Journal of West African Languages

Volume 45.1 (2018)

SEGMENT DELETION AS A MORPHO-PHONOLOGICAL PROCESS IN DAGBANI COMPOUND WORD FORMATION

Abukari Kwame
Department of Language and Culture
The Arctic University of Norway, Tromso
abukarikwames@yahoo.com

Abstract

This article discusses segment deletion as a morphophonological process in Dagbani compound word formation from a quantitative linguistics perspective. Three independent variables: the nature of constituent contact margin (C#V, V#C, N#C, V#V), compound type (NN, NA, NV, VN) and word length were considered to determine which of them is more likely to predict segment deletion (C, CV, V, Ø) in compound formation. Classification and regression trees (CART) and Random forest analysis were performed on 265 sampled compound words. The analysis revealed that the nature of constituent contact margin was more likely to predict segment deletion in Dagbani compound formation than compound type and word length. Furthermore, the commonly deleted segments in compound formation were found to be vowels and CVs. The study concludes that these processes are due to the combined interaction of Dagbani phonology and morphology to ensure that the new words formed adhere to the phonotactics and morpheme structure constraints of the language.

1. Introduction

Dagbani is a Gur language which belongs to the Oti-Volta subfamily of languages spoken in Northern Ghana. Other languages in this subfamily which Bodomo (1993) referred to as 'Mabia languages', include Mampruli, Dagaare, and Kusaal. This lable (ma 'mother' and bia 'child') is based on the sibling relationship between these languages. Dagbani has three dialects: the Eastern dialect (Nayahili) spoken around Yendi and its surrounding, the Western dialect (Tomosili), spoken around Tamale and its environs and the Nanuni dialect spoken in Nanuŋ (Hudu 2010). These dialect differences are based on tone variation and a few lexical differences. Approximately, Dagbani has about two million native speakers in Ghana.

Compounding is a productive word formation process in Dagbani, as in many other languages across the world. As a word formation process, compounding is visible within the morphology-phonology interface on one hand and the morphology-lexicon on the other. Compounding offers us very useful insights into how lexical development and linkages occur among language users. Booij (2007, 2009:3) asserts that compounding is a construction at the morphology level which has a systematic pairing of form and meaning and can be recursive. He further observes that some compounding process involves both compounding and derivation at the same time, thus suggesting that compound formation also involves syntax-morphology interaction (Booij 2009:14-15, 2010, 2012). In another observation, Fabb (1998) argues that compounding is subject to morphological and phonological processes which are specific to compounds or shared with other structures in a language. Again, Fabb observes that the meaning of a compound is to some extent compositional. The author, while citing Mohanan (1982), maintains that in Malayalam compounding, there is a phonological process of final nasal deletion and stem final vowel lenghtening, while other segment deletions are also attested in American sign language. From these various observations, it appears that compounding as a word formation process, engages with several different subgrammars of a language such as the lexicon, phonology, morphology, and syntax. It also shows that segment deletion in compounding is a productive morphophonological process in natural languages.

Therefore, the interesting questions to ask are: which segment gets deleted and what influences that deletion in the compound word formation process in Dagbani? Can segment deletion in Dagbani compounding be predicted by the nature of the constituent contact margins (C#C, C#V, V#C, and V#V), the word length, or the type

of compound formed (NN, NA, NV, and VN)? The purpose of this paper is to examine segment deletion as a morpho-phonological process in Dagbani compound word formation. The paper takes a quantitative approach to explore which segments get deleted during compound word formation and what predicts that. Even though the default morpho-phonological constraints of Dagbani can explain segment deletion in compounding, this current study will complement that through a quantitative approach as a novelty in this endeavour. Thus, I am interested in examining quantitatively which of the default factors can predict segment deletion in compounding. The following are my predictions:

- If segment deletion correlates with the nature of the constituent contact margin, then it is possible that Dagbani phonotactics and syllable structure types are what condition the deletion process.

- On the other hand, if segment deletion correlates with word length or the compound type, then Dagbani morphology and word structure might be what condition that.

Based on the above, section 2 of this paper will review previous studies on compounding in Dagbani to highlight some basic facts about nature of Dagbani compound words. In that section, I will note some morpho-phonological processes in compounding which will be discussed in section 3. In section 3, I will then present the segment inventory of Dagbani. In the same section, phonotactics and syllable structure types will be discussed. This background information about the language will help us understand why some segments get deleted in some environments but remain in others. Section 4 will cover the methodology used in this paper, where I will discuss the sources of data used for my study and the statistical methods used to analyse the data. Then in section 5, I will present the results of the data analysis in the form of tables, chats, and plots which will then be discussed in section 6. Thus, in section 6, I will try to relate the results to the Dagbani phonotactics and morphological processes and to other morpho-phonological processes in natural languages which may not be attested in Dagbani but are crutial for the purpose of this paper. Finally, I will conclude the paper in section 7.

Before I proceed to the next section, I will like to operationalize a few terms. The term compounding as used in this paper means the process of putting two words (either with/without derivational/inflectional morphemes) together, with at least one word being independent, to create a new word. This definition is based on the fact that in Dagbani, both nouns and adjectives have number suffixes as part of the lexical stem. One of these suffixes is retained in the compound word formation process.[1] Constituent contact margins as used in this paper will be restricted to the final segment(s)of the first word and and the initial segment of the second word, in a compound. Since this is the merger point in compound formation, it functions as the zone for segment deletion. Technically, therefore, the constituent contact margin of the form [C#C] is actually [N#C], as pointed as by an anonymous reviewer, where [N] stands for a nasal, which can either be /m/ or /ŋ/ based on syllable structure constraints of the language. Therefore, the constituent contact margin [N#C] will used throughout the analysis in this paper to reflect the Dagbani syllable structure and

[1] An anonymous reviewer asked if it is possible for two roots, one of which has a suffix, to be put together to form a compound word. Yes, this observation is correct. As far as I know, lexical items in Dagbani exist in their root forms only when their suffixes are taken away either for phonological or morphological reasons. So, in compound formation, one of the roots must have its suffix, which is always the second constituent, as I will demonstrate later in this paper. Also, Olawsky (2002:213) notes that compounds in Dagbani are at least trisyllabic based on the fact that they consist minimally of two roots and a suffix. Thus, the second word will always cosist of a root plus a suffix, as observed also by (Hudu, 2005:10).

phonotactics. Compound type means the lexical classes of the constituent words in a compound (eg. NN for a noun-noun combination).

Olawsky (1999:156) observes that world length, as a prosodic category, plays an important role in Dagbani phonology and morphology. He shows that the length of a prosodic word determines how certain noun suffix allomorphs are realized. Further more, he argues that the occurrence of homorganic nasal assimilation is also affected by word length in the language and that words longer than four syllables are rearly attested in Dagbani. This means that segment deletion in compound formation could possibly be predicted by the word length of the resultant compound formed. For instance, Olawsky (1999:157) observes that the singular form for class 3b nouns is [**ŋa**] due to a simplification of **[n + ga**], as in (1a) while the singular form for class 3c is [**ga**], as in (1b). He proposes that since there is no segmental environment that prevents [**ŋga**] in the singular form from [**ŋa**], the only reason for this difference lies in the prosodic structure of the nouns, where class 3c nouns have more than two syllables while most class 3b nouns have only two syllables.

(1) Number suffixes in class 3 nouns (Olawsky 1999:157)

a. Class 3b nouns

[baaŋa, baansi]	**/ba**:**ŋa, ba**:**nsi/**	(singer-SG/PL)
[guŋa, gunsi]	**/gʊŋa, gʊnsi/**	(kapok tree-SG/PL)
[nooŋa, noonsi]	**/no**:**ŋa, no**:**nsi/**	(bird-SG/PL)

b. Class 3c nouns

[gariŋga, garinsi]	**/garəŋga, garənsi/**	(viper-SG/PL)
[foliŋga, folinsi]	**/fɔləŋga, fɔlənsi/**	(span, gap-SG/PL)
[tuliŋga, tulinsi]	**/tuləŋga, tulənsi/**	(type of fish-SG/PL)

Accordingly, wordlength is considered as a variable here to examine how far it can predict segment deletion in Dagbani compound word formation process. Therefore, word length will be interpreted as the number of syllables in a compound before segment deletion occurs. This will help determine if wordlength can predict segment deletion. Finally, deleted segments will be V, for vowel, C, for consonant, CV, for a syllable and Ø, for empty (if nothing gets deleted).

Theoretically, the anlysis done in this paper will be influenced by the phonotactics and the syllable structure constarints of Dagbani. Also, I will use the concept of construction grammar (CG) in the discussion in section to analyse the internal structure of some compound words to provide additional evidence to support the claim that NA and NV compounds are said to belong to NN compound type in the language.

