IBERIA
An International Journal of Theoretical Linguistics

The Internal Structure of Compounds: A phase account of Aphasia[1]

Shanti Ulfsbjorninn
su1@soas.ac.uk
SOAS – University of London

Abstract: This study uses aphasia to support a phase-based derivation of compounds. Our research is nestled within the overarching and truly foundational debate between holists (Butterworth 1983, Bybee 2001, Starosta to appear) and atomists (Taft and Forster 1975, Rastle et al. 2004, Fiorentino and Poeppel 2007). The former camp maintains that compounds are stored devoid of any internal morphological structure; while the latter insist that compounds are derived by concatenation of constituent parts. Morpho-phonological analysis of the contrasting behaviour of simplex and compound words in Dinka and English (based on Kaye 1995) bears a striking similarity to the derivation by phase (Chomsky 2001) (cf. Newell and Piggott 2006, Newell and Scheer 2008, Scheer 2008, forth.). To confirm this novel phase-based account, *contra* the holists' null-hypothesis, we ran an experiment. We tested an aphasic patient (RC), who produced high error rates with trisyllabic simplex words and negligible error rates with disyllabic simplex words. The divisive question: What would trisyllabic compounds pattern with? The surface inclined holists predict they should pattern with the long simplex words; conversely, the atomist, for whom a trisyllabic compound will be processed either [[σ σ] [σ]] or [[σ] [σ σ]], predict they should pattern with the short simplex words. The latter turns out to be correct. Our experiment shows a compound is derived by independently sending its constituent parts to spell out, once there the constituent parts are no longer accessible to grammatical operations.

Keywords: Derivation by phase, Phonology, Aphasia, Compounds, Analytic Morphology, Phase Impenetrability Condition, Dinka.

---

[1] Thanks to RC for letting himself be frustratingly tested in the interest of phase theory. I also thank Dr Michele Miozzo for allowing me unfettered access to RC. A big thanks to Professor Bert Vaux for his teaching and company while I was at Cambridge. Critical thanks go to Professor Tobias Scheer for really putting the research in perspective in that Serbian eatery in Hungary, and to (the recently Dr) Bridget Samuels for so often thinking on different lines about the same topics… and then getting her lines all published. All errors are my own, except for where clearly referenced.

Resumen: Este trabajo usa el fenómeno de la afasia para apoyar una derivación de los compuestos basada en el concepto de fases. Nuestra investigación se enmarca dentro del debate general y fundamental entre holistas (Butterworth 1983, Bybee 2001, Starosta en prensa) y atomistas (Taft y Forster 1975, Rastle et al. 2004, Fiorentino y Poeppel 2007). Los primeros sostienen que los compuestos son almacenados sin ningún tipo de estructura morfológica interna; por el contrario los últimos insisten en que los compuestos se derivan a través de la concatenación de ciertos constituyentes. El análisis morfo-fonológico del comportamiento paradójico por parte de las palabras simples y compuestas en Dinka e Inglés (basado en Kaye 1995) muestra una similitud chocante con el fenómeno de la derivación por fases (Chomsky 2001) (cf. Newell y Piggott 2006, Newell y Scheer 2008, Scheer 2008, en adelante.). Para verificar esta nueva versión basada en la noción de fases, como contradicción a la hipótesis-nula llevada a cabo por los holistas, realizamos un experimento. Probamos con un paciente afásico (CR), el cual tuvo un alto porcentaje de errores con palabras simples de tres sílabas así como un promedio de error insignificante con palabras simples de dos sílabas. La pregunta divisoria sería la siguiente: ¿Con qué se corresponderían los compuestos de tres sílabas? Los que en apariencia apoyan a los holistas sugieren que estos deberían tener un comportamiento similar a las palabras simples y largas; por el contrario, los atomistas, para quienes un compuesto trisilábico ha de ser procesado bien como [[σ σ] [σ]] o [[σ] [σ σ]], establecen que estos se asemejan a las palabras simples y cortas. Estos últimos resultan ser los que están en lo cierto. Nuestro experimento corrobora que un compuesto se deriva a través del envío de sus constituyentes por separado a la fase de materialización, de tal manera que una vez allí dichos constituyentes dejan de ser accesibles a operaciones gramaticales.

Palabras clave: Derivación por fases, Fonología, Afasia, Compuestos, Morfología Analítica, Condición de Impenetrabilidad de Fases, Dinka.

Resumo: Este estudo recorre à afasia para confirmar a derivação por fases de compostos. A nossa pesquisa enquadra-se no debate global e verdadeiramente fundacional entre holistas (Butterworth 1983, Bybee 2001, Starosta a surgir) e atomistas (Taft e Forster 1975, Rastle et al. 2004, Fiorentino e Poeppel 2007). Os primeiros defendem que os compostos são armazenados desprovidos de qualquer estrutura morfológica interna; enquanto os últimos insistem que os compostos derivam da concatenação de partes constituintes. A análise morfo-fonológica do comportamento contrastante de palavras simples e compostas em Dinka e em Inglês (baseada em Kaye 1995) apresenta uma semelhança assinalável com a derivação por fase (Chomsky 2001) (cf. Newell e Piggott 2006, Newell e Scheer 2008, Scheer 2008, etc.). Para confirmar esta nova abordagem baseada em fases, *contra* a hipótese nula dos holistas, levámos a cabo uma experiência. Testámos um paciente afásico (RC), que produziu elevadas taxas de erro com palavras trissilábicas simples e taxas de erro pouco significativas com palavras dissilábicas simples. A questão decorrente: Que padrão seguem os compostos trissilábicos? Os holistas, baseados na superfície, predizem que estes seguem o padrão das palavras simples longas;

inversamente, os atomistas, para quem um composto trissilábico é processado como [[σ σ] [σ]] ou [[σ] [σ σ]], predizem que seguem o padrão das palavras simples curtas. Os últimos estão correctos. A nossa experiência demonstra que um composto é derivado, enviando independentemente as suas partes constituintes para serem decifradas quando estas não se encontram mais acessíveis a operações gramaticaisFormato de letra *Palatino Linotype* tamaño 12; interlineado de 1'2 y espacio entre párrafos de 6ptos. Formato de letra Palatino Linotype tamaño 12; interlineado de 1'2 y espacio entre párrafos de 6ptos. Formato de letra *Palatino Linotype* tamaño 12; interlineado de 1'2 y espacio entre párrafos de 6ptos; interlineado de 1'2 y espacio entre párrafos de 6ptos.

Palavras-chave: Derivação por fase, Fonologia, Afasia, Compostos, Morfologia analítica, Condição da Impenetrabilidade da Fase, Dinka.

1. Theories on Compounds

1.1. The Holistic Camp

The holistic camp encompasses a number of researchers, but is perhaps epitomised by the following approaches. The 'full listing hypothesis' (Butterworth 1983) argues that the entire word, irrespective of its morpho-phonological complexity, is stored as a whole in the lexicon. Other modern proponents of the holistic camp include the 'whole word' morphologists (Ford et al. 1997, Starosta to appear), and Joan Bybee's (2001) statistically driven mental lexicon. These approaches, irrespective of the details, all claim that compounds, even the most transparent, are stored in the lexicon without any special internal morpho-phonological structure. Compounds, in this view, are essentially no different to simplex words.

