW User Population. *Communications of the ACM,* 39 (6).

Krashen, S. D. (1998) *Condemned without a trial: Bogus arguments against bilingual education.* Portsmouth: Heinemann.

Lee, J. S. & Oxelson, E. (2010) "It's Not My Job": K-12 teacher Attitudes Toward Students' Heritage Language Maintenance. *Bilingual Research Journal of the National Association for Bilingual Education,* 30:2, pp. 453-477.

Macias, R. (2004) The role of schools in language maintenance and shift. *Heritage Language Journal,* 1. Available online: http://www.international.ucla.edu/lrc/hlj/article.asp?

MacKinnon, K. (2000) *The MacKinnon Report*: *An Independent Academic Study on Cornish*. GOSW available online: www.gosw.uk/publications/cornish.

MacKinnon, K. (2010) Development of Cornish Language and Literature. In: D. O'Neill (ed), *Rebuilding the Celtic Languages,* Talybont: Y Lolfa Cyf.

MAGA: The Cornish Language Partnership (2008) *Maga Business Plan 2009-2011 (no attribution).* Truro: Cornwall County Council.

Mantero, M. & McVicker, P. (2006) The Impact of Experience and Coursework: Perceptions of Second Language Learners in the Mainstream Classroom. *Radical Pedagogy,* ISSN: 1524-6345.

McKinerney, V. (2003) Multiculturalism in Today's schools: Have Teachers' Attitudes Changed Over Two Decades? *Annual meeting of the Australian Association for Research in Education, Auckland, New Zealand, November 2003.* Available

online: http://www.aare.edu.au/03pap/mci03767.pdf
Nespor, J. (1987) The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, 19 (4), pp. 317-328.
Nieto, S. (2002) *Language, Culture and Teaching.* New Jersey: Erlbaum.
Oliver, R., Rochecouste, J., Vanderford, S. & Grote, E. (2011) Teacher Awareness and Understandings About Aboriginal English in Western Australia. *Australian Review of Applied Linguistics*, 34 (1), pp. 60-74.
Pajares, M. F. (1992) Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62 (3), pp.307-332.
Pallant, J. (2007) *SPSS Survival Manual. A Step by Step Guide to Data Analysis using SPSS for Windows.* 3rd ed. Maidenhead: Open University Press.
Renkó-Michelsén, Z. (2013) Cornish in the Family: facilitating intergenerational transmission. University of Helsinki, Finland.
Romaine, S. (2002) The Impact of Language Policy on Endangered Languages. *International Journal on Multicultural Societies (IJMS)*, 4, (2) pp. 194-212.
Romaine, S. (2006) Planning for the survival of linguistic diversity. *Language Policy*, 5, pp. 441-73.
Sallabank, J. (2010) What are the aims of language revitalization? A critical review of the situation in Guernsey (Channel Islands). In: H.G. Lewis & N. Ostler (eds.) *Reversing Language Shift. How to Re-awaken a Language Tradition. Proceedings of the Fourteenth FEL Conference, Carmarthen, Wales.* Bath: The Foundation for Endangered Languages.
Sayers, D. (2012) Standardising Cornish: The politics of a new minority language. *Language Problems and Planning* 36 (2) pp. 99-119.
SGRUD Research (See MacKinnon, K.): (2000) Transcription of an audio recording of a Cornish Language Focus Group meeting at the Church Rooms, Lostwithiel, led by Professor Kenneth MacKinnon. www.sgrud.or.uk/focus_groups/cornish_focus
Smith, M. A. & Leigh, B. (1997) Using the internet as an alternative source of subjects and research environment. *Behaviour Research Methods, Instruments and Computers*, 29, pp. 496-505.
Stoyle, M. (2002) *West Britons: Cornish identities and the early modern British state.* Exeter: Exeter University Press.
Thornton, M. & Bricheno, P. (2000) Primary School Teachers [sic] Careers in England and Wales: the relationship between gender, role, position and promotion aspirations. *Pedagogy, Culture and Society*, 8 (2), pp. 187-206.
Underwood, D., Maier, M. & Kim, H. (2000) To mail or to web: Comparison of survey response rates and respondent characteristics. *Paper presented at the 40th Annual Forum of the Association for Institutional Research, Cincinnati, OH, May 21-24, 2000.*
Walker, A., Shafer, J. & Iiams, M. (2004) "Not In My Classroom": Teacher Attitudes Towards English Language Learners in the Mainstream Classroom. *NABE Journal of Research and Practice*, 2 (1) pp. 130-160.
Willett, J. (2008) Cornish Identity. Vague Notion or Social Fact? In: P. Payton (ed.), *Cornish Studies* 16, Exeter: University of Exeter Press, pp. 183-205.
Wimmer, U. (2010) *Reversing Language Shift: The Case of Cornish.* Published version of a PhD thesis, University of Graz (2006). Truro: Cornish Language Board.

Woehrling, J. (2005) *The European Charter for Regional or Minority Languages. A Critical Commentary*, (translation by Patricia Wheeler). Strasbourg: Council of Europe.

Wong-Fillmore, L. (2000) Loss of family languages: Should educators be concerned? *Theory into Practice*, 39 (4), pp. 203-210.

Glossary, Abbreviations and Terms

AS	Attitudinal statement in survey (see Appendix)
CCC	Cornwall County Council
Charter:	The European Charter for Regional and Minority Languages
Committee:	The Committee of Experts appointed to monitor compliance with the Charter
COE	Council of Europe: an international organization which promotes co-operation human rights, democracy and cultural matters between European states.
Q	Background question in survey (see Appendix)
Maga:	The Cornish Language Partnership, established to implement the Strategy for the Cornish Language
MFL	Modern Foreign Language
SPSS	Statistical Package for the Social Sciences
SWB	South West Britain

Appendix: Questionnaire

1. What is your position in the school?…….
   Email address if willing…….....
2. Approx. number of pupils on roll…….

3. Are you
   Male □
   Female □

4. Please indicate which age bracket best describes you:
   Up to 25 □
   26 - 40 □
   41 - 55 □
   56 + □

5. a) Do you consider yourself to be Cornish? Yes □ No □

   b) If not, how might you describe your nationality/ethnicity?

   …….....