2. Previous studies on Dagbani compound words

Compound formation is a very productive morpho-phonological process in Dagbani (Hudu 2014a; Abdul-Rahman 2013; Olawsky 1999, 2002). Olawsky (1999, 2002) claims that most Dagbani compounds are endocentric with the rightmost lexical element being the morphological head of the compound. He observes that in plural forms, the plural morpheme is always realised on the head, and in other cases, the head may also contain a derivational suffix to transform a verb into a noun. Despite these observations about compound words in Dagbani, there has not been any comprehensive research on compounding in Dagbani. The only few studies in Dagbani that have something to do with compounding are reviewed below.

In a study on elision in Dagbani, Abdul-Rahman (2013) analyzes segmental deletion as a prominent syllable structure process in Dagbani from an autosegmental phonology perspective. He argues that segment elision in Dagbani, especially in noun-noun and noun-adjective compounding as well as in plural formation, involves

several morpho-phonological phenomena. This means that segment deletion is a productive process in Dagbani compound word formation, as in (2). Generally, the author maintains that the segments that commonly delete in these processes (compounding and plural formation) are vowels, consonants and sometimes entire syllables. The data in (2) is taken from Abdul-Rahman (2013:222) with some modifications in the spelling.

(2) Segment deletion in compounding[2]

	A	B	C (A + B)	*D
a.	**lu.ŋa (drum)**	**do.li (stick)**	**lun.do.li (drum stick)**	**luŋadoli**
b.	**sa.a (rain)**	**gba.ni (cloud)**	**sa.gba.ni (rain cloud)**	**saagbani**
c.	**noo.ŋa (bird) nooŋatiɛʔʊ**	**tʲɛ.ʔʊ (nest)**	**noo.tʲɛʔʊ**	**(bird nest)**

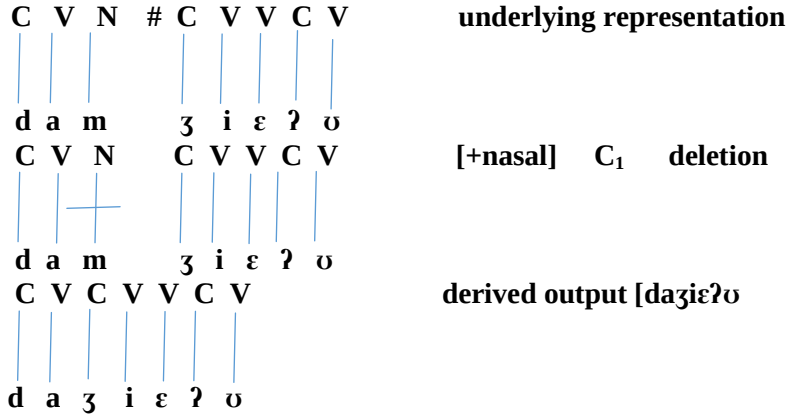
Abdul-Rahman proposes that the deletion of a vowel in the first word before it combines with the second word during compounding is influenced by the phonology of the language to allow for the right syllable structure to surface. Thus, he notes that the compound [**lun.do.li**] 'drum stick' now has a CVN.CV.CV structure, where the first syllable CVN functions as the stem of the CV.CV word [**lu.ŋa**]. Even though this analysis is right, I argue that the deletion of the number suffix [a] from [**lu.ŋa**] is not due to a phonological reason, but for a morphological reason, as I will illustrate further in section 3.3. Additionally, other processes noted in the data are the homorganic nasal assimilation and palatalization processes, which will also be discussed in section 3.3.

Furthermore, in analysing the formation of the compound word [**daʒiɛʔʊ**] as in (3), the author supposes that the deletion of /m/ in the first word is motivated by the phonology. Abdul-Rahman proposes that since the bilabial nasal is a tone bearing unit, it can function as a syllable on its own, which is prohibited in the position it occupies in the compound.

(3) Nasal elision in Dagbani compounds (Abdul-Rahman 2013:226)

a. **dam 'alcohol' + ʒiɛʔʊ 'reddish'** **daʒiɛʔʊ** 'pito – a local alcoholic drink'

b. matrical representation of the compound



[2] Note that in some examples, I enclose the Dagbani words in square brackets while in others I do not. The reason for this is to allow cited data to be as close to its source as possible.

Again, this analysis seems problematic given that /n, ŋ/ are also nasals which are allowed as root codas in [**lun.do.li**] 'drum stick' and in [**na.pɔŋ.tam**] 'footprint', where [**napɔŋ**] is 'foot' and [**tam**] is 'soil'. Unless the nasal /m/ has a different status (as a tone bearing unit, or a mora) than /n/ and /ŋ/, which I suspect is not so. Other compounds in which a nasal is part of the first word are [**ɡbampiɛlli**] 'skin white' (Western person) and [**bimpiɛlli**] 'thing white' (white calico) (Olawsky 1999:118). Note that there is homorganic nasal assimilation (henceforth HNA) in both [**ɡbampiɛlli]** and [**bimpiɛlli**]. More discussion on HNA in this section and in section 3.3. The plausible explanation, as Olawsky (1999:102-103) proposes, is that /-m/ can be a nominal suffix (perhaps for singular number marking), as can be observed in (4), where the nasal is separated from the root. Based on Olawsky proposal, then /m/ can be deleted as a number suffix so that the second word bears the number suffix as a morphological constraint in NA compounds.

(4) /-m/ as a nominal suffix (Olawsky 1999:102-103) with my own additions

a. **[kom, ko.tul.li] /ko-m, ko.tul.li/** 'water, water.hot.SG'
b. **[ʒim, ʒi.tul.li] /ʒi-m, ʒi.tul.li/** 'blood, blood.hot.SG'
c. **[kpam, kpa.biɛ.ʔʊ] /kpa-m, kpa.biɛ.ʔʊ/ '**oil, oil.dirty.SG'

Another study that does some analysis on Dagbani compound words is Hudu (2014a), who examines a phonological wordhood in Dagbani. In that study, Hudu (2014a:8) notes that compounds can differ from other complex constructions in that some compounds can have compositional meaning, where the meaning of the compound is based on the meanings of the constituent words while others could have semantically non-compositional meaning, where the meaning of the compound is not based on the meanings of the constituent words. Furthermore, he explains that the final vowel of the first noun in a compound can be reduced to [ɨ] or deleted all together due to its position in the compound, as in (5). However, he concludes that the deletion of this suffix is not phonologically motivated. For example, Hudu (2014a) proposes that the word [**nah-u zʊʔ-ʊ**] 'cow's head' can be pronounced as [**nah zʊʔʊ**] while [**baʔ-a kɔl-gʊ**] 'soothsayer's sack' can also be pronounced as [**baʔ kɔl-gʊ**] in causal speech, but such a reduction is not phonological, since it does not affect the phonology in any way. He maintains that with the exception of the vowel elision or reduction, the first root of the compound remains phonologically distinct from that of a complex word in all other ways.

(5) Segment reduction/deletion in Dagbani compound (Hudu 2014:8)

a. **nahʊ + zʊʔʊ = nahzʊʔʊ** 'cow's head'
b. **baʔa + kɔl.gʊ = baʔkɔlgʊ** 'soothsayer's sack'

Additionally, Hudu (2014, p. 16) observes that in Dagbani, homorganic nasal assimilation is a productive phonological process in clitics, affixes and in bound roots as in (6). The data in (6) and (7) are taken from Hudu (2014a:12) with some modifications.

(6) Nasal assimilation on nasal cardinal prefixes

[n-dam] /n-dam/ 'one'
[n-yi] /ɲ-ji/ 'two'
[n-woi] /ŋ-wɔi/ 'nine'
[n-nu] /n-nu/ 'five'

(7) Nasal assimilation in reduplicant prefixes

<pum-poŋɔ> /pʊm-poŋo/ 'right now'
<bum-boŋ> /bʊm-bɔŋ/ 'extreme darkness'
<kuŋ-koŋ> /kʊŋ-kɔŋ/ 'can/tin'

<dun-dɔŋ> **/dʊn-dɔŋ/** ‘court yard’

However, the process is blocked in associative compounds even when a final vowel in the first word of the compound is deleted. Instances of blocked homorganic nasal assimilation are presented in (8) adopted from Hudu (2014a:16).

(8) Nasal assimilation is blocked

Stem	Imperf	Perf	Nom	Phrase
[kɔŋ] ‘lose’	**kɔŋ-da**	**kɔŋ-ja**	**kɔŋ-bʊ**	**kɔŋ-li ‘**loose it’
[baŋ] ‘know’	**baŋ-da**	**baŋ-ja**	**baŋ-bʊ**	**baŋ-li** ‘know it’
[tam] ‘forget’	**tam-da**	**tam-ja**	**tam-bʊ**	**tam-li** ‘forget it’

The morpho-phonological reasons for the behaviour of the nasal in the above analysis will be provided in section 3.3 and are based on Hudu’s (2014a) proposal.