(1) Compounds holistically stored with no internal structure

/accordion/	(n.) a musical instrument
/black/	(adj.) colour
/bird/	(n.) feathered animal
/blackbird/	(n.) male is a black bird, yellow beak, expect to find in UK
/cat/	(n.) 'miao', long tail, feline, purrs
/cats/	(n.) more than one 'miao', long tail, feline, purrs
/house/	(n.) building to live in
/flap/	(n.) something on a hinge
/cathouse/	(n.) ditto: whorehouse
/catflap/	(n.) a door for cats to enter a house

In this lexicon, there is no differentiation between simplex and compound words. This is at the core of all holistic approaches to compounds, no matter how richly detailed (see Butterworth 1983, Bybee 2001, Starosta to appear). Conceptually, it is easy to understand why some uphold the holistic hypothesis. Why should compounds be treated differently to simplex words? Even the most transparent of compounds have a rather independent meaning to the sum of their parts. The word /blackbird/, for instance, refers to its species, not simply to birds that are black, the word refers to its pink chicks, and to its brown females and juvenile males, or to the (presumably purplish) plucked blackbird. It is intuitive to think of /blackbird/ as having a semantic identity independent from /black/ and /bird/; so, *ceteris paribus*, /blackbird/ should have its own lexical entry, an analogue to any simplex word. We take the holistic view, however, as nothing more than the null hypothesis.

(2) H_0: compounds are stored and processed like simplex lexical words.

The null hypothesis in (2) creates a number of theoretical predictions, the most basic of which is that compounds and simplex words should behave identically in tests specifically designed to show putative internal morphological structure (broadly defined). Examination of this null hypothesis leads to the formation of the second camp, the proponents of compositionality.

1.2. The Atomistic (Compositional) Hypothesis

The essence of this camp is that compounds have an internal, morphological, representation which specifically interacts with their processing (Taft & Forster 1975, 1979, Marlsen-Wilson et al. 1994, Kaye 1995, Rastle et al. 2004, Fiorentino & Poeppel 2007). This camp (either explicitely or inferentially) differentiates transparent from opaque compounds (Fiorentino & Poeppel 2007 and references therein); the latter are indistinct from simplex words.

Evidence for this compositionality, as we see it, comes from two quarters the experimental and the theoretical. We will review some key literature of both camps, and present what the latest morpho-phonological theory defines compounds as.

1.2.1. Experimental Evidence for the Atomists

The experimental evidence all comes from an exploration of the null hypothesis in (2).

Marlsen-Wilson et al. (1994) show that transparent compounds are primed by their putative internal components. If a patient is exposed to /black/ and /bird/ this will decrease the response time in recognising the stimulus /blackbird/; this same exposure will have no effect on the response rate in recognising the stimulus /bookshelf/. Opaque compounds, however, were not primed by stimulus which did not appear to be morphologically related, so: /cod/ and /piece/ would not prime /codpiece/. Rastle et al. (2004) conducted a similar study where they showed that /cleaner/ primed /clean/, /corner/ primed /corn/, but /brothel/ did not prime /broth/. Although, Rastle et al. (2004) did not explore compounds, the results point to the same overarching point: lexical form and meaning are insufficient as organising principles of the lexicon, rather, morphological structure is critical to the organisation of the lexicon and, by inference, to the processing of compounds.

Likewise, Fiorentino & Poeppel (2007) carried out an important study demonstrating differences in processing between opaque compound words and transparent compound words. Concluding their thorough MEG study, Fiorentino & Poeppel (2007) support the atomist's findings[2], stating that their results are compatible with a dual-route model, where, *specifically*, the internal morphological structure of compounds may be accessed by activation of the internal parts (constituents).

These representative empirical studies, explicitly or not, take the null hypothesis in (2) and find it wanting. They seem to conclude that transparent compounds contain internal morphological structure, which affects the internal organisation of the lexicon and, by inference, any related processing of compounds.

1.2.2. Morpho-Phonological Support

Morpho-Phonological theory also comes out in support of an atomistic view of compound morphology.

[2] Especially those of Taft and Forster (1975; 1979).

Take the following compounds from Dinka (Nilo-Saharan), the problems for the holist come from contrasting what we see in Dinka compounds and what word-structures and consonant clusters (CC) Dinka allows.

(3) Compounds of Dinka (Malou 1988:18)

(a)	tʷoŋdʒaŋ	'Dinka language'
(b)	nanco:l	'black calf'[3]
(c)	nʲakdur	'early morning'

The holist would have to claim these lack any internal constituency; as a consequence, they would have to posit that Dinka lexical domains can be of the shape, CGVCCVC.

(4) Word-Structures of Dinka (the holists' view)

(a)	CV	CV:	C(G)V:
(b)	CVC	CV:C	C(G)VC
(c)			C(G)VCCVC

This holist analysis completely ignores the observation that the root (4c) is found exclusively in compounds; no simplex lexical domain of this shape would be tolerated as a word of Dinka.

To understand the presence of (4c) (which is to say: Dinka compounds), the holist must posit a special word structure unique to these compounds. To avoid flagrant overgeneration the holist would also have to posit a meta-rule/meta-constraint barring this word structure for simplex words. Clearly, the necessity of such ad-hoc meta-rules can be seen as signal of failure to the holist' stance (c.f. Maslova (2004) for exactly the same argument against the 'redundant' OT grammars).

An analysis of the consonant clusters in Dinka compounds produces more problems for those who claim compounds have no internal morphological structure. The holist would have to claim that Dinka allows CCs: /ŋ-dʒ/, /n-c/, and /k-d/ within their lexical domains. However, Dinka presents these (or indeed any) consonant clusters exclusively in compounds, and, exclusively at the centre of compounds.

This very special distribution is easily explained by an analysis which states that Dinka has no underlying consonant clusters at all, rather, they only

[3] These are not, as the English gloss might suggest, Adj-N phrases.

seem to be present when a /C(G)V(:)C/ lexical domain is compounded by another lexical domain which is consonant initial (5).

(5) Consonant cluster mirage

[	[	C	(G)	V	(:)	C]	[	C	V	C]	]
		n	j	a		k		d	u	r	

Holists would have to state that Dinka allows the consonant clusters found in (3a-c), and, on top of this (to forbid their presence within simplex lexical domains), also has a meta-rule stating that no consonant clusters may be found in simplex words.