6. How long have you lived in Cornwall?……

7. In which district is your school?……..

8. Where did you train as a teacher?…….

9. Aside from English, what other languages, if any, do you speak?
   Fluently…….....….....
   Basic understanding…….....
   A few words or phrases…….....
   None……

10. Has your school received any Cornish language information or teaching materials in the past? Yes □ No □ Not sure □

11. If so, have you felt able to make use of them? Yes □ No □

12. If not, can you briefly describe why not?.....

13. Does your school have any policy regarding teaching Cornish? Yes □ No □

14. Does your school have a staff member with responsibility for provision of Cornish?
    Yes □ No □

PLEASE SHADE ONE CIRCLE PER STATEMENT

	disagree strongly	**disagree somewhat**	**neutral**	**agree somewhat**	**agree strongly**
1. The Cornish language should be preserved	O	O	O	O	O
2. Cornish should be maintained because it is a symbol of Cornish identity	O	O	O	O	O
3. Cornish is a language worth learning	O	O	O	O	O
4. I like hearing Cornish spoken/seeing Cornish used	O	O	O	O	O
5. There is no point keeping the Cornish language alive	O	O	O	O	O
6. Cornish is irrelevant to modern life in Cornwall	O	O	O	O	O
7. Learning Cornish is unnecessary because Britain is predominantly English speaking	O	O	O	O	O
8. Both English and Cornish should be important in Cornwall	O	O	O	O	O
9. Cornish should be limited to use with friends and Family	O	O	O	O	O
10. Money spent on the promotion of Cornish could be better spent elsewhere	O	O	O	O	O
11. Knowing Cornish is an intellectual advantage	O	O	O	O	O
12. Schools should provide the opportunity for pupils to learn Cornish	O	O	O	O	O
13. Cornish should be compulsory in Cornish schools	O	O	O	O	O
14. There are more useful languages to learn than Cornish	O	O	O	O	O
15. Children would get confused learning Cornish	O	O	O	O	O
16. People only need to know one language	O	O	O	O	O
17. There is no capacity in the curriculum for Cornish	O	O	O	O	O
18. I would like to incorporate some Cornish into the curriculum	O	O	O	O	O
19. I would welcome provision and support for this	O	O	O	O	O
20. Learning to deliver Cornish would assist the professional development of staff	O	O	O	O	O

SOAS Working Papers in Linguistics Vol. 17 (2015): 147-167

Writing System Mimicry in the Linguistic Landscape

Paul Sutherland
619191@soas.ac.uk

1. Introduction

The study of the linguistic landscape of cities, introduced by Landry and Bourhis (1997), has so far attended to a variety of different uses of written language in the public sphere. One use it has yet to address in significant detail is what I term *writing system mimicry*, and it is the principal aim of this paper to begin to address this gap.

Writing system mimicry will be defined in more detail in Section 3, but may be loosely described as the choice to make a text in one writing system superficially resemble text in another. Figure 1 is an example of this from West London, in which the word <bollywood>[1] (written in Latin script and clearly intended to be read as English) mimics certain features and graphemes of the Devanagari script, e.g. the horizontal connecting bar and the grapheme <ठ> /ʈʰ/ that resembles Latin <o>.

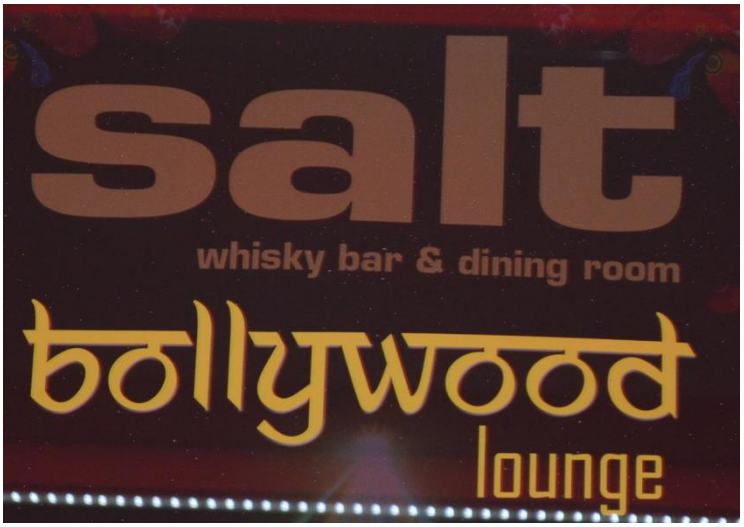


Figure 1: A typical example of writing system mimicry.[2]

In this paper I investigate the possibility that the choice to mimic another writing system may have a sociolinguistic explanation. In particular, I attempt to find answers to the following:

1. How have various academic fields and non-academic discourses described and accounted for writing system mimicry?
2. How might it fit into existing theories and frameworks of linguistic landscape studies?

[1] Throughout this article, I use the following standard transcription conventions: /phonemic/, [phonetic] and <orthographic>.
[2] Source for all figures: the author.

3. What is the extent of writing system mimicry in the linguistic landscape?
4. What methodological issues arise, and areas for further research become apparent, following initial research?

First I give an overview of the focus and some key findings of linguistic landscape studies so far (Section 2). Then I formulate a working definition of writing system mimicry (Section 3). I address question 1 with a summary of academic and non-academic mentions of the phenomenon (Section 4), and present the findings from a survey of the linguistic landscape of London (Section 5). Finally, I address questions 2, 3 and 4 with a discussion (Section 6) and conclusions and further research questions (Section 7).

2. Overview of the linguistic landscape literature

Linguistic landscape, at a basic definition, refers to the "linguistic objects that mark the public space" (Ben-Rafael et al. 2006: 7). Typically this has meant studying signs, but recent attempts to widen linguistic landscape research to include virtual and mobile spaces (e.g. the internet) have expanded its scope (Kasanga 2012). The term is thought to have originated with Landry and Bourhis (1997), although Spolsky and Cooper's (1991) research into language choice in Jerusalem was pioneering. Linguistic landscape study is one area of the flourishing field of the sociolinguistics of writing (see Lillis 2013; Sebba 2012), in which written language is seen no longer as neutral but as reflecting and defining sociolinguistic relationships.

2.1. Dichotomies of agency

Since the introduction of the term, two dichotomies that often run parallel have both been used productively and criticised for their lack of nuance. These are *top-down/bottom-up* and *official/non-official*. They highlight the important roles of the *agents* of a sign: those who enact control over it, also referred to as *authorship* (Malinowski 2009).

Top-down refers to "elements used and exhibited by institutional agencies which in one way or another act under the control of local or central policies" (note the explicit reference to language planning and policy, normally perceived of as top-down, with its terminology). Bottom-up elements, on the other hand, are "utilised by individual, associative or corporative actors who enjoy autonomy of action within legal limits" (Ben-Rafael et al. 2006: 10), and tend to include all manner of non-governmental social organisations and corporations. Generally (although this may not always be the case) top-down authorship is synonymous with official use, and bottom-up with non-official use.

On the one hand, these dichotomies have been informative. Genuine differences are often found between the uses of street signs employed by top-down/official contexts and bottom-up/non-official actors. The former tend to be much more regular, reflecting the language practises of the dominant linguistic group in the country, while non-official signs are much less regulated (Backhaus 2006), tending to reflect *solidarity* with others of the same ethnolinguistic (often minority) background in cities (Spolsky and Cooper 1991) through the incorporation of groups' language(s) (Huebner 2006).

