Title
Grammatical gender agreement with nominal compounds in Somali

Permalink
https://escholarship.org/uc/item/14k8p79b

Author
Kaldhol, Nina Hagen

Publication Date
2021-01-16

Grammatical gender agreement with nominal compounds in Somali

Nina Hagen Kaldhol
University of California San Diego

1   Introduction

Compounds pose a puzzle in languages with grammatical gender: if two nouns with different genders are compounded, what is the gender of the compound? For some languages, it has been argued that gender is determined by the *head* of the compound. While this may be a straightforward generalization in some cases, the situation is not always that simple: the present paper illustrates that in Somali (Cushitic, the Horn of Africa), noun-noun compounds with gender mismatches control variable gender agreement. As the example in (1) shows, a feminine-masculine compound may control different gender values on different agreement targets, reminiscent of the behavior of *hybrid controllers* (Corbett 2006, pp. 160–165).

(1)   Wadda + hálaq-u   waa   fiicán tahay
      road + snake-**M**.DEF.NOM DECL good 3SG.**F**.COP.PRS
      'The way snakes go/move is good.'   cf. *waddá-du (F)*, *hálaq-u (M)*

In the present paper, I will argue that noun-noun root compounds in Somali are semantically left-headed. This is straightforwardly shown if the relationship between the two members is one of hyponymy: for example, a *laf + dhábar* 'spine (lit. bone-back)' is a type of *láf* 'bone', not a type of *dhábar* 'back'. However, Somali has postposed definite articles, as in *dhábar-ka* 'back-DEF' and *hálaq-u* 'snake-DEF'. That is, while the semantic head of a compound is to the left, this agreement target surfaces on the right edge. If headedness determined the gender of

the compound, one would expect the gender agreement to be consistently predictable from the gender of the semantic head. I will show that while this holds true for subject-verb agreement, the definite article shows variation, and can either reflect the gender of the semantic head of the compound, or the gender of the member to which it is linearly adjacent.

Somali furthermore exhibits *tonal gender marking*, such that there is a correlation between the gender of nouns and their tone pattern. In noun-noun root compounds, it is the *final* member which determines the tone pattern. If the gender of this member is different from the initial member, the result is thus a mismatch between the gender of the semantic head on the one hand, and the tone pattern of the compound on the other.

Previous research has identified *false universals* in compounding related to the notion *head*, and there is a need for more in-depth research on compounds, both typologically and in particular languages (Guevara and Scalise 2009, pp. 122–124). The present study aims to take up this cue, and investigates the role of headedness in determining gender in compounds in Somali. It will be argued that the properties related to headedness (semantic head, morphosyntactic head, morphosyntactic locus) can be split between different elements within a morphological construction, and thus cause the variable agreement patterns that we see. Furthermore, it will be argued that a complete understanding of the agreement variation cannot be based on a single construction: rather, the attested variation in the definite article is due to competing properties of the Somali system as a whole. These properties work together to allow for competing generalizations: the location of the semantic head relative to the definite article, the tonal gender marking, the fact that the article is phonologically bound to the final member of the compound (the non-head), and the coexistence of left-headed noun-noun root compounds and other, right-headed morphological constructions within the same system.

The present paper is structured as follows. In section 2, the Somali gender system and compound strategies are presented, followed by some background on the notion *head*. In section 3, novel data is presented, showing agreement patterns with noun-noun root compounds, and revealing variability and gender mismatches involving both agreement and tone. In section 4, these agreement patterns are compared to those of other types of gender mismatches reported from other languages, such as those involving hybrid nouns, conjunct agreement, and agreement attraction. In section 5, it will be argued that the agreement variation in Somali is the result of several factors within the system which work together to allow for competing generalizations to be made by speakers. Section 6 concludes.