Other studies of Dagbani in which the issue of compounding is discussed are mainly the works of Olawsky(1999, 2002). In a study to determine a wordhood in Dagbani, Olawsky (2002) briefly discusses compounding. Olawsky observes that in Dagbani, compounds usually consist of two to three lexical roots but compounds longer than three roots are rare. He notes that in Dagbani the common compound forms are NN, NA and VN and that most of these compounds are either lexicalised or semantically transparent (2002:213-214). He maintains that for lexicalised compounds, the meaning of the compound is not dependent on the constituent word meanings, hence, their semantic structure is fossilised, which give a new meaning to the whole new word, as in (9).

(9) Sample lexicalised compounds in Dagbani (Hudu 2014a:7-8)

a. **[wab-gu ma]** ‘a type of rat’
b. **[saa paʔa]** ‘dragonfly’

In (9a), [**wab-gu**] ‘elephant’ and [**ma**] ‘mother’ would have a compositional meaning as ‘elephant’s mother’ but this is not the case. The same compositional meaning is lost in (9b) where [**saa**] ‘rain’ and [**paʔa**] ‘wife’, give the meaning ‘rain’s wife’. Additionally, it has been argued that based on Dagbani morphology, the head of a compound, which is the rightmost constituent, bears the number suffix in NN, and NA compounds. This behaviour of the constituent head leads Olawsky (2002) to argue that NA compounds are of the form NN. Another interesting observation about Dagbani compounds, especially, NA compounds is the behvoiur of some morphemes which Olwasky calls bound adjectives (10a) and how they pattern with regular adjectives in compounding (10b) in terms of their number suffixes.

(10) Bound adjectives in Dagbani (Olawsky 2002:215-216).

a. **/-la-a/** ‘male-SG’ **/-la-hi/** ‘male.PL’
/-lo-gu/ [lɔʔʊ] ‘male-SG’ **/lo-ri/** ‘male-PL’
/-ɲaŋ/ [nyaŋ] ‘female-SG’ **/ɲam-a/** ‘female-PL’
/sa-a/ ‘female-SG’ **/sa-hi/** ‘female-PL’

b. **bu-a** ‘goat-SG’ **nyaŋ** ‘female’ **[bu-nyaŋ]** ‘female goat’
no-o ‘fowl-SG’ **lɔʔʊ** ‘male’ **[no-lɔʔʊ]** ‘cock’
buŋ-a ‘donkey’ **nyaŋ** ‘female’ **[buŋ-nyaŋ] /buɲ-ɲaŋ/** ‘female donkey’

Even though the rightmost word is taken as the headword in most compounds in Dagbani, it seems that the adjectives in NA compounds in (10b) cannot function as the head of the compounds, given that in many of these compounds, for instance in [**bu-nyaŋ**] ‘goat-female (female goat), the modfiying word is [**-nyaŋ**] and not [**bu-a**].

[3] Olawsky (2002:216) calls these kinds of adjectives intermediate words because he observes that they are not lexical items since they cannot stand isolated and nor are they clitics, since they are morphologically complex (in that they take number suffixes). Furthermore, they are not suffixes, since they are composed of lexical root elements and number suffix endings.

The above observations indicate that in Dagbani compound word formation, different phonological processes may occur, including segment loss, assimilation or other neutralization processes, some of which I turn to in section 3.3.

3. Dagbani segment inventory, syllable structure types and phonological processes

In this section, I present the segment inventory of Dagbani in section 3.1 and the syllable structure types in 3.2. Then in section 3.3, I discuss a few phonological processes in Dagbani which are relevant for the purpose of the analysis done in this paper. The focus of this section is to lay the basis for a better understanding of both the statistical analysis and later discussions.

3.1 Dagbani consonant and vowel inventory

Dagbani has about 20-22 contrastive consonants and 6 short contrastive vowels, five of which have their long counterparts (Olawsky 1999; Hudu 2008, 2010, 2013). Both the consonants and vowels have their surface variants (allophones). Hence, all segments in square brakets [ ] are the surface variants of the segments they are closer to. The consonant inventory of Dagbani is shown in (11) while the vowel inventory is in (12). In the consonant inventory presented below, I adopt proposals made by Hudu (2010), where [ʔ], rather than [ɣ], is an allophone of /g/.

(11) Dagbani consonants (Hudu 2010:8) with some modifications

	labial	alveolar	palatal	velar	lab-vel	glottal
plosive	p b	t d[ɾ]	ʧ ʤ	k g	kp gb	[ʔ]
nasal	m	n	ɲ	ŋ	ŋm	
fricative	f v	s[ʃ] z[ʒ]		[x]		[h]
lat.		l				
glide			j		w[ʊ]	

(12) Dagbani vowels cited from Hudu (2010:15; 2013:49; 2014b:2)

[3] An anonymous reviewer pointed out that in current syntactic theory and analysis for many languages including Dagbani, it is assumed that D is the head of the DP rather than the noun. Based on this, he/she questions my assumption that the adjectives in NA compounds in (10b) could be problematic. I agree that D is the head of the DP in syntactic analysis. However, in the case of the NA compounds analysed here, I am particular about the semantics of the compounds. The entire meaning of the NA compounds is depended on the meaning of the noun. For instance, in a compound like [**bikurli**] 'grown child/big child', it the meaning of the noun [**bia**] 'child' which controls the meaning of the entire compound and distinguishes it from other compound such as [**tikurli**] 'old medicine', [**sa?ikurli**] 'old TZ', **[nahkurli**] 'old cow', and so it. This further confirms Olawky's (1999) observation that NA compounds are similar to NN in form, and perhaps in meaning in my opinion.

[i] iː	ɨ	ʊ uː [u]
ɛ eː [eː]		ɔ oː [o]
	a aː [ə]	

The short vowels are further put into two sets based on Advanced Tongue Root (ATR) features. The [+ATR] set is /i, u, e, o/ while the [-ATR] set is /ɛ, ɨ, ɔ, ʊ/. Hudu (2013:50) indicates that the [+ATR] variants [i, u, e, o] only surface in CV words whereas the [−ATR] vowels [ʊ, ɛ, ɔ] surface in CVC words and in all non-final positions outside of harmonic contexts. Hudu (2010, 2013) observes that /a/ has a [+ATR] variant based on [+ATR] harmony. Vowel length in Dagbani is contrastive and consists of the set /iː, eː, aː, oː, uː/.

3.2 Dagbani phonotactics and syllable structure types

Dagbani has several syllable structure types in the language but the most preferred types are shown in (13). Other types which are usually found in loanwords are noted. The common syllable structure types based on the available literature (Olawsky 1999, 2002; Hyman and Olawsky 2004; Hudu 2010) are stated below.

(13) Dagbani Syllable structure types

V	**CV**	**CVV**	**CVN**	**CVC**
a '2P.sg' **o** '3P.sg' **a.la.hi.chi** 'sin'	**ba** 'father' **tu** 'insult' **ku** 'kill' **chi** 'millet' **zo** 'friend'	**doo** 'man' **puu** 'farm' **nyii** 'bite' **nooŋa** 'bird' **noo** 'hen'	**kom** 'water' **koŋ** 'lose' **nem** 'grind' **poŋ** 'rotten' **dem** 'play'	**namda** 'sandal' **liʔri** 'money' **kɔbli** 'bone' **tandi** 'soil'

Hyman and Olawsky (2004:97-98) argue that verbal roots can also have strutures such as CVCi [**labi**] 'return', CVCCi [**zagsi**] 'refuse' and CVCiC [**lagim**] 'meet', where /i/ is predictable. They also assert that the underlying form for most verb stems are CV, CVN, CVC, CVV, and CVCC. Especially, words such as those in (14) are analysed to be CVV with the final /i/ being a verbal suffix (Olawsky 2002:212; Hyman and Olawsky 2004:97). On the contrary, Hyman and Olawsky (2004:98, footnote 2) maintain that if a verb stem is of the form CVV with the root vowel being /ii/ as in [**pii**] 'choose', then the verbal suffix /i/ will not occur. On the other hand, if the CVV has the root vowel to be /ee/ as in [**tee**] vs. [**teei**] 'remember', the occurrence of the suffix /i/ is assumed to be dialect variation.

(14) Verbs with CVV structure with an /i/ suffix

[looi] /loː**i/** 'cross'
[tooi] /toː**i/** 'fetch'
[tuui] /tuː**i/** 'mash'
[yooi] /joː**i/** 'open'
[deei] /deː**i** 'collect'

Additionally, Olawsky (1999) observes that syllable structure types such as (VC, CVVN, CCV, CVCC and N) also exist in the language, but most are found in loanwords. Nevertheless, it has been argued that Dagbani does not allow complex structures in its syllable forms (Hudu 2010; Gurindow M-minibo 2014), hence, CCV and CVCC structures will have vowel insertion to break the cluster. Finally, the only

consonants in word final coda position are /m, ŋ/ while word medial coda consonants are / **m, n, ŋ, r, l, b, ɣ**/ (Olawsky 1999, 2002).