However, they have been criticised for over-simplifying the relationships agents have with each other. They imply a two-tier hierarchy in which both major corporations and

fly-posters (two extremes of the bottom-up actors) are seen as equal. Not only might major corporations have considerably more control over signage than fly-posters, they may be actually more powerful than governments in this respect (Huebner 2009), reflecting the transfer of power from governments to corporations as a product of globalisation. Far from enjoying "autonomy of action within legal limits", bottom-up actors are sometimes subject to further constraint, especially by governments who wish to control an ethnic minority, or by city planners who wish to alter how an area is perceived. An example of the latter, blurring the lines between actors, is reported by Lou (2010). She finds that the high presence of Chinese glyphs on business signs in Chinatown, Washington D.C., is due to a combined higher corporate (non-official) and government (official) language planning that attempts to maintain the Chinese character of the area.

2.2. Agents and audiences

In their groundbreaking research into languages of historic and present Jerusalem, Spolsky and Cooper (1991: 81-85) formulated three conditions for writing in certain languages but not others on public signs. Essentially they state that choice of language is dependent upon the agents and audiences: a sign is written in a language known to the agent(s), presumed to be understood by the audience(s), and in a language by which the agent wishes to be identified. As Backhaus (2007) notes, the first two are based on practicality, while the third has a distinct social and political basis. The latter two considerations have had particular importance in conceiving the development of linguistic landscape studies as a field.

Given that the linguistic landscape is dominated by advertising in some form or another (Leeman and Modan 2010; Spolsky 2009), it is vital that audience be considered for practical purposes. Audiences are inherently multiple, making it difficult for agents to conceive how their signs may be viewed, and how effective they may be at delivering the intended message.

The assertion that agents consider how they wish to be identified, and thus express themselves symbolically in solidarity with their ethnolinguistic group through language on signs, has been many times validated since then (see, for example, Part IV of Shohamy and Gorter 2009).

2.3. Focuses

Generally, studies have focused on power relations between ethnic groups, in particular ethnic minorities' linguistic vitality in the landscape as indexing their local sociolinguistic status. In terms of the geographic regions, studies have been fairly sporadic depending on researchers' locations and interests, but tending to focus on urban areas. Studies of the British Isles so far appear to be limited to Ireland and Northern Ireland (Kallen 2009; Kallen and Ní Dhonnacha 2010).

2.4. Methodologies

Methodologies have so far been fairly consistent, involving photographing signs or groups of signs in an area, coding them for significant features, and making statistical comparisons to draw conclusions about the relationships between ethnolinguistic groups. While this has yielded some interesting results, it essentially ignores the agents' actual considerations when making the signs, something addressed by a special issue of the *International Journal of the Sociology of Language* (Zabrodskaja and Milani 2014)

entitled ‘Signs in Context’. The push towards qualitative methodologies such as in-depth interviews is exemplified by Malinowski’s (2009) conversations with Korean business owners.

3. Working definition of writing system mimicry

The term *typographic mimicry*, as coined by Coulmas (2014), provides a useful starting point for my definition of writing system mimicry. This was a first attempt at defining the concept for sociolinguistic purposes, after having found no mentions elsewhere (Coulmas p.c., December 2014). Despite disliking a proliferation of terms in academic literature, I choose not to continue with the word ‘typographic’. It usually refers to the act of arranging type, and is therefore too broad for current purposes, as it could reference any design elements of typography. The mimicry of Devanagari in Figure 1 is of very specific features of the script, rather than just any design elements. ‘Script’ could be used instead of ‘typographic’, covering many common examples of mimicry (e.g. faux-Arabic, faux-Cyrillic, faux-Devanagari), however that would exclude mimicry of another language that uses the same script but a different orthography, e.g. the grapheme <ö> in the logo of the heavy metal band Motörhead. Therefore the term ‘writing system’, defined according to Coulmas (2003: 35, first definition) as being “of an individual language” is preferred.

Coulmas (2014: 18) notes that design features being mimicked are salient to the mimicker – they are “conspicuous features” that distinguish the mimicked writing system from the base writing system. They may be real features of the writing system, or may be imagined by the mimicker, potentially indicating stereotypes held about the writing system. The ‘base writing system’ here is the lens through which the text is intended to be read (English in Figure 1). It is taken for granted that phonetic properties of graphemes mimicked are ignored, although it is possible that that may not always be the case.

With this in mind, I propose the following definition of writing system mimicry:

> The mapping of (real or imagined) design features and/or graphemes of a mimicked writing system onto a base writing system, so that the base writing system somewhat resembles the mimicked writing system while retaining legibility.

4. Mentions elsewhere: discourse analysis and literature review

4.1. Overview and methodology

This section is an overview of current terminology for, and explanations of, writing system mimicry. This takes the form of two methods: a brief discourse analysis of English-language internet websites and a cross-disciplinary academic literature review. The (limited) discourse analysis is intended to identify some non-academic or ‘folk’ terms and explanations for the phenomenon, while the literature review aims to assess how it has been treated in various academic fields. It is hoped that the findings from non-academic discourse may inform present and future research by scholars.

There are some important things to clarify from the outset. A first is that this research is exploratory: mentions and explanations are scattered in both surveys, suggesting that it has not been systematically assessed before. This informed the methodology, because due to the lack of a standard term, I had to employ a scattergun/snowball approach to

search terms on internet search engines and academic journal databases. For example, at one point it became clear that in the discourse the term *faux Cyrillic* was used to describe Latin-script text that mimicked the Cyrillic script, at which point I searched using this and derived terms (e.g. *faux Arabic*, *faux Chinese*) in both locations. This method was productive in that it enabled me to slowly identify which terms were in common use, but it obviously means that research could never be 'complete', as it is likely that more terms (and interpretations) exist, and furthermore, it limits my corpus to the English language. More specifically, for non-academic discourse I searched through the first 20 or 30 entries on popular search engine Google[3], and used Primo Central[4] for academic journal searches.

A second issue is that although this paper focuses on the linguistic landscape, writing system mimicry is not restricted to this domain. Several of the academic articles that appeared in the literature review analyse instances of writing system mimicry that do not form part of a classic linguistic landscape survey, e.g. product branding (Kurland 2004; McMichael 2009) and online communications (Miller 2011).

4.2. Academic and 'folk' terms

Despite (or perhaps because of) the lack of a standardised term for this phenomenon, various terms are identifiable. These include:

- *Faux x*, with *x* being either the name of the script mimicked or the language typically associated with the writing system. Examples of the former include *faux Cyrillic* (Know Your Meme 2014; McMichael 2009), *faux Devanagari* (Chachra 2014) and *faux Runic* (Page Studio Graphics 2015), and examples of the latter include *faux Russian* (Kurland 2004), *faux Arabic* and *faux Japanese* (Page Studio Graphics 2015).
- *Foreign look font* (Dafont 2015a);
- *Ethnic type* (Shaw 2009);
- *Mimicry typefaces* (Wikipedia 2015);
- *Wonton font* (Dafont 2015b) and *Chop Suey font* (Yang 2012) for the Chinese script;
- *Pseudoscript* (Coulmas 2014; Patel 2005);
- *Typographic mimicry* (Coulmas 2014).