2 Background

2.1 The gender system in Somali

Somali (Cushitic, the Horn of Africa) has a gender system with two genders, feminine and masculine. One of the cues to gender is the choice of definite article:[1] when the noun is masculine, the definite article is *ka*. When the noun is feminine, the definite article is *ta*. The articles further undergo various sandhi alternations, such that they each have four different allomorphs, as illustrated in (2)–(3).[2] Long vowels are represented here with two consecutive vowels (aa) and High tones are marked with acute accents (*á*).[3]

(2) **Masculine nouns**
    a. *dáb-ka*   [dábka]   'the fire'
    b. *qálbi-ga*   [ɢálbiga]   'the heart, spirit'
    c. *aabbá-ha*   [aabbáha]   'the father'
    d. *gó'-a*   [góʔa]   'the portion'

(3) **Feminine nouns**
    a. *xeéb-ta*   [ħeébta]   'the shore'
    b. *taariíkh-da*   [taariíχda]   'the history, date'
    c. *gabádha*   /gabáɖ-tɑ/ – [gabáɖɖɑ]   'the girl'
    d. *úsha*   /úl-ta/ – [úʃa]   'the stick'

The choice of allomorph of /ka/ and /ta/ is predictable from the final segment of the noun. The patterns are as follows: The [ga] allomorph of the masculine definite article (2b) is found after masculine nouns ending in [g] or a vowel, except the vowel [e]. If the noun ends in [e], as in *aabbé* 'father', the definite article takes the form [ha], and the final vowel of the noun turns into [a] (2c). The initial segment of the masculine definite article is deleted after any of the consonants [ɢ χ ħ ʕ ʔ h] (the "guttural" consonants), as in (2c). The form [ka], illustrated in (2a) is the "elsewhere" case and is found after any of the remaining possible segments, which are [b d ɖ f tʃ l n r ʃ].

---

[1]There is no indefinite article in Somali. Nouns without any determiner (article, demonstrative, possessive) have an indefinite meaning.

[2]There is variation between and within speakers in the realization of singleton and geminate stops which is not indicated here – see Bendjaballah and Le Gac (2019) for a discussion and a recent production experiment.

[3]In the remainder of the paper, words will be transcribed following the Somali orthography. The deviations between the orthographic symbols and the IPA symbols are as follows: *c* represents a voiced pharyngeal fricative [ʕ], *dh* is a retroflex stop [ɖ], *j* is a postalveolar affricate [tʃ], *kh* a uvular fricative [χ], *sh* a postalveolar fricative [ʃ], *x* a voiceless pharyngeal or epiglottal fricative, [ħ] or [ʜ], *y* is a palatal glide [j], and ' is a glottal stop [ʔ].

As for the feminine definite article, the [da] allomorph (for some speakers [ða]) illustrated in (3b) is found after feminine nouns ending in [d], a vowel, or any of the consonants [ɢ χ ħ ʕ ʔ h] (the "guttural" consonants). The form is [ɖa] if the nouns ends in [ɖ] (3c), and if the noun ends in [l], that [l] coalesces with the initial [t] of the feminine definite article and they surface as [ʃ] (3d). The form [ta] (3a) is the "elsewhere" case and is found after any of the remaining possible phones, which are [b f g n r s]. For more examples and details, see Barillot (2002, pp. 231–240); Green and Morrison (2018); Saeed (1999, pp. 28–29).

Besides the choice of definite article, another cue to gender is subject-verb agreement. When the subject is third person singular, the form of the verb indicates the gender of the subject, as illustrated with the copula verb in (4)–(5), whose form is *yahay* when the subject is a masculine noun (4), and *tahay* when the subject is a feminine noun (5).

(4) **Masculine nouns**

   a. Libáax-u   waa xún yahay.
      lion-**M**.DEF.NOM DECL bad 3SG.**M**.COP.PRS
      'The lion is bad.'

   b. Málab-ku   waa fiicán yahay.
      honey-**M**.DEF.NOM DECL good 3SG.**M**.COP.PRS
      'The honey is good.'

(5) **Feminine nouns**

   a. Shimbír-tu   waa fiicán tahay.
      bird-**F**.DEF.NOM DECL good 3SG.**F**.COP.PRS
      'The bird is good.'

   b. Saliíd-du   waa xún tahay.
      oil-**F**.DEF.NOM DECL bad 3SG.**F**.COP.PRS
      'The oil is bad.'

These examples illustrate gender agreement with *singular* nouns. When the subject is a plural noun, the gender is not indicated through subject-verb agreement, as the verb instead has a single third person plural form (*yihiin* in the case of the copula verb). The only cue to gender for plural nouns is thus the bound determiners, such as the definite article. However, as the examples in (6) illustrate, the choice of definite article is "switched" in the plural forms of these nouns, such that the masculine noun *libáax* 'lion' gets a feminine definite article in its plural form (6a), while the feminine noun *shimbír* 'bird' gets a masculine definite article in its plural form (6b). This phenomenon is called *gender polarity* (see e.g. Hetzron 1967, 1972), and this type of patterning is common across Cushitic languages (Mous 2012).