Dinka casts a very significant shadow over the holists' understanding of compounds, but it is not alone; exactly the same arguments could have been made about Lhasa Tibetan (Ulfsbjorninn 2007), Thai (Denwood 1999), Cantonese and Mandarin Chinese (Goh 1997), and Vietnamese (Pham 2003). In all these languages, the canonical word-structures correspond to CV:, CVC roots. That is, small lexical domains that present no consonant clusters. All these languages, however, have compounds, these are bigger than any possible simplex word and may, unlike any simplex word, manifest consonant clusters (although exclusively at their centre). To a holist, this should present a mystery: why would the grammar encode such strict restrictions on word-size and violation of consonant clusters while taking no action to either prohibit compounds or to repair them in some way?

(6) Hypothetical repair of compound

*CGVCCVC		C(G)V(:/V)C
njakdur	→	nja:r or njaur

Other morpho-phonological facts show that compounds must be atomistic, for instance, in English we see that within a lexical domain, obstruent-obstruent and s-obstruent clusters must share a 'voicing' (7a), (7b).

(7) Obstruent Clusters in English

(a)	apt	skrɪpt	opt	aptɪtʃu:d	'apt', 'script', 'opt', 'aptitude'
	*abd			*apdɪtʃu:d	
(b)	rɪsk	bəhɛst	pəust		'risk', 'behest', 'post'
	*rɪsg				
(c)	rag + z	→	ragz		'rag', 'rags'
	kat + z	→	*katz	kats	'cat', 'cats'

In English, this rule is violated exclusively in compounds; to be specific, in the consonant clusters of compounds which straddle the boundary between its putative parts. This is the very same, and very specific, environment where (to the holist) rules that hold everywhere in a language break down.

(8) Obstruent-Obstruent Clusters in English

(a)	blakbɜ:d	*blagbɜ:d	*blakpɜ:d
(b)	hausdog	*hauzdog	*haustog

The holists' explanation, whatever it may be, to explain why these consonant clusters do not share 'voicing' will, by definition, ignore the fact that this violation occurs exactly where you might predict it to occur if compounds are atomistic. Strong morpho-phonological evidence for the atomism of compounds comes from (9 a-c).

(9) Stress and Morphology in English (Kaye 1995)

(a)	párent + al	→	paréntal	
(b)	párent + hood	→	*parénthood	párenthood
(c)	párent + trap	→	*parénttrap	párenttrap

The contrast between (9a) and (9b) is just one example of many processes (see Kaye 1995) which shows that, in some cases, when morphology is added to a stem it appears invisible to that stem (Scheer 2008, forth.).

The following data from Hampstead Street-Urchin English explain what we mean by visible. The process involved is that of l-vocalisation. In simplex words we see this process occurring finally (10b-c). Before a vowel, however, we see the clear /l/ allophone (10d). The non-words (10e) and (10f) show the alternation's productivity.

If the /l/ 'sees' the end of a word it vocalises. If the /l/ 'sees' an ensuing vowel then it surfaces with its clear alternant /l/. The compounds show something interesting (10g-i). Look at (10i) contrasted with the non-compound (10j). The /l/ in (10j) behaves normally, it sees the vowel that follows it and surfaces as /l/. The /l/ in (10i) is also followed by a vowel, but it appears not to 'see' it, as such it surfaces as it would if it was at the end of a word.

(10) Hampstead Street-Urchin English

(a) [l] → [ʊ] / __]

	Standard Register		Hampstead Street-Urchin[4]	
(b)	/batl/	[ba.tl̩]	[ba.tʊ]	‘battle’
(c)	/sɪkl/	[sɪ.kl̩]	[sɪ.kʊ]	‘sickle’
(d)	/atləs/	[at.ləs]	[aʔ.ləs]	‘atlas’
(e)	/makl/	[makl̩]	[ma.kʊ]	non-word
(f)	/makli/	[mak.li]	[ma.kli]	non-word
(g)	/mɛtl pot/		[mɛ.tʊ.poʔ]	‘mettlepot’
(h)	/katl prod/		[ka.tʊ.prod]	‘cattleprod’
(i)	/batl ɛɹiə/		[batʊ.ɛɹiə]	‘battlearea[5]’
(j)	/mɛtəlɜ:dʒi/		[mɛtəlɜ:dʒi]	‘metallurgy’

Look back at the (9a-c). The contrast between (9a) and (9b-c) both show what (10i) and (10j) show. In (9a) and (10j) all parts of the string are visible to all other (relevant) parts of it; while in (9b), (9c), and (10i) some parts of the string are not visible to other (relevant) parts of it.

Kaye (1995) formalised this by the introduction of ‘domains’, and the positing of a restriction on computational power: Kaye’s Phase Impenetrability Condition (for the terminology see Scheer 2008).

(11) Phonological phases and the PIC

(a) [párent, al]	→	[paréntal]	non-analytic
(b) [párent] , [hood]	→	[[párent][hood]]	analytic
(c) [párent] , [trap]	→	[[párent][trap]]	analytic[6]

The bracketing in (11b) and (11c) reflects the stem’s independent processing. Kaye also restricted the phonological grammar’s power by axiomatising a prohibition against any ‘look back’ operations: Kaye’s PIC.

The stress shift and l-vocalisation in compounds of Hampstead Street-Urchin English show them to be of the analytic type. Before we show a

[4] The details are similar to other London varieties of English but not in every detail, so we use this label. One significant variant seems to differ from other London varieties as the /t/ only debuccalises word-finally and in coda position, never intervocally: *[bɑʔɐ], [bɑtə] ‘butter’.

[5] One example of this would be ‘red pitch’ outside of Hampstead School, N.W. London.

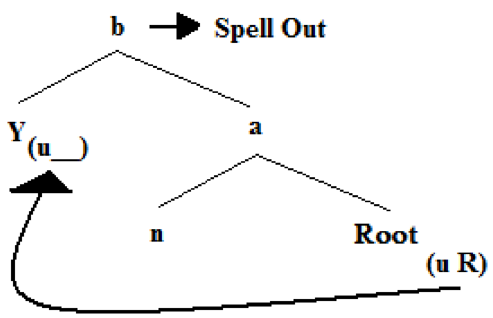
[6] Kaye (1995) also presents various ‘meter-words’ to show how they have different stress patterns depending on whether the grammar treats them as compounds or not, see the British vs. American pronunciation of ‘altimeter’ [[álti][meter]] vs. [altímeter] (see (33) and (34)).

representation of what we mean by this we must briefly discuss the fascinating advances that theoretical phonology has made in this area since Kaye (1995).

Recent research on domains and the PIC (Newell & Piggott 2006, Newell & Scheer 2007, Scheer 2008, forth.) has exposed them as the *in toto* precursors to syntactic phases and the *phase impenetrability condition* (PIC) of minimalist syntax (Chomsky 2001).

This understanding of the PIC shows it not to be a principle at all: that is, if a principle is an encoding of information in the grammar which suppresses/inhibits or activates/stimulates the grammar's operations [7] or maintains the well-formedness of structural and other relations. Rather, the PIC, as originally formulated, is nothing but an *emergent generalisation* based on the interaction between phonological computation and the phonology's own syntactic structure. Just as in Kaye (1995) and Scheer (2008; forth.), the interpretation triggering affix spells out its sister.