All but one of these (typographic mimicry) refers to the product of the phenomenon rather than the phenomenon itself, implying that there is some kind of standard process for creation, and more-or-less uniform degrees of implementation.

4.3. Interpretations

4.3.1. From discourse analysis

Internet discourse repeatedly offers two themes of discussion: that it is a marketing tool, and that it relies on stereotypes.The terms *wonton font* and *Chop Suey font* (above) are both named after popular Chinese foods in the west, indicating an association with Chinese restaurants. This is corroborated by design historian Shaw (2009), who adds that it may be necessary as an advertising strategy: "fail to use this kind of lettering and

[3] http://www.google.com

[4]http://mlplus.hosted.exlibrisgroup.com/primo_library/libweb/

you run the risk of being overlooked … so there's a commercial incentive for takeout places to use this kind of typeface" (quoted in Yang 2012). A commenter on the same article concurs, saying that it is "designed to get customers". Seemingly contradictorily, given that it is unlikely that business owners would wish for their businesses to be portrayed negatively, writing system mimicry is at the same time seen by many potential customers as relying on negative stereotypes (see Shaw 2009). These are especially prominent when coupled with Chinese caricatures (see Strasburg 2002; Coulmas 2014: 18), causing some to accuse it of being a racist practice (Yang 2012, especially comments section).

4.3.2. From literature review

Findings from the literature review generally support the suggestions in the discourse analysis, in addition demonstrating evidence that writing system mimicry has an identity function.

Foreign branding is a marketing strategy used by businesses to draw attention to the foreignness (particularly the connection to a place) of their product, because specific products have positive associations with their country or region of origin. Kurland (2004) studied American consumers' perceptions of Russian vodka by manipulating two variables on the branding: the incorporation of the word 'Russian' and the writing system used. Participants favoured vodka that called itself Russian, unsurprisingly; interestingly for our purposes, out of the three writing systems (Latin, Cyrillic, and mimicked Cyrillic) they favoured mimicked Cyrillic, followed by Latin and then real Cyrillic. This can be explained as follows: real Cyrillic is "too confusingly unfamiliar to be found appealing" (Kurland 2004: no page), while a mimicked Cyrillic font occupies a comfortable middle ground between legibility and authentic Russianness.

Although not directly referencing writing system mimicry, the following comments from Leeman and Modan's (2010: 92) analysis of advertising in the linguistic landscape are illustrative:

> The use of a 'foreign' language as a selling point is heightened when that language has a different orthography from the language of the target consumer [...] For a viewer of an unfamiliar orthography, the linguistic valence of the writing system becomes backgrounded, and the aesthetic qualities become more salient [...] [which can be] capitalized on through font design.

In Kurland's study, the "font design" of branding on the vodka was more important than the "linguistic valence of the writing system".

Evidence to suggest that it perpetuates negative stereotypes is rarer. McMichael (2009:339) negatively describes the use of "faux-Cyrillic" on a late Soviet Russian album released for US audiences as a "cultural cliché", implying that some potential buyers or fans of the band may be put off by its over-stereotyping through mimicking Cyrillic.

The remainder of the references in academic sources discuss how writing system mimicry is a symbol of group identity – ethnic and social. A survey of the linguistic landscapes of three cities in Maharashtra, India, found combined Devanagari and Latin

scripts on shop fronts. Rubdy (2013: 44) categorises this as an example of *language mixing*, calling it an expression of *hybrid identity* that "[enables] them to navigate between global identification and local cultural practices". In this context, the global is expressed by indexing English, and the local by indexing Marathi. The desire to express multiple identities through writing is also stated by Patel (2005), who highlights font design students' attempts to create fonts that might be used by Indian diasporas that identify with multiple cultures.

Similarly, American Jews' mimicry of the Hebrew script in local flyers enables the diaspora to maintain a relationship with its own heritage. Similarly to the Russian-branded vodka example above, writing system mimicry enables the post-vernacular American-Jewish community (and those outside of it) to read text that looks distinctively Hebrew without being able to read the Hebrew script, or as Shandler (2006: 156, quoted in Benor 2009: 251) puts it, "[marking] the words as distinctively Jewish while integrating them into a more widely familiar communicative code".

A final example of group identity is found in Japan through what is termed by Miller (2011) *girl graphs*. This is the substitution of Japanese graphs with certain symbols and graphs from other scripts (especially Greek and Cyrillic letters that bear a resemblance to them). As the name suggests, it is done by young girls, partially to reaffirm group identity but also to subvert mainstream orthographic expectations.

4.4. Summary

A review of academic and non-academic discussions of writing system mimicry identifies the following. There is no uniform term for it, either in academic or non-academic discourses. Its use as an advertising or marketing tool is recognised in both surveys. It is perceived by some to perpetuate negative stereotypes, while in academia sporadic mentions have been made of its connection to identity. Not enough evidence yet exists to make broad claims about the truth of any of these, and connections to established sociolinguistic research areas are impoverished. I now turn to an account of a linguistic landscape survey of London.

5. Survey of the linguistic landscape of London (UK)

5.1. Background

Between December 2014 and January 2015, I conducted an on-foot survey of signs in several limited geographic areas within London's city centre. It is my understanding that this may be the first sociolinguistic survey of London's linguistic landscape, as I am unable to find any others. London has a long and rich history of immigration that has resulted in an abundance of multilingual signage, with specific areas being home to specific ethnic minorities. The survey aims to assess the prevalence of writing system mimicry in an urban city, and to act as a basis from which further similar research might develop.

5.2. Methodology

This survey generally followed the methodologies of other classic linguistic landscape studies (e.g. Ben-Rafael et al. 2006). Despite the clearly important role of agency in writing system mimicry, I did not choose to conduct interviews because in this initial survey I felt it was important to answer more basic questions, in particular how prevalent the phenomenon is, before researching in detail the motivations involved.

For each geographic area I would count and tally every relevant sign unit, coding them according to category of business they appeared on, and I would take photographs of every instance of what appeared to be writing system mimicry on these sign units (assessing the truth of these judgments afterwards). This provided me with essentially as complete as necessary a data set of each geographic area, enabling me to calculate writing system mimicry as a percentage of all relevant sign units. A pilot survey was done first to iron out initial methodological issues.