There are also some phonological processes which are triggered when segments occur in certain sequence under some phonological environments. A few of such processes are discussed below.

3.3 Some phonological processes in Dagbani

In this subsection, I present a few morpho-phonological processes which may be relevant in compound formation in Dagbani.

Homorganic nasal assimilation. This is a phonological process in natural languages where a nasal assimilates to the place of articulation of the following consonant. For instance, in the compound formation process for [**lun.do.li**], as in (2a), it was noted that the nasal in the root [**luŋ**-] gets assimilated to the following [d] in [**do.li**] after the deletion of the vowel in [**luŋ-a**]. Another instance of a homorganic nasal assimilation, as pointed out by an anonymous reviewer, can be found in the compound word [**alizim-biɛʔʊ**], after the number suffix /-i/ has been deleted from the word [**alizini**] 'jinn'. The same reviewer further proposed that there is a homorganic nasal assimilation in [**bachiniŋda**], hence, the compound word should be [**bachininda**]. However, I disagree with the reviewer that there is assimilation process in the word [**bachiniŋda**] for reasons already provided in the data in (8). This compound is formed from the noun [**bachi**] and the verbal word [**niŋ-da**], where [**niŋ**] is a verb 'do' and [**niŋ-da**] is the imperfective form of the verb. Hence, [**niŋda**], based on Hudu's (2014) analysis, will not have any assimilation process between the /ŋ/ and /d/. Hudu (2014) argues that homorganic nasal assimilation is productive in bound root, clitics, and other affixes but underapplies in free standing morphemes, as exemplified in (8) above. As Hudu maintains, the basis for this underapplication is that in the contexts where nasal assimilation occurs, the nasal is at the right edge of the morphological bound unit but in the contexts where it fails, the nasal is at the right edge of a free-standing word (2014a, p.16), as illustrated in the data in (15).

(15) Nasal assimilation is blocked

a. Possessor	Possessed	Compound	Gloss
/kɔn-ga/ [kɔŋ:]	**bi-a**	**kɔŋ-bia**	'a leper's child'
/zɔn-ga/ [zɔŋ:]	**napɔŋ**	**zɔŋ-napɔŋ**	'a bat's foot'
[dagbana]	**bi-a**	**dagban-bia**	'a Dagomba's child'

b. **[dagban-]** 'dagomba' **[bi-a]** 'child' **[dagbambia]** 'Dagomba child'

The assimilation process noted in compounding, however, is motivated by the morphology of Dagbani. For instance, Hudu agrees that in many NN and NA compounds, the morphology of Dagbani predicts that the number suffix (and other root suffixes which are usually inflectional morphemes) in the first nominal constituent in the compound will be deleted. Hence, deletion in this sense is not motivated by phonology but rather by the morphology of the language. For instance, in the compound word [**alizimbiɛʔu**], the first noun [**alizin-i**] has a singular number suffix which is deleted so that the adjective (/**biɛ-gu**/ - [**biɛʔu**]) can have the number suffix. Thus, this deletion is motivated by the morphology of Dagbani. Nonetheless, after the deletion, a phonological environment is created for the nasal /n/ to undergo homorganic nasal assimilation to become /m/ in the compound [**alizimbiɛʔu**] or in other cases segment alternation may occur (Personal communication: 15 Jan. 2018). Furthermore, in (15b), Hudu (2014a, pp. 16-17) argues that [**dagban**-] is a bound root which is why /n/ undergoes homorganic nasal assimilation in the compound [**dagbambia**] due to the influence of the following bilabial stop.

Vowel epenthesis. Another phonological process attested in compound formation, especially involving loan words, is vowel epenthesis. Thus, vowels are usually inserted into consonant clusters to break them in loanwords or in roots before number suffixes are added or before compounds words are formed with those words. Accordingly, this insertion process is motivated by the phonotactics and syllable structure constraints of Dagbani. For example, the word *school* is adapted into Dagbani as **[shikuru**], where both segment insertion and alternation occur, as in (16a) and in the compound word in (16b).

(16) Vowel insertion in loan words

a. **/skuːl/ → [ʃikuru] ‘**school’
/bɔl/ → [bɔlli] ‘ball’

b. **[shikuru] ‘**school’ + **[duu] ‘**room’ **[shikuru-duu] /ʃikuru duː/ ‘**classroom**’**
[bɔlli] ‘ball’ + **[ŋme-ra] ‘**player’ **[bɔl-ŋmera] ‘**football player’

Segment alternations. Another productive phonological process in Dagbani is segment alternation. Some consonants change their form in certain environments due to phonological reasons. For example, /g/ →[ʔ]/V-V. Hence, /paga/ →[paʔa] ‘woman’. Other examples taken from (Hudu 2010:13) are in (17) below.

(17) Segment alternation in Dagbani

/s/ → [ʃ]	**/sia/**	**[ʃia]**	‘bee’
/k/ → [ʧ]	**/kilim/**	**[ʧilim]**	‘delay’
/z/ → [ʒ]	**/zili/**	**[ʒili]**	‘load’
/g/ → [ʤ]	**/gɛlinsi/**	**[ʤɛlinsi]**	‘hatred’

Some of these processes are co-created by the morphology and phonology of Dagbani and have relevant implications in compound formation. However, the most relevant of all these phonological processes is segment deletion (the main focus of this paper) which will be subjected to a quantitative analysis.

Having given this background information about Dagbani, where the segment inventory of the language, the syllable tructure types and some phonological processes are discussed, the next section will cover the methodology used in this paper.

4. Methodology

4.1 Data

To analyse the compound word formation processes in Dagbani, a random sample of 265 compound words of various types was obtained from the Dagbani-English Dictionary (Blench 2004). The sample was sent to a few other literate native Dagbani speakers, who were mostly teachers, for them to check whether those words are compound words attested in the language. The purpose of this verification was just to be sure that those words are actual words they hear in their daily use of the language. The sampled data consisted of NN compound type, NA compound type, NV compound type and VN compound type. In addition, there was some variation in the constituent contact margins across these compound types. The compound words were typed in an excel file document and converted into a comma-separated csv file which was fed into the R Software (R Core Team 2015). A sample of the data is shown in (18).

(18) Table 1: A sample of Dagbani compound words used in the analysis

	COMPWORD	WORDGLOSS	W.LENGHT	COMP.TYPE	DEL.SEG	C. C. MARGIN
1	**adakabila**[4]	small box	5	NA	Ø	V#C
2	**alizimbiɛʔu**	bad jinn	6	NA	V	V#C
3	**bachibia**	punctuation mark	4	NN	Ø	V#C
4	**bachiniŋda**	verbs	4	NV	Ø	V#C
5	**baʔbia**	soothsayer's stone	4	NN	V	V#C
6	**baʔgbee**	soothsayer's stick	3	NN	V	V#C
7	**bahisuŋ**	good life	4	NA	CV	V#C
8	**bahibiɛʔu**	bad life	5	NA	CV	V#C
9	**bahiyoli**	worse life	5	NA	CV	V#C
10	**ʤiliŋgbaŋ**	prayer mat	3	NN	Ø	N#C
11	**banyaŋ**	female dog	2	NA	V	V#C
12	**banyini**	dog's tooth	3	NN	V	V#C

In (18) above, the compound words are numbered. The word gloss for each compound is provided, the number of syllables in the word before deletion, the compound type it belongs to, the type of segment(s) that is/are deleted and the nature

[4] All compound words in this data are typed according to Dagbani orthography and are the surface forms after segment deletion. Also writing the compound types as NN or N+N is just a matter of style.

of constituent contact margin for each compound type is provided. Also, note that both wordlength and constituent contact margins of each compound are determined before segment deletion. For instance, in a compound like [**banyini],** the undelying form is [**baa + nyini**], where /a/ is deleted in [**baa**] in the compound word [**banyini**]. Hence, worldlength and constituent contact margins are determined from the underlying form of the compound before segment deletion.

4.2 Data analysis

Since the focus of this paper is to investigate what predicts segment deletion in compound formation process through a quantitative approach, the above data was subjected to Classification and Regression Tree (CART) and Random Forest analysis (Tagliamonte and Baayen 2012) using the R add-on package "party". Both the *ctree* (Hothorn, Hornik and Zeileis 2006) and *cforest* (Strobl Carolin et al. 2008) functions were used. Accordingly, to determine how each of the explanatory variables (compound type, word length, and nature of constituent contact margin) can predict what segments get deleted (the response variable) in compounding, the data, in a *csv file* format, was read into R and CART analysis was performed using the *ctree* function. In addition, to determine variable significance between the explanatory variables and their ability to predict the response variable, Random forest analysis was done using the *cforest* function.