(12a) [parental] , non-analytic morphology

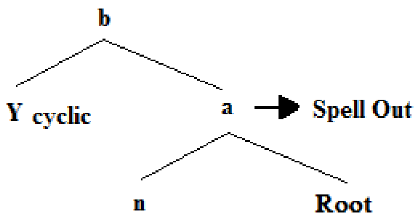


In our interpretation of non-analytic morphology (see Scheer 2008, forth. for a related view), we have a head: /n/ and a Root: Root. These merge as alpha. Alpha, merges with a non-interpreting affix (Y). Y is annotated with an unvalued feature which probes for the root. The root moves to Y to check the probe. Once checking is complete, Y's projecting node /b/ is sent to spell out.

[7] INSERT, DELETE, COPY (Samuels to appear), or AGREE, MERGE (Ulfsbjorninn 2008, forth.), and/or the licensing forces Scheer 2004).

This way, the root and any non-analytic suffixes are sent to spell out as a single string: [X Y].

(12b) [[parent][hood]]



The interpretation-triggering (hence interpreting) affix (Ycyclic), on the other hand, does not contain any unvalued features and is thus not a probe. In line with Scheer (2008), the cyclic affix spells out its sister /a/. This means that the root will be sent to spell out in isolation from its affix. The lack of lookback devices means that the root will be forever frozen from anything outside [X]. Spell out will read [X] in isolation from Y.

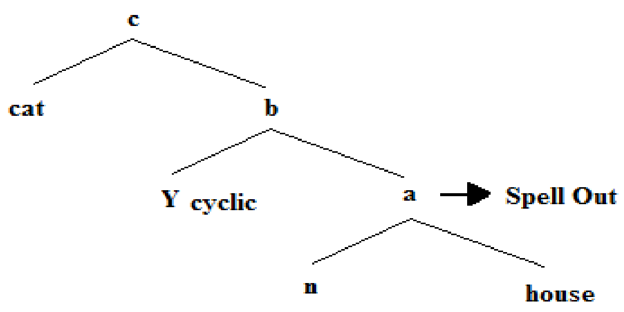
Due to this understanding of morpho-phonology, there is no ‘look back’ in /parenthood/ because the strings /parent/ and /hood/ are sent to spell out independently. Thus the PIC is a generalisation that emerges from the way the phonology’s syntax and the computation interact.

Compounds pattern with the interpreting suffixes (9a-c, 10b-i) although the compound’s two parts are major class lexical items: /house/, /cat/. Neither part of which is an interpreting affix. We posit, therefore, an interpreting (empty) affix (Y) which merges with the non-head part of the compound (/house/ in ‘housecat’). The interpreting affix has the head part of the compound as its specifier (/cat/ in ‘housecat’). The ‘head’[8] will inevitably c-commanding the ‘non-head’. This structure ensures that that the ‘non-head’ (/house/) will surface initially and with main stress (as /house/ does in ‘housecat’ and as /parent/ does in ‘parenthood’ (12b))[9]:

[8] Of the compound.

[9] This does mean a slight but interesting reanalysis from the normal understand of compounds: the ‘head’ of a compound is not actually its structural head, in fact, the compound’s structural head is a piece of empty morphology, the cyclic Y.

(13) The Compound 'housecat'



The string this creates is one where /house/ and /cat/ are sent to spell out independently as monosyllabic strings.

2. Synthesis and Predictions

The current experimental support of the atomist position (1.2.1.) is inherently limited because they lack any reference to specific theoretical (morpho-phonological) claims. Concomitantly, the theoretical camp is often criticised for a lack of experimental application[10].

This paper serves as a bridge between these two positions and is intended to show to all camps, that claims, irrespective of any detail, claiming compounds to be structurally equivalent to simplex words are profoundly misguided (c.f. Butterworth 1983, Bybee 2001, Starosta to appear). We are able to do this because the structure proposed for compounds in (13) creates strong and testable predictions, and we have had access to a patient with a specific and (seemingly) rare form of aphasia (henceforth RC).

2.1. The Patient (setting up a prediction)

RC is a patient linked to Dr Michele Miozzo's 'Sound to Sense' research program at the University of Cambridge. RC is in late middle age. He has lost the use of his right arm, and, although is right leg is also affected, he can walk aided with a stick, crucially, RC has no facial paralysis and no motor-articulatory deficit.

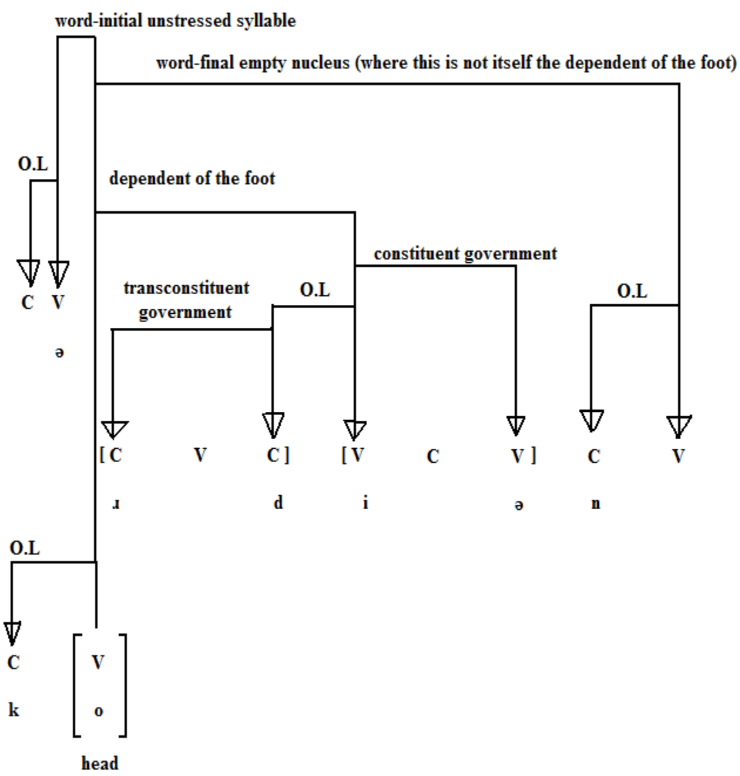
[10] This is not to concede that theoretical linguistics does not rely completely on less obvious types of experiments, such as grammaticality judgments (Hyman p.c.) or the use of morphemes to alter the phonological environments.

Preliminary tests had already established that the length and syllabic complexity of a word were significant and triggered errors.

In naming and repetition tasks, the patient had a high error rate in words of three or more syllables (30/96, 31.3%)[11]. RC also showed comparatively negligible error rates in words of two syllables (or less) (23/399, 5.8%). The difference between error rates is extremely significant (z test: $p < .000$).