Definitions for the previous paragraph are:

- Sign unit. Following Hult (2009), every store front equalled one unit, including in that definition signs or groups or signs for businesses not at street level. I included all writing visible at the store front, including menus, the names of objects being sold and advertisements. Store fronts were counted rather than individual signs as first, the latter would be too much over several geographic areas, and secondly, it was assumed that a store front reflected the decisions of one or a group of agents working collaboratively to make decisions about how the business is perceived (see Section 2.2).
- What to tally. Unlike most linguistic landscape surveys, which often look for signs of ethnic groups' written language vitality and function, I was looking for a particular feature of minority (immigrant) language use. Therefore I only tallied units that showed signs of being owned by, or providing a service to, ethnic minorities. Due to some ambiguous examples in the pilot, I chose only to tally units that demonstrated the minority's language in some form, either in the original or mimicked writing system or a Latin transliteration. I chose to focus on the dominant minority for each area because I wanted to find out if some immigrant groups are more likely to use writing system mimicry than others. Without approaching the owners of individual stores (not feasible due to time constraints on the research), this unfortunately involved approximating which ethnic group owned which store (to be discussed later).
- Choice of geographic areas. Geographic areas were chosen that had a dominant immigrant minority, and where the language had a significant visual presence in the local linguistic landscape. This could be a street or a network of streets. Some places (e.g. Chinatown and Brick Lane) were initially selected through my own folk knowledge about their ethnic composition, which was confirmed through figures available from the latest UK census (Nomis 2014). Because street-level census data is not available, postcode area data had to be used, which is assumed to equate approximately to business ownership in the local area. The dominant immigrant minorities were selected based on having a uniform non-Latin script for their language(s).
- Coding units. Based on evidence from the literature and discourse analysis about the importance of audience and advertising in writing system mimicry and the

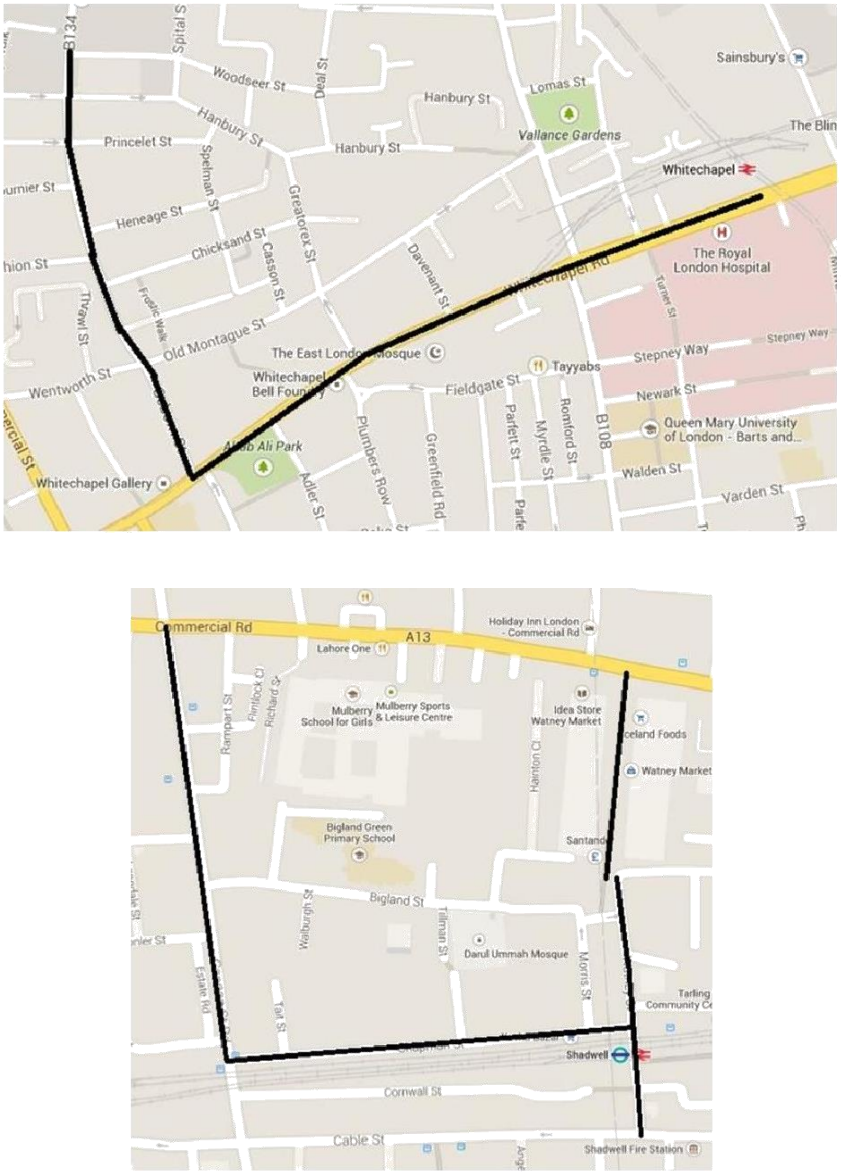
linguistic landscape, I decided to code for two broad categories of commercial outlets, again aiming to see if there is a difference in the prevalence of writing system mimicry. Such assumptions are not without objection, but had to be made on the basis that: there is no standard coding system used in historic linguistic landscape studies; I was not able to approach individual stores for information about their customers; and that this research was preliminary and exploratory. The two categories coded were:

  - Eateries. These were any sit-down or takeaway food or drinks establishments whose primary products associated with the ethnic group's culture. It was assumed that these, of all commercial types, were most likely to be aimed at those outside the ethnic group.
  - Other services. This encapsulated a wide range of commercial and cultural activity centres including ethnic grocery stores, religious buildings, massage parlours, clothing stores and travel agents. It was assumed that these were mostly frequented by those within the ethnic group.
- Identifying writing system mimicry. Despite the working definition, this became an issue at the analysis stage (at which point I erred on the side of caution, only including results that I was certain of), but at the time of data collection I decided to photograph all instances of suspected mimicry.

The three geographic areas chosen for the survey were (speaker data from Nomis 2014):

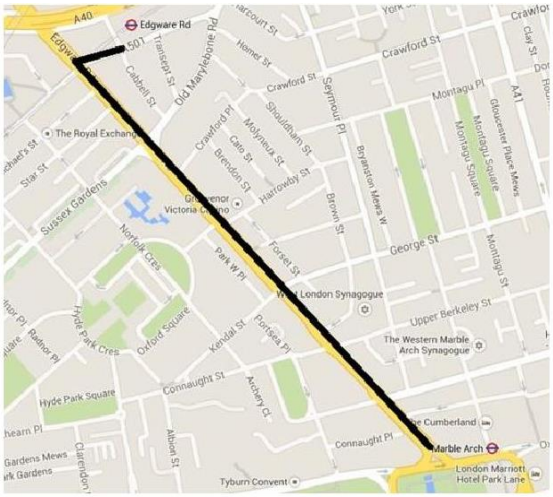
- East London (Brick Lane and surrounding areas, popularly referred to as 'Banglatown'). 13% are speakers of Bengali/Sylheti/Chatgaya (Chittagonian) as a main language (all in the Bengali-Assamese branch of the Indo-Aryan language family, typically written with the Bengali script). Postcode areas: E11/E12/E15/E16. 2.44km surveyed.
- Central London ('Chinatown'). 4.4% are speakers of a Chinese language (written using Chinese glyphs) as a main language. Postcode areas: W1D5/W1D6/WC2H7/WC2H0. 2.17km surveyed.
  West London. 6.6% are speakers of Arabic (using Arabic script) as a main language. Postcode areas: W21/W22. 1.08km surveyed.