(6) **Plural nouns**

  a. Libaaxyá-du   waa xún yihiin.
      lion.PL-**F**.DEF.NOM DECL bad **3PL**.COP.PRS
      'The lions are bad.'

  b. Shimbirú-hu   waa xún yihiin.
      bird.PL-**M**.DEF.NOM DECL bad **3PL**.COP.PRS
      'The birds are bad.'

However, it has been argued by Nilsson (2016b) that the definite articles do not show gender agreement synchronically, as they are suffixes and phonologically bound to their nouns, rather than independent words. The argument is based on the idea that gender is "reflected in the behavior of associated words" (Hockett 1958, p. 231) – that is, while nouns have inherent gender (they are agreement *controllers*), the evidence for it does not come from the noun itself, but rather "from agreement markers attached to other sentence elements" (Corbett 1991, p. 147) (that is, to agreement *targets*). Nilsson proposes that the switch that we see from e.g. *libaax-u* 'the lion' (M) to *libaaxya-du* 'the lions' ("F") is a property of the paradigms of these nouns and their inflectional classes, rather than examples of gender agreement. Certain nouns do not show such a switch (e.g. *míis-ka* 'the table' (M) – *miisás-ka* 'the tables' (M), *úgax-da* 'the egg' (F) – *ugxaán-ta* 'the eggs' (F)), and the form of the article in the plural is predictable from a combination of grammatical and morphophonological properties of the stem (Nilsson 2016b, pp. 455–456). Nilsson further argues that gender simply is not relevant in the plural because there are no agreement targets that show gender distinctions with plural nouns – that is, under the assumption that suffixes are not agreement targets.

This discussion illustrates the need for independent evidence of a noun's gender, from uncontroversial agreement targets, such as verbs. This point will be taken up again below. I remain agnostic to the question of whether the definite articles are suffixes (hence parts of the inflectional paradigms of nouns) or unbound forms (hence possible agreement targets), but I will refer to them as agreement targets for the purposes of the present paper.

The situation is different for derivational suffixes: they do not act as agreement targets – rather, they assign their own gender. This is illustrated by the examples in (7): these are two deverbal nouns with different genders, determined by their suffixes.

(7) **Derivational suffixes and gender**

  a. *qor* 'write' – *qoraal* 'text, document' (M)

  b. *qor* 'write' – *qoris* '(act of) writing' (F)

2.2   Gender, tone and declension in Somali

Another property of the gender system of Somali, is that there is a correlation between gender and tone (Armstrong 1934), and many pairs like the ones provided in (8)–(9) (Hyman 1981, p. 172), where the masculine member of the pair has a penultimate High tone, and the feminine member of the pair has a final High tone.

(8)   **Masculine nouns, penultimate H**
  a. *ínan* 'boy'
  b. *qaálin* 'young he-camel'
  c. *daméer* 'he-donkey'

(9)   **Feminine nouns, final H**
  a. *inán* 'girl'
  b. *qaalín* 'young she-camel'
  c. *dameér* 'she-donkey'

Hyman (1981) divides Somali nouns into three declensions based on their tone patterns, and I will take this categorization as a starting point here.[4] Examples are illustrated in tables 1–3, and additionally, some feminine nouns which are not mentioned by Hyman and do not seem to fit in any of these three declensions, are provided below. Note that there are different factors that determine the form of nouns in different declension (for example gender, type of focus marker, or position in the utterance). The number of logical combinations of all the relevant factors is huge, so for simplicity, only partial paradigms are provided here. The tables are furthermore formatted differently from each other, to reflect the factors relevant the declension in question.

Nouns of declension 1, illustrated in table 1, are divided into two subclasses based on gender (1a includes masculine nouns, 1b include feminine nouns), and the two subclasses have different tone patterns.

Nouns of declension 2 either end in the vowel *-e* and are masculine, or end in the vowel *-o* and are feminine. The tone patterns are the same regardless of gender, and have been investigated in detail by Le Gac (2002, 2003, 2016). As table 2 shows, the location of the High tone depends on whether the noun occurs utterance-finally or not, whether it functions as a subject or not, and whether or not it is followed by a definite article (or another determiner or modifier).