The choice to use CART and Random forest analysis is motivated by the fact that the frequencies for some of the counts were zero. This makes it difficult to perform correlation and regression analyses because of the missing data in some cells. This observation has been noted by Tagliamonte and Baayen (2012) to be one of the advantages of using Random forest in linguistic variation analysis. Levshina (2015:291-294) also notes that CART is a method that uses regression and classification approach based on binary recursive partitioning. Hence, results produced through this process can pattern well with the results of correlation and regression analyses. Furthermore, she observes that CART and Random forest test if any explanatory (independent) variables are linked to the given response (dependent) variable and selects the explanatory variable that strongly correlates with the response variable. As also noted by Hothorn, Hornik, and Zeileis (2006:670), conditional inference trees select variables in an unbiased way. Thus, the partitions induced by this recursive partitioning algorithm are not affected by overfitting. It also produces a visual output in the form of a tree structure. CART and Random forest analysis also return the *p*-value through a process similar to bootstrapping, used in non-parametric methods, thus, a means for validating the obtained results. Finally, Random forest is grown from the conditional trees and determines the variable importance measure among the explanatory variables.

Given that compound formation involves a high-order interaction of processes from the lexicon, morphology, syntax and phonology and that different segments are involved, in the case of this study, within the response variable, CART and Random forest analysis was the most suitable statistical approach to use.

5. Results

The results obtained from the data analysis in section 4.2 are reported in the form of tables, charts and plots with some descriptive analysis in this section. The overall interpretation of the results is then presented in the section on discussion, where the results are linked to Dagbani morpho-phonological properties and other compound formation phenomena in other languages.

First, to determine the relative productivity of the compound types in Dagbani, a frequency distribution computed and a bar plot generated for the compound types, as in (19).

(19) Distribution of compound types in the sample data.

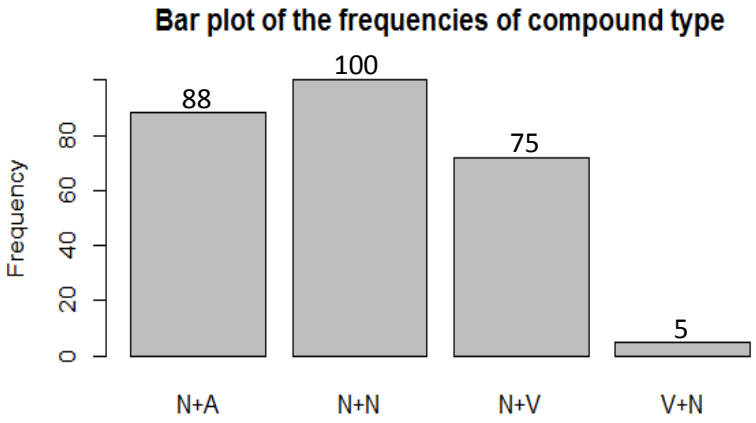


Figure 1. A Bar plot of the frequencies of compound type in Dagbani

From the figure above, it can be seen that NA and NN compound types are the most frequent compounds in Dagbani with VN compounds being the least common type. Even though NV compound also has a higher frequency, I will demonstrate later why it is not seen as a very productive form in Dagbani.

In addition, to examine the nature of the constituent contact margins in all the sampled data, a frequency table was generated as presented in (20).

(20) Table 2. The frequencies of word margin type in Dagbani compounds

Constituent contact margin	**N#C**	**C#V**	**V#C**	**V#V**
Frequency	40	1	223	1

From Table 2 in (20), it shows that the most frequently occurring constituent contact margin in Dagbani compounds is the [V#C] type. This is because most syllables in Dagbani have an onset. Also, the most preferred syllable structure type in the language is a CV, where most words begin with a consonant and end in a vowel segment. The second most occurring constituent contact margin is the [N#C] type. Since only /m, ŋ/ can occur as word final codas, it entails that word final Cs in a CVC structure are generally a nasal, which makes this constituent contact margin type less frequent. This explains why [N#C] word margins are not as widespread as the V#C type. As already stated in section 4.1, all constituent contact margins are determined in the underlying form of the compound before the segment that get deleted are noted. The statistics in Table 2 further provide support for the discussion on Dagbani syllable structure in section 3.2.

The analysis also revealed that among the deleted segments in compound formation, vowel deletion is the most frequent one followed by CV and Ø, as in (21).

(21) Segment deletion in compound types

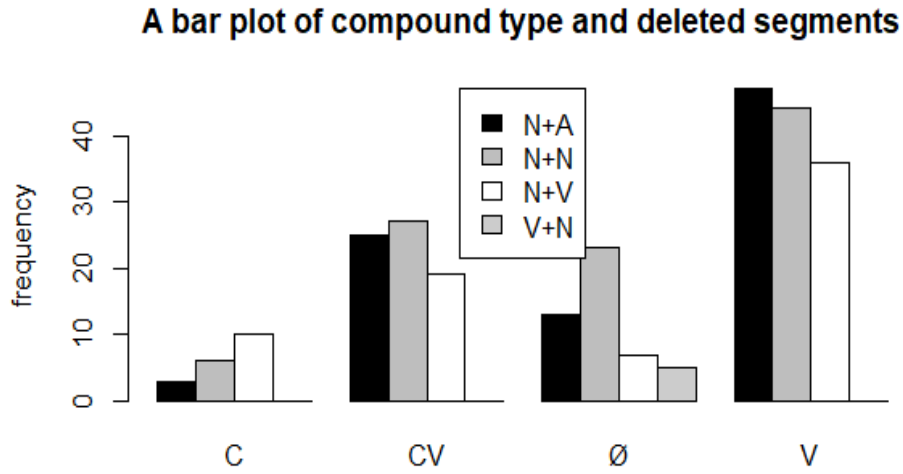


Figure 2. A Bar plot of segment deletion in the carious compound types

The pattern shown in Figure 2 above could be explained in relation to the fact that many lexical words (nouns and adjectives) in Dagbani have a number suffix which could be a vowel segment or a CV. Also, most lexical roots have /i, a, o, e, u/ vowels as suffixes (Olawsky 1999). Based on this observation, Olawsky (2002:214) noted that when a noun and an adjective are combined in a construction, the noun in the initial position loses its class suffix and what remains is the nominal root followed by the adjective. This means that, in most N+ compounds, the number suffix (at the word margin) often gets deleted. On the other hand, C or CV deletions occur to meet the preferred syllable structure of CV or CVN forms. Because, as indicated earlier, only /m, ŋ/ can occur in word final coda and /n, m, b, r, l, ŋ, ʔ/ in word medial coda positions. This limitation on coda consonants explains why C deletion is not widespread. On the other hand, where nothing is deleted at the constituent contact margins, it implies that those segments in the sequence N#C or C#V are permissible in the language, otherwise, there will be a vowel insertion to break the consonant cluster in a C#C word contact margin. These morphophonological properties of Dagbani lexical items are responsible for the kind of segment deletion exhibited in Figure 2 and as discussed in sections 3.2 and 3.3.

To examine the distribution of word length across the data, a frequency table was generated to determine the nature of word length in the sampled compound words, as presented in (22) below.

(22) Table 3: Frequency of word length in the compound words.

Word length	2	3	4	5	6	7
frequency	10	77	119	51	7	1

Even though these are compound words, it shows that out of the sample size of 265 words, 206 (77.7%) of the words fall between word length of 2 and 4 syllables. Yet, whether word length predicts segment deletion or not will be discussed later.

Additionally, to determine how each of the explanatory variables (word length, compound type and nature of word contact margins) could predict the response variable (deleted segments), three separate CART analyses were performed which returned the ctrees shown in (23).