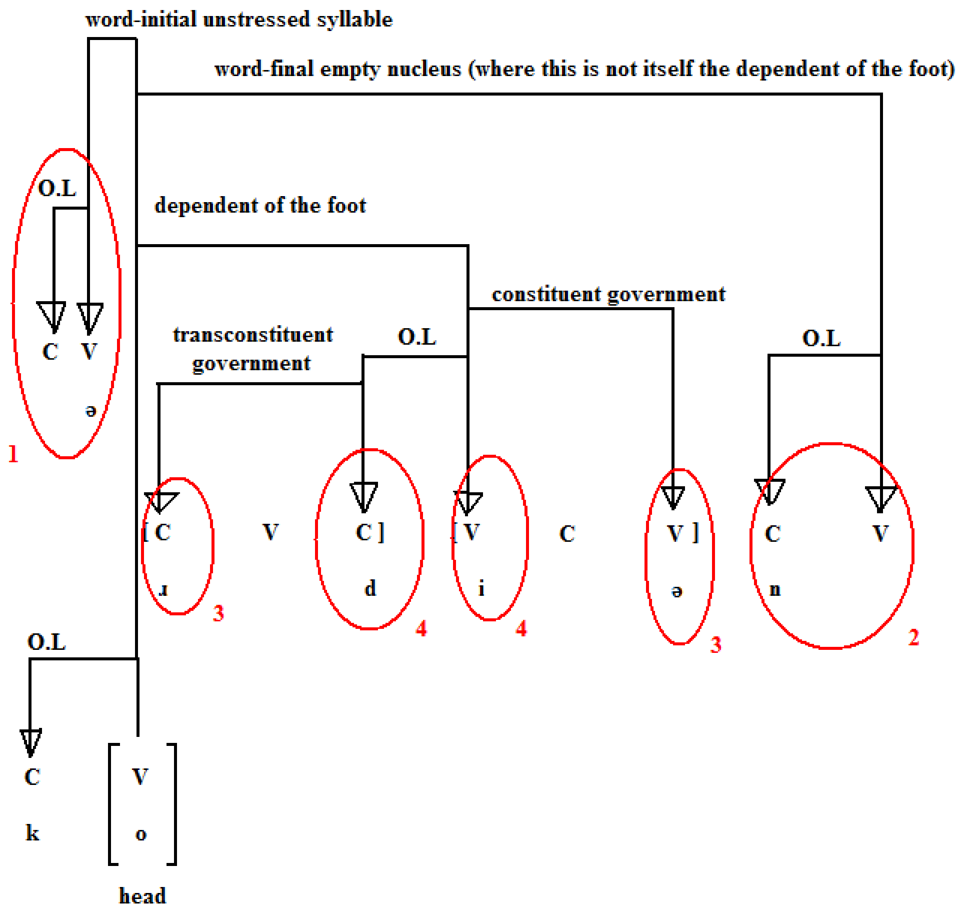
The nature of the errors was interesting also. Typically, they constituted a simplification of parts of syllabic structure familiar to other psycho-phonological disorders (Jakobson 1941, van der Lely 2005, Gallon et al. 2008). Interestingly, these problematic positions are coextensive with what Harris' (1997) A-licensing would define as weak (hence marked) positions. In (14) we see the licensing forces related to A-licensing; these were independently developed by researchers, for an overview see Charette (1991), Harris (1994), or Scheer (2004).

(14) Licensing in (a rhotic) 'accordion' /əkoɹdiən/



[11] 3 syllable long words will hence he termed: long and disyllabic words will be termed: short.

(14) (a) Commonly affected syllabic positions in SLI (van der Lely 2005)[12]



Because RC's errors typically revolved around marked syllabic structures, we had to establish that it was not the case that the long words were triggering errors because they were, relative to their length, more syllabically complex. We established a syllabic complexity score by dividing the number of the consonant clusters and vowel clusters per word by the number of its phonemes. We found that, in our sample, long words (relative to their length) were not significantly more complex than the short words (z test: $p < .89$).

[12] The numerical values symbolise the more marked and most typically damaged in phonological pathology. It is not scientifically defined but would, most probably, be met with consensus. The smaller the number, the more fragile. The affected areas are in keeping with van der Lely (2005) and Gallon et al. (2007): initial unstressed syllables >> word-final consonants >> vowel and consonant clusters >> dependents of the prosodic head.

With the issue of complexity established, RC provides the perfect experimental testing ground for our theoretical structure of compounds. RC shows high error rates in long words where he simplifies consonant clusters. Conversely, RC correctly produces disyllabic words, even when they contain[13] consonant clusters.

2.2. The Prediction

Given what we know, we can predict exactly how RC will produce trisyllabic compounds should the holists or atomists be correct.

The holist must maintain that, as there is no statistically significant difference between length and phonological complexity of long simplex and compound words ([σσσ]). These must pattern together; both triggering significantly higher error rates than either the short simplex or short compound words.

(15) Holist Predictions (of syllabically complex words)

No Errors		Errors	
Short simplex word	/picture/	*Long simplex word*	/stethoscope/
Short compound	/light bulb/	*Long compound*	/picture frame/

Atomists supporting (12-13), would maintain that trisyllabic compounds will be processed in one of two ways: [[σσ] [σ]], [[σ][σσ]]. The phonological implication is a prohibition on 'look back', therefore, when processing trisyllabic compounds, the phonology will never have to process a trisyllabic string. As such, they will be unaffected by RC's deficit. The atomist's startling prediction, therefore, is that phonologically complex, trisyllabic, compound words (/picture frame/) will pattern with the short words (/picture/).

(16) Atomist Predictions

No Errors		Errors	
Short simplex word	/picture/	*Long simplex word*	/stethoscope/
Short compound	/light bulb/		
Long compound	/picture frame/		

[13] …what in long words are problematic…

3. Experiment Design

3.1. Method

We ran a picture naming task, an experiment used extensively to test production of words by aphasic patients (Snodgrass & Vanderwart 1980, Ferrand et al. 1994). The patient was presented with a picture (printed onto a sheet of A4) and asked to produce the corresponding, target noun.

Repeated productions: 'apple tree… yes… apple tree' were both recorded as productions. If the patient was unable to guess the target noun, short semantic clues were provided. Productions facilitated by occasional phonological priming (giving the initial consonant and a schwa) of the target noun were discounted. The productions were recorded as sound files, and transcribed once in IPA (during the experiment) and again, after the experiment, from the tapes (transcribed using the Sound to Sense laboratory transcription key). On average, there would two breaks per naming session.

3.2. Materials

The pictures used to elicit the target nouns where a set of long and short simplex words, long and short compound words, and monosyllabic fillers. The long and short simplex words, and long and short compounds were matched for complexity. The compounds are significantly more phonologically complex than the simplex words (reinforcing later conclusions). The pictures were taken partly from the Snodgrass & Vanderwalt's (1980) standardised set and partly by public domain images taken from Google searches. All the pictures where printed in black and white and, in total, the target nouns numbered 207. Recording was on a SONY Digital Voice Recorder, as regularly used by the Sound to Sense laboratory.