---

[4]There are substantial disagreements about the number of declensions needed for Somali, partly because different types of inflectional categories or properties (number, case, tone patterns) are taken into account by different scholars, and partly because their works are couched within different frameworks (ranging from Andrzejewski (1964) which proposed eight declensions based on descriptive generalizations, to CV Phonology and Distributed Morphology by Lampitelli (2013), who argues that the concept "noun class" is redundant, at least for non-derived nouns, and aims to do away with declensions and gender completely). In the present paper, plural forms and their part in the inflectional classes and gender system are left aside for simplicity. Hyman's categorization, which is based on tone patterns, is chosen here because it captures the relevant data.

		Example	Gender	Tone
1a.	INDEF	ínan	M	Penultimate H
	INDEF.NOM	inan		No H
	DEF	ínan-ka		(Stem-) penultimate H
	DEF.NOM	ínan-ku		(Stem-) penultimate H
		'boy'		
1b.	INDEF	inán	F	Final H
	INDEF.NOM	inan-i		No H
	DEF	inán-ta		(Stem-) final H
	DEF.NOM	inán-tu		(Stem-) final H
		'girl'		

Table 1: Partial paradigms for nouns of Declension 1

		Masculine	Feminine	Tone
[+Focus]	[+Final]	waraábe	abeéso	Penultimate H
	[-Final]	waraábe ~ waraabé	abeéso ~ abeesó	Final/penultimate H
[-Focus]	[+Subject]	waraábe	abeéso	Penultimate H
	[-Subject]	waraabé	abeesó	Final H
DEF		waraabá-ha	abeesá-da	(Stem-) final H
DEF.NOM		waraabú-hu	abeesá-du	(Stem-) final H
		'hyena'	'python'	

Table 2: Partial paradigms for nouns of Declension 2 (nouns ending in *-e/-o*)

Nouns of declension 3 consists mostly of plural nouns (which go beyond the scope of the present study), but also a "a limited number of exceptional [singular] nouns" (Hyman 1981, p. 180). They are "exceptional" because in spite of being masculine, they do not show the same tone patterns as the masculine nouns in declension 1 or 2, but rather the ones illustrated in table 3 (which also are found with the aforementioned plural nouns) (see also Banti 1988b, 2016). Andrzejewski (1964, pp. 35–38) lists 91 such singular nouns. Unlike the nouns in declension 2, which all end in *-e/-o*, nouns in declension 3 do not have anything in common phonologically that can predict their membership in this class (examples include *maroodi, abti, wax, sambab*, etc.).

There is a class of nouns which do not fit into any of the declensions above, namely feminine nouns with a penultimate High tone (10). There seems to be only a handful of such nouns in the language (Andrzejewski 1964, pp. 39–40). They do not occur in any of the compounds included in the present study, but they are listed here to illustrate that the correlation between tone and

Environment	Example	Tone
Absolute and [+Final] Focused by *ayaa* DEF DEF.NOM	maroodí maroodí ayáa maroodí-ga maroodí-gu	Obligatory (stem-)final H
Citation form Elsewhere	maroódi ∼ maroodí maroodí ∼ maroodi 'elephant'	Final/penultimate H Optional final H

Table 3: Partial paradigms for nouns of Declension 3

gender in Somali is not absolute: *sábti-da* 'the Saturday' and *inán-ta* 'the girl' are both feminine nouns, but *sábti-da* has a penultimate High tone, and *inán-ta* has a final High tone.

(10)   **Feminine nouns with a penultimate High tone**

    a. *sábti-da* 'the Saturday'

    b. *úgax-da* 'the egg'

    c. *sáddex-da* 'the three'[5]

    d. *Máryan* (female name)

While tone is a cue to gender in Somali, it is not a perfect one: the gender and declension of a noun cannot be predicted solely based on its tone pattern. As we have seen, both *inán* 'girl' and *maroodí* 'elephant' have a final High tone, but the former is feminine, and the latter is masculine. The gender of a noun is determined through agreement, as explained in section 2.1.[6] The assumptions in (11) will guide the analysis of tone patterns in compounds in section 3.5 below.

(11)   **Guiding assumptions**

    a. The gender of a noun is reflected in the form of its agreement targets.

    b. The form of a noun (including its tone pattern) is a property of its inflectional class or form-class, not its gender.