(23) Classification and regression tree (CART) plots

a. Tree structure for compound type and deleted segment

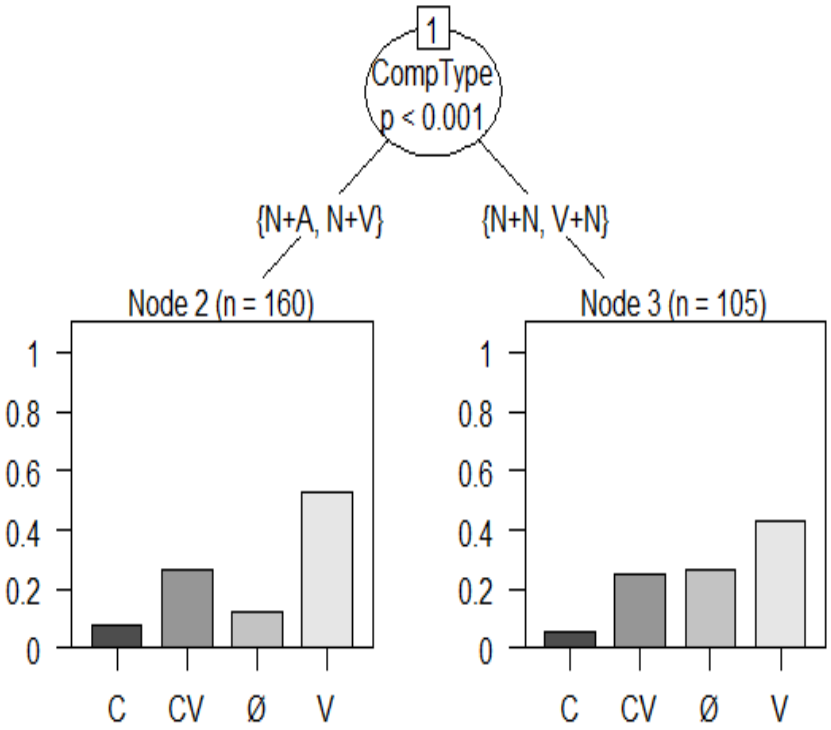


Figure 3. Classification and regression tree for compound type and deleted segments

b. Tree structure for constituent contact margin and deleted segment

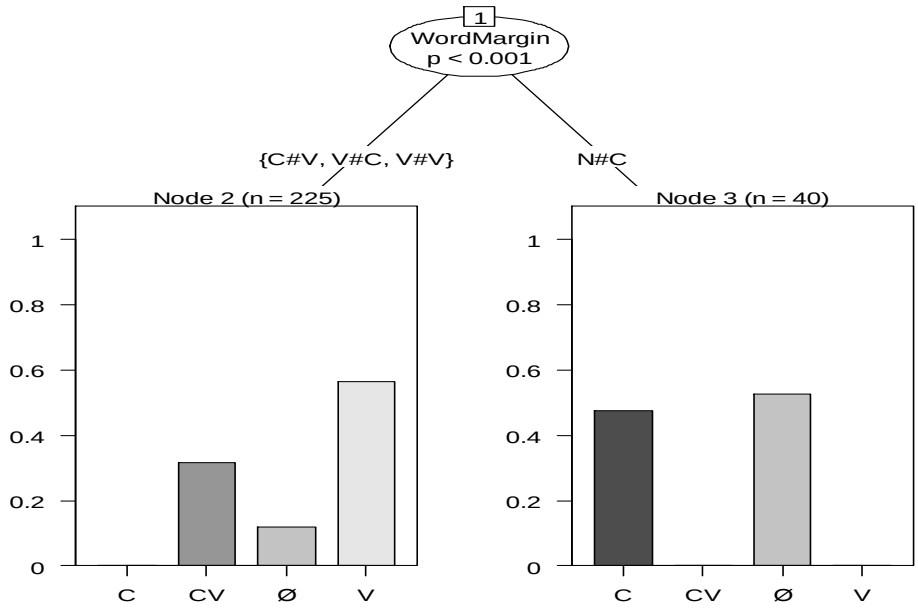


Figure 4. Classification and regression tree for nature of word margin and deleted segments

c. Tree structure for word length and deleted segments

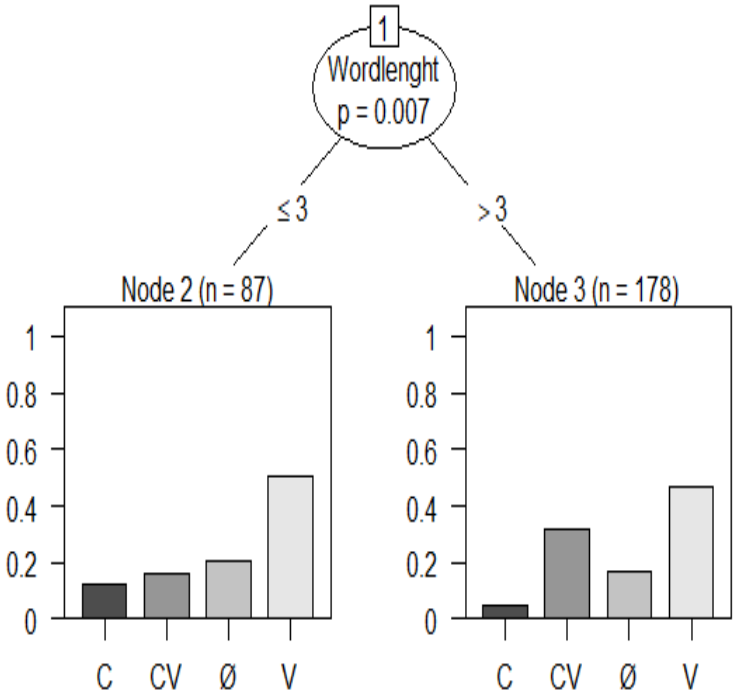


Figure 5. Classification and regression tree for word length and deleted segments

From the plots in (23), we can see that all the explanatory variables; word length, compound type and word margin are significant (with a *p*-value of < 0.001 for both compound type and constituent contact margin and a p-value of < 0.01 for word length) in predicting what segments will be deleted in Dagbani compound formation. In (23a), the tree shows that in compound type of NA and NV (Node 2), a V or a CV is more likely to be deleted (84 vowel deletion and 43 CV deletions) than in a compound type of NN and VN (Node 3) (where there are 45 V deletions and 26 CV deletions). This pattern can be explained by the morphology of Dagbani verbs. Both nouns and adjectives have number suffixes and given that the first word in NN and NA compounds are nouns, these suffixes will be deleted. In VN compounds, there will not be much segment deletion, since most verbs are monosyllabic with a few of them having a verbal suffix (usually a vowel). This explains why there are even less CV deletions in that cluster. On the other hand, in (23b), the tree shows that a CV and a V are more likely to be deleted in word contact margins [C#V, V#C, and V#V] (Node 2), whereas in [N#C] word contact margin, a C and Ø are more likely to be deleted. In (23c), it shows that in compounds with up to 3 syllables or less, a C is more like to be deleted (11 deletions) than in compounds with more than 3 syllables, where V (82 deletions), CV (57 deletions) and Ø are likely to be deleted, indicating that in longer words (words with many syllables) segment deletion is more common than in shorter words. Again, all these patterns show that, in general, segment deletion is influenced by the phonotactics and syllable structure constrains of Dagbani.

Notwithstanding, when the three explanatory variables are put together, the analysis shows that the nature of constituent contact margin is significant, at a *p*-value of (< 0.001), in predicting the response variable as shown in (24).

(24) Tree structure for all the explanatory variables and segment deletion

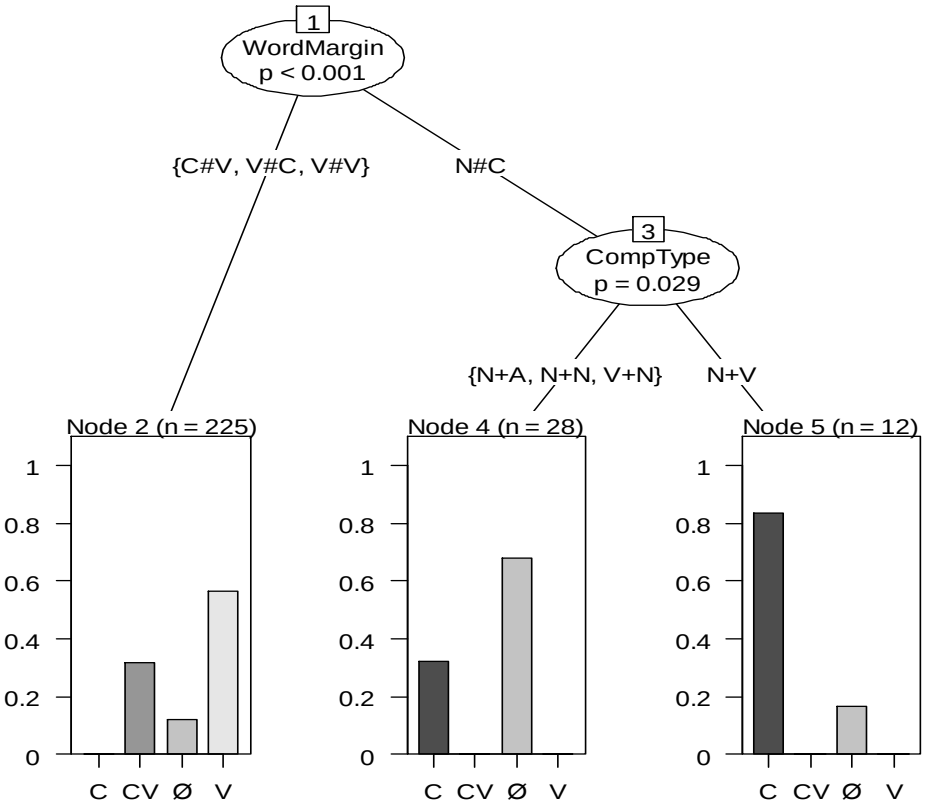


Figure 6. Classification and regression tree showing how compound type, word length and nature of word margin can predict segment deletion in Dagbani compound word formation

From Figure 6 in (24), the CART analysis shows that constituent contact margin (with a *p*-value < 0.001) is the variable that most likely can predict the kind of segment deletion in compound formation shown in Figure 2 above. It shows that in [C#V, V#C, V#V] (Node 2) word contact margins, a CV and a V are more likely to be deleted than in an [N#C] word margin. Furthermore, in all compound types with an [N#C] word contact margin (Node 3), nothing Ø is likely to be deleted in NA, NN and VN compound types (Node 4) than in NV compound type. On the other hand, in an [NV] compound type, a C is more likely to be deleted (Node 5) than in [NA, NN, VN] (Node 4) compound types. As can be seen, in an N#C constituent contact margin, the only possible segment to be deleted is a C or nothing is deleted at all while in [C#V, V#C, and V#V] word contact margins, a vowel or an entire syllable is more likely to be deleted, which again reflects the work of syllable structure constraints.