3.3. Error Analysis

The target words were split into three categories: short, simplex, and compound (the latter two categories both minimally 3 syllables and matched for length and syllabic complexity). We recorded all the errors in all words, however, the target for error was much higher in long words than in short words (more phonemes to trigger errors); therefore, we separately present the errors the patient made with CCs (the percentage of which was roughly equal in all word sets) and only drew conclusions from these CC errors (although all data is supportive of one hypothesis).

3.4. Remaining

The results are a collation of data recorded by the author, combined with a smaller sample of data collected by Dr Miozzo, what we term, the pilot experiment. These were incorporated into the study and, as they have not been previously unpublished. I again thank Dr Miozzo for allowing me to use his data. The analysis of all data is the author's own.

4. Results and Analysis

The following are all the errors (of any type) from all the patient's productions.

(17)

Pilot Experiment	Errors	Words in Sample	Error %
Short Words	1	60	1.6
Long Compounds	3	51	5.8
Long Simplex Words	30	96	31.3

(18)

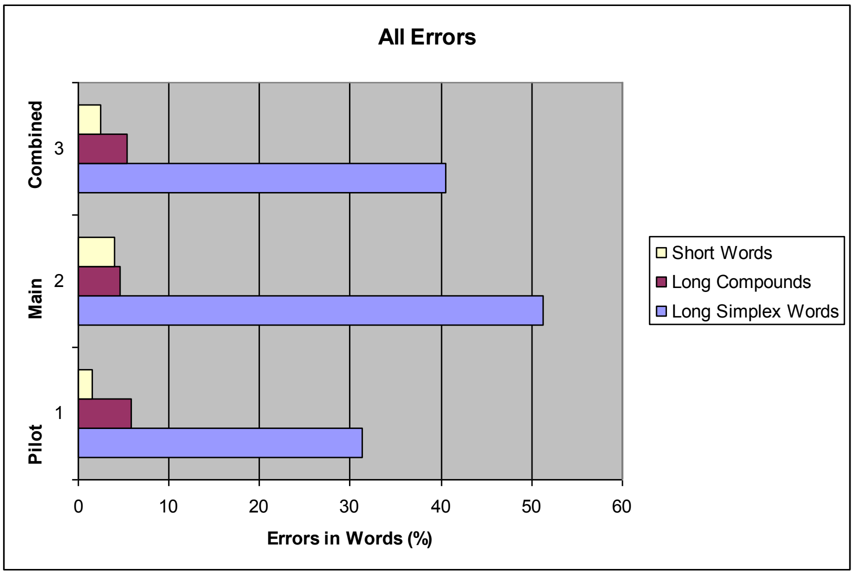
Main Experiment	Errors	Words in Sample	Error %
Short Words	1	25	**4**
Long Compounds	2	43	**4.6**
Long Simplex Words	**43**	**84**	**51.2**

(19)

Combined Errors	Errors	Words in Sample	Error %
Short Words	2	85	**2.4**
Long Compounds	5	94	**5.3**
Long Simplex Words	**73**	**180**	**40.5**

We plotted the percentage of error columns against the word-type to produce the graph in (20).

(20) Errors (%), Pilot, Main and Combined



Although we will only take the data from errors in CCs as conclusive, we point out here that long simplex words were far more likely to trigger errors than the long compounds of equivalent length and phonological complexity. The results are extremely statistically significant (z test: $p < .000$). There was also a difference between the compounds and the short words, this however, was not statistically significant (z test: $p < .63$). The ensuing tables are the errors with consonant clusters.

(21)

Pilot Experiment	CC Errors	Tot CCs in Sample	Error %
Short Words	2	66	**3**
Long Compounds	1	91	**1.09**
Long Simplex Words	**27**	**112**	**22.1**

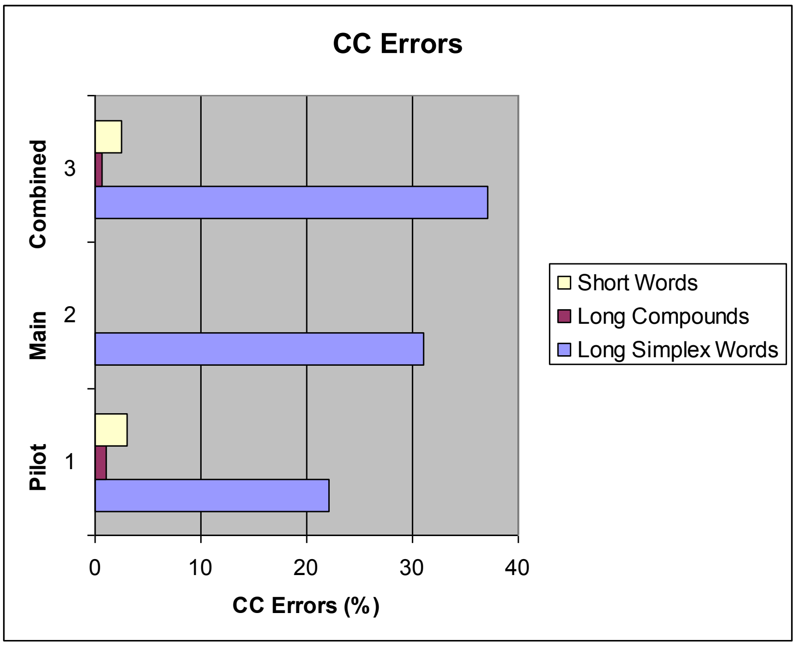
(22)

Main Experiment	CC Errors	Tot CCs in Sample	Error %
Short Words	0	13	**0**
Long Compounds	0	52	**0**
Long Simplex Words	**19**	**62**	**31**

(23)

Combined Errors	CC Errors	Tot CCs in Sample	Error %
Short Words	2	79	**2.5**
Long Compounds	1	143	**0.69**
Long Simplex Words	**46**	**124**	**37.1**

(24) Consonant Cluster Errors, pilot, main, and combined.



Taking the combined results, the long simplex words were far more likely to contain CC errors than the compounds of equivalent length and phonological complexity.

The results turn out to be extremely statistically significant (z test: p < .000). There was also a difference in the combined study between the long compounds and the short words. This difference, however, was not statistically significant (z test: p < .7).

Surface based, holist, accounts which ignore internal, morpho-phonological structure would have predicted the long compounds to pattern with the long simplex words, which are (from a surface perspective) are equivalent in length and phonological complexity. We see the reverse, however. The data overwhelmingly show that long compounds pattern with the short words. The data is entirely consistent with the atomistic, representation (13). Supportive evidence comes from an investigation of the representative error types (4.1.); these help reveal exactly how the representation (13) is the etiology behind the data.

4.1. Representative error types

The most representative error types are also of interest in that they suggest the very phonological reason why simplex long words are damaged (while simplex short words are not). It also explains why, due to representation (13), long compounds are unaffected. It is important to note that RC performed all sorts of errors, substitution of syllables and phonemes, some cases of intervocalic voicing, and even introduced new consonant clusters and unstressed initial syllables to some words. However, the vast majority of his errors were the following familiar types.