Maps 1-3 illustrate the extent of each geographical area. Roads surveyed are highlighted with a thick black line. The total area surveyed was 5.7km.



Map 1: East London survey area. Source: Google (2015).

Map 2: Central London survey area. Source: Google (2015).



Map 3: West London survey area. Source: Google (2015).

It is important to note some features of these areas. Two short sections of the East London area (a stretch of Whitechapel Road and Watney Market), in addition to having regular shops, are host to daytime markets catering primarily for the Bangladeshi diaspora. These had to be ignored for consistency – none of the other areas had them – but such markets might exhibit different patterns of language use. Additionally, the Brick Lane area of East London and the entire Central London area are famous for their restaurants as tourist attractions, with the latter especially catering to theatre- and cinema-goers.

5.3. Results

In total, 513 shop fronts were counted, of which 21 (4.1%) exhibited signs of writing system mimicry, with representation in every geographic area and business type surveyed, indicating consistently moderate use throughout. See Table 1 for full details. The Appendix contains all instances of relevant writing system mimicry documented, including information about geographic area, business category, the mimicked writing

system, and any limitations to the extent of mimicry. An overview of some of the distinctive features of the results follows.

Table 1: Prevalence of writing system mimicry in surveyed London areas.

	Eateries		Other services		Total	
	No mimicry	Mimicry	No mimicry	Mimicry	No mimicry	Mimicry
East London (*n*=225)	78	5 (6.0%)	138	4 (2.8%)	216	9 (4.0%)
Central London (*n*=172)	89	4 (4.3%)	78	1 (1.3%)	167	5 (2.9%)
West London (*n*=116)	32	6 (15.8%)	77	1 (1.3%)	109	7 (6.0%)
Total (*n*=513)	199	15 (7%)	293	6 (2.0%)	492	21 (4.1%)

For every area, mimicry occurred more in eateries than in other services. Mimicry within each area ranged between 2.9% and 6%. The areas displayed some internal differences. Of particular note is that over 15% of the relatively few Arabic-language eateries in West London exhibited mimicry, which is considerably more than in the other areas. However, three of these were restaurants owned by the same company 'Maroush' (see Appendix). Counting only one of those instances instead of all three – as the primary actor (business owner) is presumably the same across the chain – the percentage is lowered to 11.1%, which is still much higher than other areas. The East London area exhibited considerably more writing system mimicry in the north (top Map 1, which includes Brick Lane and Whitechapel Road) than in the south (bottom Map 1).

In terms of the languages mimicked, Arabic dominates the data. West London (Arabic-dominant) mimicked only the Arabic script. Central London mimicked predominantly the Chinese script (with the exception of Manchu on a Manchurian restaurant). Strikingly, only one East London example mimicked the Bangla script, while the rest mimicked Arabic. Finally, although this was not initially coded for, analysis reveals that every instance of writing system mimicry was a very prominent sign on the store front, with all but one being the business name itself.

6. Discussion

In this section I consider everything mentioned so far in a discussion of my findings. I first address the results of the survey and the literature review, and then highlight methodological concerns and attempt to contextualise these results within the linguistic landscape literature.

The prevalence of writing system mimicry across London not only validates it as a phenomenon worth studying, but suggests that the potential deterrent of negative stereotyping (identified in Section 4.3) may not be strong. The prominent size of such signs also suggests that where it is employed it is done so very purposefully, either for advertising or identity purposes, or both.

That it is present in eateries more than non-eateries could be due to advertising - the business owners may wish to target those outside the ethnolinguistic group - or it could

be an act of solidarity with other members of the group, if eateries are considered to be a more communal space within that culture.

The dominance of Arabic in the data in general – both in terms of the quantity of mimicry in West London and the use of Arabic mimicry by Bangladeshi East London – deserves further research. A simple possible answer is that perhaps Arabic is deemed to be easy to mimic, and is therefore more incorporable into font design than other writing systems, although this does not explain its use in East London. According to census data (Nomis 2014), only 0.5% of the population in the East London area report speaking Arabic as their main language, so it is not likely that writing system mimicry here expresses a language they use. It may, however, express a language they identify with but not speak. Fairly similarly to American Jews' use of Hebrew to express identity (Benor 2009), the predominantly Muslim Bangladeshis may identify with Arabic writing because of religious association with the Qur'an. This is supported by the use of writing system mimicry with words that evoke Arab or religious associations: 'Arabian', 'Aladin', and especially 'bismillah' ('in the name of Allah') (see Appendix, Table 2).The pairing of the Arabic script with Islam is perhaps most visually explicit in the West London Syrian restaurant 'Abu Zaad' (Figure 2), in which features of the Arabic script are combined with stylised minarets. A strong connection of language to culture through religion may indeed explain the dominance of Arabic mimicry in West London eateries, though not fully.

Figure 2: Incorporating religious iconography.

In predominantly Chinese Central London, one example of mimicry that was not included in the data (as non-Chinese agency was assumed), but was still initially photographed, was a very overt example of mimicking a different writing system. Figure 2 shows a Vietnamese restaurant, the sign for which mimics the Chinese script. It can be assumed that the sign-makers did not consider the audience to know that Vietnamese does not use Chinese characters. Why, though, incorporate Chinese at all? It could be a marketing motivation, to associate with the 'Chineseness' of the surrounding restaurants. If the owners are Vietnamese, it may be the case that here, the desire to advertise overrides the desire to express identity through writing. Alternatively, the owners could be ethnically Chinese (from Vietnam or elsewhere), wishing to identify as such through their business sign.

Figure 3: Mimicry of another writing system.[5]

Another example in Chinatown – a Manchurian restaurant called 'Manchurian Legends' – mimics the Manchu script. The letter <L>, for example, resembles multiple graphs including a rotated <ᡴ>. It is unlikely that many members of the potential audience would recognise the Manchu script in the sign, so once again it is possible that this is an expression of Manchu identity, or perhaps just a way to make the letters look distinct from others in the area.

Entering the field, with little background evidence to suggest otherwise, it was assumed that writing system mimicry would be a simple phenomenon to identify: either a sign exhibits it, or it does not. Most popular terms for it (e.g. *faux-script*) refer to the product of mimicry rather than the process itself, suggesting that there is typically wholesale incorporation of writing system mimicry.

However, my research suggests otherwise. Although font designers may develop so-called pseudoscripts (e.g. Dafont 2015a; Page Audio Graphics 2015), the majority of the examples identified in this survey incorporate a limited amount of features. In fact it is arguable that defining a pseudoscript is impossible, as the logical extreme of mimicking as many features and graphs as possible is to essentially recreate the writing system. Writing system mimicry is therefore necessarily a compromise, and any attempt to define a pseudoscript (against something that is not a pseudoscript) is at best subjective and at worst unproductive.

Two analyses of the data illustrate these points: an assessment of which features and graphs of each script are commonly mimicked, and a discussion of some ambiguous examples.