To evaluate the variable importance, as predicted in the CART analysis, the Random forest measure of variable importance confirmed that word margin was more likely to predict segment deletion in Dagbani compound formation than compound type. The result of the Random forest analysis is presented in (25).

(25) Random forest measure of variable significance

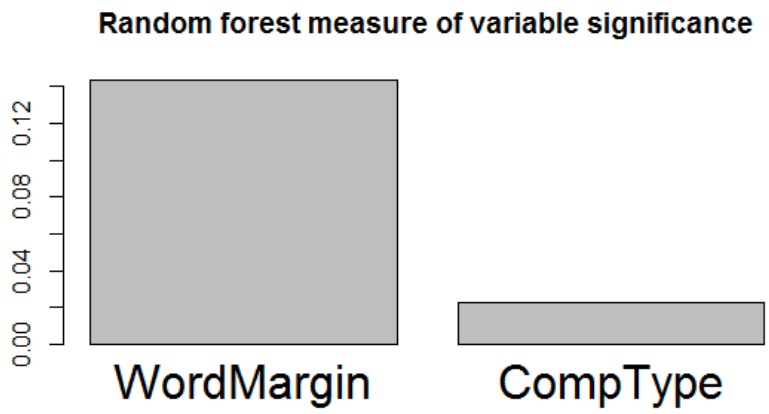


Figure 7. A bar plot of Random forest measure of variable significance

From Figure 7 above, the Random forest measure of variable significance confirms the CART analysis which indicates that word margin is the most likely predictor of segment deletion (with a *p*-value of < 0.001) in compound formation in Dagbani. Furthermore, the variable importance measure for all three variables shows that the nature of constituent contact margin is still significant in predicting segment deletion in Dagbani compound word formation as shown in the dot chart in (26) below.

(26) Variable significance measure.

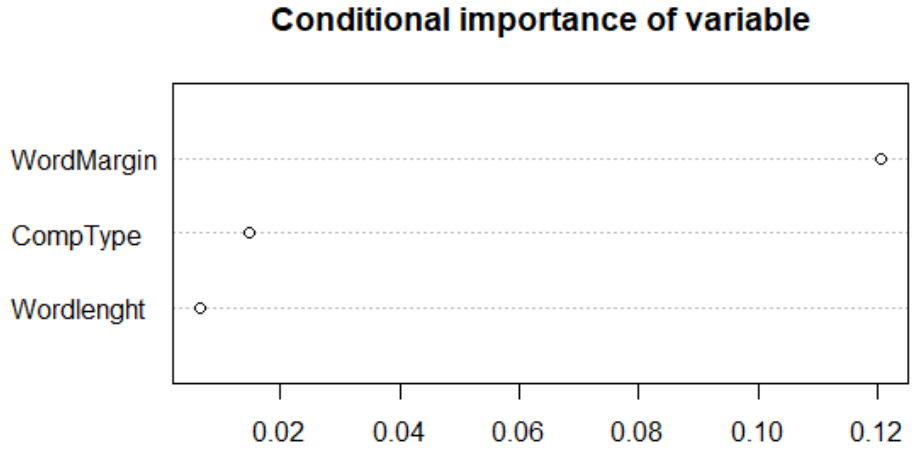


Figure 8. A dot chart showing the Random forest measure of variable significance

Figure 8 shows that the nature of word margin is the most significant variable that is likely to predict segment deletion in compound word formation followed by compound type and word length as the least predictor of segment deletion.

6. Discussion

A few interesting observations can be made about compound formation in Dagbani based on the results presented above. In relation to compounding as a productive word formation process in Dagbani, Olawsky (1999, 2002) notes that the most productive compound forms in Dagbani are NN and NA. This has been confirmed by the analysis done in this paper as shown in (19) and (20), which is also consistent with observations made in other languages, as Lieber (2005:378) argued that in English compounding, [NN, NA, AA, AN] compounds are all productive but with NN compounds being by far the most productive form. Fabb (1998) agrees with this observation about English as well as the Punjabi language. Additionally, it has been observed that in Dagbani compound formation, vowels (V) and syllable (CV) segments are the most deleted segments at the word contact margins (Abdul-Rahman 2013, Olawsky 1999, 2002; Hudu 2014a). This observation is borne out through Dagbani phonotactics and syllable structure constraints, since in the language, a V and a CV can both function as syllables (phonological units) or morphemes (morphological units). Cross linguistically, this phenomenon has been attested in other languages such as Fante, a language widely spoken in Ghana (Abakah 2004:182-183), where one of the vowels gets deleted in V#V word margin. Abakah argues that this phonological process occurs because these segments are intrinsic properties of morphemes in an isolated style, thus suggesting some kind of morphology-phonology interactions.

This notion of interaction between phonology and morphology then brings the discussion to the issue of the variable significance in prediciting segment deletion. Even though both the Random forest measure of varibale significance (25) and the CART tree (24) show that the nature of word contact margin (constituent contact margin) is more likely to predict segment deletion, both compound type and word margin seem to be good in prediciting segment deletion when considered individually as supported by (23b) and (23c). What this means is that both phonology (in terms of syllable structure types and processes) and morphology (in terms of morpheme structure constraints) in Dagbani are closely related. Most morphemes in Dagbani (both root and affixes) can be a CV, V or CVC. This explains why in all compound types and word contact margins of the form [C#V, V#C, V#V], V and CV deletion are more common than C or nothing (Ø). Patterns of this nature are what motivate Abdul-Rahman, (2013:230) to argue that final vowel (as a number suffix) deletion in Dagbani compound formation is motivated by the fact that there is a tight relationship between morphology and phonology. As a result, he concludes that segment deletion in compounding is to ensure that the resulting word observes both the phonotactics and morpheme structure constraints of the language.

From the analysis of the data, it was shown that word length is less significant in predicting segment deletion when combined with the other explanatory variables, but may predict segment deletion when examined alone (Figure 5, with a p-value of = 0.007). Even though Olawsky (1999) observes that word length, as a prosodic category, has some effects on Dagbani morpho-phonological process, that seems not to be the case in compound word formation. That said, it is possible that its effects are significant on lexical items but not on compound words, since compound words have their own internal semantics, syntax and morphology, which may be distinct from simple lexical items. Moreover, Olawsky (2002:210) observes that although compounds in Dagbani represent single grammatical words, each constituent in the compound maintains its vowel harmony values. This assertion further points to the fact that compounds may behave differently from lexical items in relation to some phonotactics of the language.

In addition, the observation made in the tree structure in (24) (Node 3) where [NA, NN, VN] are grouped together in (Node 4) from [NV] (Node 5) is very interesting. The fact that NA is grouped together with compounds that have their head as N (in NN and VN compounds) isn't a surprise. In Dagbani, both nouns and 'true'

adjectives have the same morphology in the sense that they are made of a root and a number suffix (for singular and plural), hence, they can be grouped into the nominal classes (Olawsky 1999:72). As Olawsky claims, NN and NA compounds in Dagbani have similar strutures (Olawsky 1999). This means that the adjective in [NA] compound has a nominal function as the head. The internal structure of these compounds can be analysed based on Booij's (2007, 2009, 2012, 2013) Constructional Morphology schema (a component of the Construction Ggrammar). According to Booij (2010:544), constructional morphology assumes that complex words, which are the outputs of morphological operations, can be listed in the lexicon as abstract schemas, which give a generalization about the form-meaning relationship in complex words. Booij mainatins that morphological schemas, therefore, have two functions: they express predictable properties of existing complex words and indicate how new ones can be coined. This perspective could help us resolve Olawsky (1999) observation that NN and NA compound in Dagbani are similar in structure. Examining the internal structure of Dagbani compounds could also provide some basis for my argument that the adjective in NA compounds is not the head, since all NA compounds have nominal functions in the language. Finally, it can also help in resolving the claim that verbs are not constituents of compounds with other lexical forms in Dagbani (Hudu 2010:38), even though VN and NV compounds are attested in Dagbani (Olawsky 1999) and in other languages.