(25) CC reduction

flamingo	LSW	fəmɪŋgəu
propeller	LSW	pəpɛlə

(26) Final C reduction

Abacus	LSW	abəkə
Unicorn	LSW	ju:nikɔ

(27) Initial unstressed σ-deletion

Mechanic	LSW	kanek
Decanter	LSW	kantə

(28) VV reduction

accordion	LSW	əkɔ:dɪnɪn
sombrero	LSW	sombrɛlə

(29) Reduction to a trochaic foot

clarinet	LSW	kaɹi, karin, kanə
limousine	LSW	zim, zimə

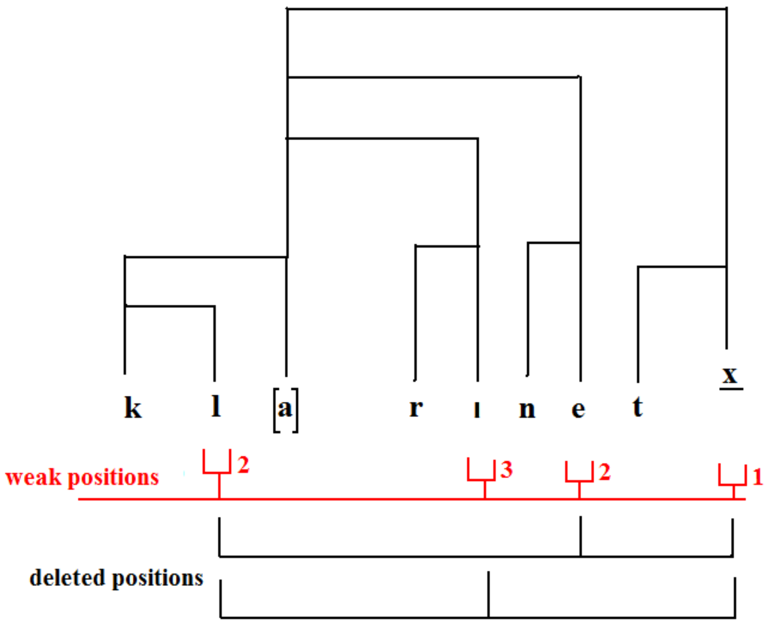
5. Discussion

As briefly alluded to in (2.1.), RC's errors are very similar to those found in grammatical SLI (van der Lely 2005, Gallon et al. 2007) and mirror many common syllabic structure-related error types in child language acquisition (Smith 1973, Bernhardt & Stemberger 1998, Kager et al. 2004). This link between pathology and acquisition is reported as early as Jakobson (1941) who comments that the structures which are damaged in phonological pathology are often the latest to be acquired by typically developing children. The reason for this, from a phonological perspective, is to understand syllabic markedness and positions of strength and weakness as universally drawn from what Harris (1997) calls A-licensing.

In A-licensing, the nucleus which is the head of the domain, although not of the phase (c.f. Scheer 2008), which bears the main word-stress, licenses all other positions in that domain. As such, the more distant the licensing is from the core licensing (the core main word stress CV) the weaker, or more marked, the syllabic position will be.

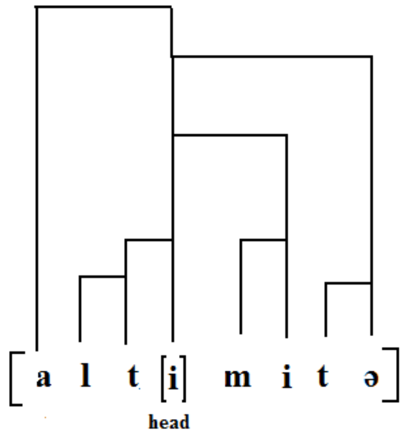
Harris conceived of A-licensing to account for neutralisation phenomena, although its effects in phonological pathology are clear. Take the word /klarInEt/ 'clarinet'. If you were to select the weak positions as defined by A-licensing, and deleted some or all of the weak and marked syllabic positions, you would generate precisely the types of the errors produced by RC, patients with grammatical SLI (van der Lely 2005, Gallon et al. 2007), and those products of typically developing L1 phonological acquisition.

(30) A-licensing in 'clarinet' (see (29)), weak positions and deleted positions

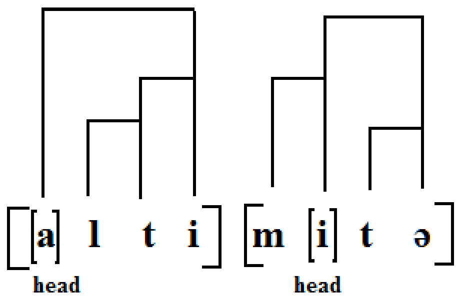


The weakness (and markedness) of a syllabic position is defined therefore, by its relation to a head within a domain. We can compare trisyllabic simplex words and trisyllabic compounds as they would appear in their independently processed strings, as defined by the representation in (13).

(33) Long Simplex Word, US pronunciation of 'altimeter' (Kaye 1995)



(34) Compound, UK pronunciation of 'altimeter' (Kaye 1995)



The structure of compounds given in (13) means that the resultant strings are small and the positions within those domains are therefore more strongly licensed than the positions in the long simplex word.

Our phase based account of compounds accounts not only for why they pattern with short words, but also incorporates an explanation as to the common, predominant error types.

Interestingly, the very few errors found in the compounds themselves revealing (although not in any way conclusive).

(35) Errors in Compounds

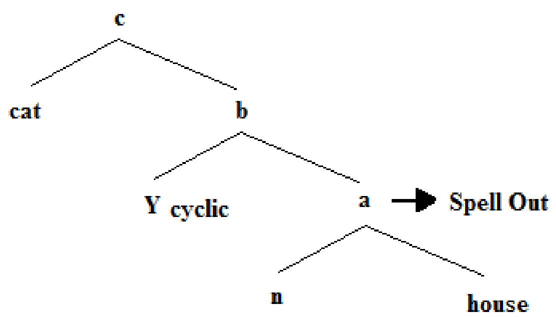
ironing board	"	onibo:d
typewriter	"	write - typer

Look at the error in /ironing board/; it is quite possible that, although /ironing board/ was included as a trisyllabic compound, the patient interpreted its first part as: /ɑijənɪŋ/. If that was the case, this compound would have been comprised of a long simplex portion and a short portion, which could explain its apparent cluster reduction. An examination of other such compounds such as /newspaper boy/ would reveal whether this is the case.

6. Conclusion

Having introduced the long-running debate between the holists and atomist regarding the internal status of compounds, we showed that previous experimental studies have cast doubt on the holist's thesis that compounds do not contain any internal morpho-phonological structure. We also showed that arguments from theoretical morpho-phonology seem to conclude with an atomist position, that compounds, unlike simplex words, have an articulated internal morpho-phonological structure. Various facts about compounds in Dinka and English (supported by Thai, Cantonese and Mandarin Chinese and Vietnamese), pointed to a specific internal representation of the compound.