The first is an admittedly post-hoc identification of features. One very prominent issue with current methodology and state of research is that there is of yet no explicit identification or listing of these. Further, as Coulmas (2014: 19) mentions, an assessment of which features and graphs are mimicked by different speech communities may indicate perceptual stereotypes about the writing system. This does not fall within the scope of my study, nor am I equipped with the typographic knowledge to attempt it. A casual observation reveals the following. Mimicking Arabic tends to incorporate: diacritic dots; a bottom line typical in handwritten joined Arabic; changes to letter thickness representing calligraphic pen strokes; overt, slightly exaggerated curves; and occasional graphs, including <ﺲ>and<ﺳ> for Latin <s>. Mimicking Chinese tends to incorporate fewer unique features, mostly limited to 'rough', approximate calligraphy, representing handwritten characters, and predominantly straight lines. Only one usable example was found that mimicked Bengali script, incorporating most noticeably the

[5]As of May 2015, this sign has now been replaced with one using plain Latin letters.

horizontal top bar, which is indicative of several Indic scripts, and the connection of <N> to the bar with a loop that resembles a vowel in Bengali.

As it is not yet clear which features are typically mimicked, or even whether or not all the examples found were even intended by the actors, identifying intended writing system mimicry in the linguistic landscape proved very difficult. A number of ambiguous examples were found and later excluded from the data. It is likely that I photographed them because I was overanalysing the signs around me, trying too hard to identify that, say, curved writing in Figure 4 (Bangladeshi East London) was a mimicked feature of Arabic. Less frequently, as in Figure 5, there were signs that may actually have been non-English languages but that I took to be attempts at mimicry.

Figure 4: Is there any Arabic here?

Figure 5: Is there any English here?

A number of methodological concerns arose from this research. One clear conclusion is that a focus on the decision-making processes of agents, which could be gleaned through interviews, is essential to truly understanding motivations for incorporating writing system mimicry. Another issue with the current methodology comes from the nuances of its employment, in particular how to identify and describe the mimicry of individual features. There is also the question of who exactly the sign-makers are. The methodology used here forced me to approximate the agency of some signs, naturally (most likely) resulting in errors.

Writing system mimicry may be roughly contextualised within the linguistic landscape literature as follows. As it is not apparent on any government signs, rather only on private businesses, and appears to have identity and solidarity functions, it seems to form part of the bottom-up and non-official parts of the dichotomies of Section 2.1. It

does not exclusively serve the purposes of identity and solidarity, though: due to being read as the dominant language, it also serves an instrumentality function, as such being a very productive compromise between those goals that are often only achievable through bilingual signage.

7. Conclusions and further questions

This research began with four initial questions:

1. How have other academic fields and non-academic discourses described and accounted for writing system mimicry?
2. How might it fit into existing theories and frameworks of linguistic landscape studies?
3. What is the extent of writing system mimicry in the linguistic landscape?
4. What methodological issues arise, and areas for further research become apparent, following initial research?

All of these are too broad to answer fully within the scope of one article. The intention was to present preliminary research that should be expanded upon and clarified by other scholars. Assessments of academic literature and non-academic discourses revealed a lack of collaboration between, or even recognition of, previous studies; a lack of uniform terminology; and perceptions that writing system mimicry relates to identity, relies on stereotypes, and is often a marketing tool. An analysis of the linguistic landscape literature suggested some dichotomies it may fit into; that it likely considers both instrumentality and symbolism; and that the field's recent methodological focuses on signs' constructions, contexts and agents may be productively applied here.

Writing system mimicry is applied by immigrants to London to a limited but still significant amount of business signs, with variation apparent across type of business and ethnolinguistic group. It is promising, given a general lack of research elsewhere, that so many additional nuances about the implementation of writing system mimicry were revealed. Writing system mimicry has been shown to be a distinctive feature of the linguistic landscape that is a rich source of sociolinguistic research. Far from answering any of the research questions fully, this paper has identified a large range of follow-up questions, including:

- In any given situation, what weighting do marketing and identity have on the decision to incorporate writing system mimicry?
- Why is writing system mimicry more commonly used by owners of eateries than non-eateries? Research into those businesses' clienteles may prove informative here.
- Are the trends identified here consistent with those of other cities?
- What are the motivations to mimic writing systems used by other ethnolinguistic groups, or those that are unlikely to be recognised by the majority of the audience (e.g. the Arabic/Bengali and Manchu examples)?
- What effect, if any, does negative stereotyping have on the use of writing system mimicry?
- Which features are mimicked by which groups, and why? I echo Coulmas (2014) here.

Appendix: All examples of writing system mimicry identified in the survey

Table 2: East London

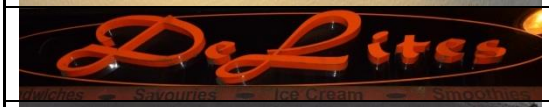
Example	Business category	Mimicked writing system	Limitations
	Other service	Arabic	
	Other service	Arabic	
	Other service	Arabic	
	Other service	Arabic	
	Eatery	Arabic	Only the bottom of graph <L>
	Eatery	Arabic	
	Eatery	Arabic	
	Eatery	Bengali	
	Eatery	Arabic	

Table 3: Central London

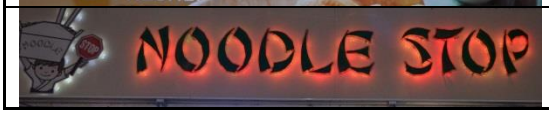
Example	Business category	Mimicked writing system	Limitations
	Other service	Chinese	
	Eatery	Chinese?	
	Eatery	Manchu	Only the graphs <M> and <L>
	Eatery	Chinese	
	Eatery	Chinese	

Table 4: West London

Example	Business category	Mimicked writing system	Limitations
	Other service	Arabic	
	Eatery	Arabic	
	Eatery	Arabic	
	Eatery	Arabic	Mostly only the graphs <I> and <R>
	Eatery	Arabic	
	Eatery	Arabic	
	Eatery	Arabic	

References

Backhaus, P. (2006) 'Multilingualism in Tokyo: A Look into the Linguistic Landscape'. *International Journal of Multilingualism* 3(1), pp. 52-66.

Backhaus, P. (2007) *Linguistic Landscapes: A Comparative Study of Urban Multilingualism in Tokyo*. Clevedon: Multilingual Matters.

Ben-Rafael, E., Shohamy, E., Hasan Amara, M. and Trumper-Hecht, N. (2006) 'Linguistic Landscape as Symbolic Construction of the Public Space: The Case of Israel'. *International Journal of Multilingualism* 3(1), pp. 7-30.

Benor, S. (2009) 'Do American Jews Speak a 'Jewish Language'? A Model of Jewish Linguistic Distinctiveness'. *Jewish Quarterly Review* 99(2), pp. 230-269.