For instance, it appears that both N+A and N+V compounds have similar internal structures, where in NA compound, the structure is [[[N]+[A]]A-x]N where x is a number suffix, as in (27), and in NV compounds, the internal structure is [[[N]+[V]]V-x]N where x is a nominal derivative suffix such as -rim, -li, -ga, -gu, etc., as in (28). Both data are taken from Olawsky (2004:135) but with some modifications.

(27) Examples of NA compounds (where SG = singular, class 1 or 4 nouns)

i. **tim + kpilli → [ti-kpil.li] ‘**medicine-round-SG1**(**medical pill**)**
ii. **bini + piɛlli → [bim-piɛl.li] ‘**thing-white-SG1 **(**shroud, white calico**)**
iii. **waʔu + ʒeʔu → [waʔ-ʒe.ʔu ‘**snake-red-SG4' **(**type of snake**)**
iv. **gban + piɛlli → [gban-piɛl.li] ‘**skin-white-SG1' (Western person)

(28) Examples of NV compounds (where NS = nominal derivative suffix)

i. **kom + nyurim → [ko-nyu-rim] ‘**water-drink-NS (drinking water)
ii. **tim + dirim → [ti-di-rim] ‘**medicine-eat-NS (medicine to be eaten)
iii. **suhu +kabli → [suhu-kab-li]** 'heart-break-NS (broken heart)

The data presented in (27) and (28) clearly illustrate the schema used by (Booij 2009) in the analysis of Dutch compounds in which NA has [NA]A-x]N and NV has [NV]V-x]N internal structures. These are called syntactic compounds, since they involve both compounding and derivation (Booij 2009:14-15)[5]. However, I leave the issue of internal structure of compounds in Dagbani for further research, since my focus in this paper is on segment deletion.

7. Conclusion

In conclusion, this paper examines segment deletion as a morphophonological process in Dagbani compound word formation. The paper found that NN and NA are the most productive compound types in Dagbani. It has also demonstrated that segment deletion is a widespread process in Dagbani compound formation. Vowels

[5] For details on the analysis of internal structure of complex and compound words in many other languages including English, interested readers are referred to the publications of Booij (2007, 2009, 2010, 2012, 2013, among others) and references therein.

and syllables (CVs) are deleted more than consonants across all compound types and are more likely to be deleted in [C#V, V#C, V#V] constituent contact margins than in an [N#C] constituent contact margin. Additionally, the nature of word contact margin was found to be the strongest predictor of segment deletion in compound formation than compound type and word length. However, both explanatory variables seem to be active in determining what segments get deleted, thus reflecting the close link between Dagbani morphology and phonology. The paper also argues that, even though NA and NV compounds would appear to be headed by an adjective and a verb respectively, their internal structures and functions may be that of nouns, which explains why all these compound types have nominal functions. Finally, this paper concludes that segment deletion in Dagbani compound formation is instantiated by both the phonotactics and morpheme structure constraints of the language.

However, there are a few limitations of the study. Due to the fact that there is no corpus data for the language, it was not possible to consider the impact of word frequency in compound formation and whether that has an influence on segment deletion. Another limitation is that the study did not delve much into the kinds of vowels that are deleted the most and whether these vowels are phonemes or morphemes. In future studies of this kind, considerations should be given to these factors so as to determine how much they play into segment deletion in compound formation. Finally, it could also be taken into account when does a C gets deleted in an N#C constituent contact margin and what might be responsible for that, since in an N#C word margin the first C gets deleted more in [NV] compounds than in [NA, NN, VN] compounds, where nothing gets deleted. Future studies should also consider analyzing compounding in Dagbani from CG/CM theoretical perspectives, since only a sketch of this is used in the current study.

Acknowledgement

This paper was initially presented as part of the course requirement in the course Quantitative Methods in Linguistics. I wish to express my appreciation to Prof Laura A. Janda, Prof Natalia Levshina and the students who took the course for their comments, suggestions and criticisms. I also want to thank both Janine T. Strøm and Mathew Burian for their comments on my language and style of argumentation. Finally, I extend my appreciation to the two anonymous reviewers for their impressive and thought-provoking comments and suggestions. All remaining errors are mine.

References

Abakah, N. E. (2004). Elision in Fante. *Journal of Africa & Asia, 4*, 181-213.

Abdul-Rahman, F. (2013). Elision in Dagbani. *International Journal of Linguistics, 5*(1), 219-230.

Blench, R. (2004). *Dagbani-English dictionary.* With contributions from Harold Blair Tamakloe, Harold Lehmann, Lee Shin Chul, André Wilson, Maurice Pageault, Knut Olawsky and Tony Naden. Tamale.

Bodomo, A. (1993). *Complex predicates and event structure: An integrated analysis of serial verb constructions in the Mabia languages of West Africa, Master's thesis.* Trondheim: University of Trondheim. Working Papers in Linguistics, no. 20.

Booij, G. (2007). Construction Morphology and the Lexicon. In Montamini, F., Gilles B. and Babil H. (eds.). *Morphology in Toulouse, Selected Proceedings of the 5th Decembrettes* (pp. 34-44). Somerville, MA: Cascadilla Proceedings Project.

Booij, G. (2009). Compounding and construction morphology. ResearchGate, pp. 1-21). doi. 10.1093/oxfordhb/9780199695720.013.0010

Booij, G. (2010). Construction Morphology. *Language and Linguistics Compass, 4*(7), 543-555. doi:10.1111/j.1749-818x.2010.00213.x

Booij, G. (2012). Construction Morphology, a brief introduction. *Morphology, 22*, 343-346. doi:10.1007/s11525-012-9209-x

Booij, G. (2013). Morphology in Construction Grammar. In T. Hoffmann and G. Trousdale (eds.). *The Oxford Handbook of Construction Grammar* (pp. 255-273). Oxford: Oxford University Press.

Fabb, N. (1998). Compounding. In Spencer, A. and Arnold Z. (eds.). *Handbook of Morphology* (pp. 63-83). Oxford: Blackwell.

Gurindow M-minibo, J. I. (2014). The Reality of Consonant Clusters in Dagbani Syllables. *International Journal of Linguistics, 6*(3), 231-242.

Hothorn, T.; Kurt Hornik, and Achim Zeileis. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics, 15*(3), 651-674. doi:10.1198/106186006X133933

Hudu, F. (2008). Markedness and Positional effects in Dagbani phonology: evidence from lenition. *Conference of the Canadian Linguistics Association.* University of Bristish Columbia.

Hudu, F. (2010). *Dagbani tongue-root harmony: A formal account with ultrasound investigation. PhD dissertation, University of British Columbia, Canada.* British Columbia.

Hudu, F. (2013). Dagbani tongue-root harmony: triggers, targets and blockers. *JALL, 34*(1), 47-73. doi:10.1515/jall-2013-0002

Hudu, F. (2014a). What is the phonological word in Dagbani? A positional faithfulness account. *Ghana Journal of Linguistics, 3*(1), 1-44.

Hudu, F. (2014b). Dagbani vowel phonology: competition between constraint hierarchies. In K. S. Kwesi Yankah, *A Legon reader in Ghanaian linguistics* (pp. 136-161). Oxfordshire: Ayebia Clarke Publishing.

Hyman, L. and Olawsky, K. (2004). Dagbani verb tonology. *Trends in African Linguistics, 4*, 97-108.

Levshina, N. (2015). *How to do Linguistics with R: Data exploration and statistical analysis.* Amsterdam: John Benjamins Publishing Company.

Lieber, R. (2005). English word-formation processes: Observations, issues, and thoughts on future research. In Stekauer, P. and Lieber R. (eds.), *Handbook of Word-Formation* (pp. 375-427). the Netherlands: Springer.

Michaelis, L. A. and Knud Lambrecht. (1996). Towards a Construction-Based Thoery of Language Function: The Case of Nominal Extraposition. *Language, 72*(2), 215-247.

Olawsky, J. K. (1999). *Aspects of Dagbani grammar: with special emphasis on phonology and morphology.* Munich: Lincom.

Olawsky, J. K. (2002). What is a word in Dagbani? In R. M. W. Dixon and A. Y. Aikhenvald (eds.). *Word: A Cross-Linguistic Typology* (pp. 205-226). Cambridge: Cambridge University Press.

Olawsky, J. K. (2004). What is a noun? What is an adjective? Problems of classification in Dagbani. *Journal of African Languages and Linguistics (JALL), 25*, 127-148. doi:10.0167-6164/04/025-00127

R Core Team. (2015). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org/

Strobl, Carolin; Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis. (2008). Conditional variable importance for Random Forest. *BMC Bioinformatics, 9*(307). Retrieved from www.biomedcentral.com/1471-2105/9/307

Tagliamonte, S. and Baayen, R. H. (2012). Models, forest and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change, 24*(2), 135-178.