(36) Compound



No experimental evidence had yet been carried in support of this structure and so to settle the debate between holists and atomists, we ran an experiment.

RC is a patient who produced errors with words longer of three (or more) syllables, and produced virtually no errors in words of two syllables (or less). Given RC's error distribution, the holists and atomists would make exactly contrasting predictions as to whether trisyllabic compounds would be produced with errors (like the trisyllabic simplex words) or without (like the smaller words).

The holists would predict the former. And the atomists, if (13) is correct would predict the latter.

It turns out that the latter is overwhelmingly supported by the findings of the experiment. The experiment's findings were completely consistent with (13) and a discussion of the phase based account of compounds and A-licensing, not only predicted the correct error distribution in the patient, but also, explained the predominant syllable structure related error types.

Bibliography

Bernhardt, Barbara & Joseph Stemberger. 1998. *Handbook of Phonological Development from the Perspective of a Constraint-Based Non-Linear Phonology*. San Diego: Academic Press.

Bhattacharja, Shishir. On finding morphology without dividing words: A review of explorations in seamless morphology, *South Asian Language Review* XVI (2), 45-77.

Butterworth, Brian. 1983. Lexical Representation. In *Language Production, Volume 2: Development, Writing and Other Language Processes*, 257-294. London: Academic Press.

Bybee, Joan. 2001. *Phonology and Language in Use*. Cambridge: Cambridge University Press.

Charette, M. 1991. *Conditions on Phonological Government*. Cambridge: Cambridge University Press.

Chomsky, N. 2001. Derivation by Phase. In Michael Kenstowicz (ed.), *Ken Hale: a life in language*. Mass.:MIT Press.

Denwood, Anne. 1999. A Government Phonology analysis to Thai tones: observation and hypothesis, *SOAS Working Papers in Linguistics and Phonetics* 9, 173-190.

Ferrand, Ludovic., Grainger, Jonathan & Juan Segui. 1994. A study of masked form priming in picture and word naming, *Memory and Cognition* 22 (4), 431-441.

Fiorentino, Robert & David Poeppel. 2007. Compound words and the structure of the lexicon, *Language and Cognitive Processes* 22 (7), 953-1000.

Ford, Alan., Rajendra Singh & Gita Motohardjono. 1997. *Pace Panini: towards a word-based theory of morphology*. New York: Peter Lang.

Gallon, Nancy., John Harris & Heather van der Lely. 2007. Non-word repetition: an investigation of phonological complexity in children with grammatical SLI, *Clinical Linguistics and Phonetics* 21, 435-455.

Goh, Yeng-Seng. 1997. *The Segmental Phonology of Beijing Mandarin*. Taipei: Tiger Crane Publishing Co.

Harris, John. 1994. *English Sound Structure*. Oxford: Blackwell.

Harris, John. 1997. Licensing inheritance: an integrated theory of neutralisation, *Phonology* 14, 315-370.

Jakobson, Roman. 1941. *Kindersprache, Aphasie und Allgemeine Lautgesetze*. Frankfurt.

Kager, Rene., Pater, Joe & Wim Zonneveld. 2004. *Constraints in Phonological Acquisition*. Cambridge: Cambridge University Press.

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

van der Lely, Heather. 2005. Domain-specific cognitive systems: insights from grammatical specific language impairment, *Trends in Cognitive Sciences* 9 (2), 53-59.

Malou, Job. 1988. *Dinka Vowel System*. Arlington: Summer Institute of Linguistics.

Marslen-Wilson, William., Tyler, Lorraine., Waksler, Rachel., & Lynn Older. 1994. Morphology and meaning in the English mental lexicon, *Psychological Review*, 101 (1), 3-33.

Maslova, E. 2004. Stochastic OT: a model of constraint interaction.

<http://roa.rutgers.edu/files/694-1104/694-MASLOVA-0-0.PDF> [30/11/09]

Newell, Heather & Glyne Piggott. 2006. Syllabification, stress and derivation by phase in Ojibwa, *McGill Working Papers in Linguistics* 20 (2), 39-64.

Newell, Heather & Tobias Scheer. 2007. Procedural First. Paper presented at the *38th Poznan Linguistics Meeting*. University of Poznan. Poland.

Pham, Hoa. 2003. *Vietnamese Tone: a new analysis*. New York: Routledge.

Rastle, Kathleen., Davis, Matthew & Boris, New. 2004. My broth in my brother's brothel: morpho-orthographic segmentation in visual word recognition, *Psychonomic Bulletin & Review* 11 (6), 1090-1098.

Samuels, Bridget. To appear. *A Minimalist Program for Phonology*. Oxford University Press.

Scheer, Tobias. 2004. *A Lateral Theory of Phonology: What is CVCV and why it should be*. Berlin: Mouton de Gruyter.

Scheer, Tobias. 2008. Intermodular argumentation: morpheme-specific phonologies are out of business in a phase-based architecture. In Nomi Shir & Lisa Rochman (eds.), *The Sound Patterns of Syntax*. Oxford: Oxford University Press.

Scheer, Tobias. Forth. *A Lateral Theory of Phonology, Volume 2: on locality, morphology and phonology in phonology*. Berlin: Mouton de Gruyter.

Smith, Neil. 1973. *The Acquisition of Phonology: a case study*. Cambridge: Cambridge University Press.

Snodgrass, J. & M. Vanderwart. 1980. A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity, *Journal of Experimental Psychology: Human Learning & Memory* 6, 174-215.

Starosta, Stanley. To appear. Do compounds have internal structure? A seamless analysis. In Rajendra Singh and Stanley Starosta (eds.), *Explorations in Seamless Morphology*. New Delhi, London and Thousand Oaks: Sage Publications.

Taft, Marcus & Kenneth Forster. 1975. Lexical storage and retrieval of prefixed words, *Journal of Verbal Learning and Verbal Behaviour* 14, 638-647.

Taft, Marcus. 1979. Lexical access via an orthographic code: the basic orthographic syllabic structure (BOSS), *Journal of Learning and Verbal Behaviour* 18, 21-39.

Ulfsbjorninn, Shanti. 2007. Diachronic Evidence of Word-Final Empty Structure and its Effects: Raddoppiamento Sintattico in Maremmano Italian and Compensatory Lengthening in Lhasa Tibetan. Ms. BA Thesis. SOAS – University of London.< http://ling.auf.net/lingBuzz/000436> [30/11/09]

Ulfsbjorninn, Shanti. 2008. MOVE as AGREE and I-MERGE: CV metathesis and the primitive operations of phonology. Talk given at the *London Phonology Seminar*. UCL/SOAS: University of London. [Under review for *Lingua*]

Ulfsbjorninn, Shanti. In Prep. *The Outer Limits of Phonological Complexity*. PhD Thesis. SOAS – University of London.

Shanti Ulfsbjorninn
su1@soas.ac.uk
SOAS – University of London