Chachra, D. (2014) 'Kern Your Enthusiasm (10)', *HiLoBrow*, 10 Aug, [Online], Available: http://hilobrow.com/2014/08/10/kern-your-enthusiasm-10/ [2 Jan 2015].

Coulmas, F. (2003) *Writing Systems: An Introduction to their Linguistic Analysis*. Cambridge: Cambridge University Press.

Coulmas, F. (2014) 'Writing systems and language contact in the Euro- and Sinocentric worlds'. *Applied Linguistic Review* 5(1), pp. 1-21.

Dafont. (2015a) *Foreign look fonts*, [Online], Available: http://www.dafont.com/mtheme.php?id=2 [7 Jan 2015].

Dafont. (2015b) *Wonton Font*, [Online], Available: http://www.dafont.com/wonton.font [7 Jan 2015].

Google. (2015) *Google Maps*, [Online], Available: https://maps.google.co.uk/ [10 Jan 2015].

Huebner, T. (2006) 'Bangkok's Linguistic Landscapes: Environmental Print, Codemixing and Language Change'. *International Journal of Multilingualism* 3(1), pp. 31-51.

Huebner, T. (2009) 'A Framework for the Linguistic Analysis of Linguistic Landscapes'. In Shohamy, E. and Gorter, D. (eds.) *Linguistic Landscape: Expanding the Scenery*, pp. 70-87. London: Routledge.

Hult, F. (2009) 'Language Ecology and Linguistic Landscape Analysis'. In Shohamy, E. and Gorter, D. (eds.) *Linguistic Landscape: Expanding the Scenery*, 88-104. London: Routledge.

Kallen, J. (2009) 'Tourism and Representation in the Irish Linguistic Landscape' In Shohamy, E. and Gorter, D. (eds.) *Linguistic Landscape: Expanding the Scenery*, pp. 270-284. London: Routledge.

Kallen, J. and Ní Dhonnacha, E. (2010) 'Language and Inter-language in Urban Irish and Japanese Linguistic Landscapes'. In Shohamy, E., Ben-Rafael, E. and Barni, M. (eds.), *Linguistic Landscape in the City*, 19-36. Bristol: Multilingual Matters.

Kasanga, L. (2012) 'Mapping the linguistic landscape of a commercial neighbourhood in Central Phnom Penh'. *Journal of Multilingual and Multicultural Development* 33(6), pp. 553-567.

Know Your Meme. (2014) *Faux Cyrillic*, [Online], Available: http://knowyourmeme.com/memes/faux-cyrillic [9 Jan 2015].

Kurland, J. (2004) *American Perceptions of Vodka Shaken, Not Stirred: An Analysis of the Importance of Vodka's Foreign Branding Cues and Country-of-Origin Information*, [Online], Available: http://ccat.sas.upenn.edu/~haroldfs/popcult/handouts/adverts/goodones/JKurland.html [7 Jan 2015].

Landry, R. and Bourhis, R. (1997) 'Linguistic landscape and ethnolinguistic vitality: an empirical study'. *Journal of Language and Social Psychology* 16(3), pp. 23-49.

Leeman, J. and Modan, G.. (2010) 'Selling the City: Language, Ethnicity and Commodified Space'. In Shohamy, E., Ben-Rafael, E. and Barni, M. (eds.), *Linguistic Landscape in the City*, 182-198. Bristol: Multilingual Matters.

Lillis, T. (2013) *The Sociolinguistics of Writing*. Edinburgh: Edinburgh University Press.

Lou, J. (2010) 'Chinese on the Side: The Marginalization of Chinese in the Linguistic and Social Landscapes of Chinatown in Washington, DC'. In Shohamy, E., Ben-Rafael, E. and Barni, M. (eds.), *Linguistic Landscape in the City*, pp. 96-114. Bristol: Multilingual Matters.

Malinowski, D. (2009) 'Authorship in the Linguistic Landscape: A Multimodal-Performative View'. In Shohamy, E. and Gorter, D. (eds.), *Linguistic Landscape: Expanding the Scenery*, pp. 107-125. London: Routledge.

McMichael, P. (2009) 'Prehistories and Afterlives: The Packaging and Re-packaging of Soviet Rock'. *Popular Music and Society* 32(3), pp. 331-350.

Miller, L. (2011) 'Subversive script and novel graphs in Japanese girls' culture'. *Language and Communication* 31, pp. 16-26.

Nomis. (2014) *Census 2011 Main Language*, [Online], Available: http://www.nomisweb.co.uk/census/2011/wd204ew [16 Dec 2014].

Page Studio Graphics. (2015) *PIXymbols Faux Fonts*, [Online], Available: http://www.vershen.com/simfonts.html [2 Jan 2015].

Patel, M. (2005) 'Search for Vernacular Identity'. *Design Issues* 21(4), pp. 32-40.

Reh, M. 2004. 'Multilingual writing: a reader-oriented typology – with examples from Lira Municipality (Uganda)'. *International Journal of the Sociology of Language* 170, pp. 1-41.

Rubdy, R. (2013) 'Hybridity in the Linguistic Landscape: Democratizing English in India'. In Rubdy, R. and Alsagoff, L. (eds.), *Global-Local Interface and Hybridity: Exploring Language and Identity*, pp. 43-65. Bristol: Multilingual Matters.

Sebba, M. (2012) *Spelling and Society*. Cambridge: Cambridge University Press.

Shandler, J. (2006) *Adventures in Yiddishland: Postvernacular Language and Culture.* California: University of California Press.

Shaw, P. (2009.) 'Stereo Types', *Print*, 17 June 2009, [Online], Available: http://www.printmag.com/article/stereo_types/ [2 Jan 2015].

Shohamy, E. and Gorter, D. (2009) *Linguistic Landscape: Expanding the Scenery.* New York: Taylor and Francis.

Spolsky, B. (2009) 'Prolegomena to a Sociolinguistic Theory of Public Signage'. In Shohamy, E. and Gorter, D. (eds.), *Linguistic Landscape: Expanding the Scenery*, pp. 25-39. London: Routledge.

Spolsky, B. and Cooper, R. (1991) *The Languages of Jerusalem*. Oxford: Clarendon Press.

Strasburg, J. (2002) 'Abercrombie recalls T-shirts many found offensive', *SFGate*, 19 Apr 2002, [Online], Available: http://www.sfgate.com/news/article/Abercrombie-recalls-T-shirts-many-found-offensive-2849480.php [7 Jan 2015].

Wikipedia. (2015) *Typeface*, [Online], Available: http://en.wikipedia.org/wiki/Typeface#Mimicry_typefaces [7 Jan 2015].

Yang, J. (2012) 'Is Your Font Racist?', *Wall Street Journal*, 20 June 2012, [Online], Available: http://blogs.wsj.com/speakeasy/2012/06/20/is-your-business-font-racist/ [2 Jan 2015].

Zabrodskaja, A. and Milano, T. (2014) 'Signs in context: multilingual and multimodal texts in semiotic space'. *IJSL* 228, pp. 1-6.