---

[5]All numerals from 2–8 are feminine nouns and have a penultimate High tone.

[6]Similar mismatches are found in other gender systems – Harris (1991) discusses an analogous situation found with the suffixes *-a/-o* in Spanish. There is a correlation between form (suffix) and gender, however, Harris argues that analyzing these vowels as exponents of gender "errs in conflating biological sex (male versus female), grammatical gender (masculine versus feminine), and form class (*-o* versus *-a*)" (Harris 1991, p. 28). He further argues that "these are interrelated but autonomous domains of linguistic generalization" (Harris 1991, p. 28), and concludes that "Tradition is wrong: the *-o* and *-a* in question belong to a set of exponents of declensional class" (Harris 1991, p. 59) (not gender).

c. There is a correlation between tone and gender in Somali, but it is not a perfect one. The correlation follows from the correlation between gender and inflectional class.

d. Tone is not an exponent of gender in Somali; the nominal tone patterns are lexical properties of nouns and their paradigms.

2.3 Nominal compounds in Somali

Nominal compounds in Somali come in different shapes. In the simplest case, there are two noun roots, as in (12). Noun-noun root compounds have one High tone and a single definite article at the right edge, while noun-noun *phrases* (genitive constructions) have two High tones and two articles (13).[7]

| (12) | **Noun + Noun root compound** | (13) | **Noun + Noun genitive construction** |

(12) **Noun + Noun root compound**

gaari + fáras-ka
car + horse-M.DEF

'the horse cart'

(13) **Noun + Noun genitive construction**

búug-ga   macállin-ka
book-M.DEF teacher-M.DEF

'the book of the teacher'

Semantically, noun-noun genitive constructions have associative or possessive meanings, as in (13) above. In noun-noun compounds, many semantic relationships are possible, including figurative meanings, as illustrated in (14).

(14) **Non-transparent semantic relationships**

a. fool + díin
   labor.pain + turtle
   'false labor pains'

b. af + gáab
   language/mouth + shortness
   'quiet person; a smallpox-like disease'

Other types of nominal compounds include noun-adjective compounds (15), noun-verb compounds (16), and more complex examples such as noun-adposition-verb (17) or noun-adjective-verb (18) (examples from Saeed 1999, pp. 156–159).

---

[7]Noun-noun phrases may also have a single definite article, but in this case, on the *initial noun*, as in *áf-ka hooyó* 'the mother language'. The meaning here is different from one with two articles: *áf-ka hooyá-da* 'the language of the mother' would refer to the language of a particular mother.

(15)   **[Noun Adjective]** _{Noun}

    bad + weyn
    sea + big

    'ocean'

(16)   **[Noun Verb]** _{Noun}

    biya + dhac
    water + fall

    'waterfall'

(17)   **[Noun Adposition Verb]** _{Noun}

    cid + la + joog
    people + with + stay

    'domestic animal'

(18)   **[Noun Adjective Verb]** _{Noun}

    war + xun + gee
    news + bad + carry

    'large poisonous spider sp.'

The rightmost members of the compounds in (17)–(18) can be analyzed as deverbal or agentive nouns (Banti and Jama 2016; Saeed 1999, pp. 156–159). Unlike these, some compounds have overt derivational suffixes, such as *-aal* (meaning 'product of X') (19a) or *-n* (meaning 'act of X-ing') (19b).

(19)   **Compounds with derived nouns**

    a. hadal + qor-aal
        talk + write-NMLZ

        'protocol (lit. talk + document)'

    b. dal + ku + celi-n
        country + to + send.back-NMLZ

        'repatriation'

Since the object of study in the present paper is gender in noun-noun compounds, and derivational suffixes in Somali assign gender (see section 2.1), such suffixes may override any structural gender assignment patterns that the present study aims to uncover. Therefore, the focus of the paper will be on noun-noun *root* compounds. Since the vowels *-e* and *-o* in Somali assign both inflectional class and gender (see section 2.2), compounds such as the ones in (20) will similarly be left aside for the present purposes.

(20)   **Compounds with nouns ending in *-e* and *-o***

    a. il + tuke
        eye + crow

        'crow's eye (plant sp.)'

    b. liin + bambeelmo
        citrus.fruit + grapefruit

        'grapefruit'

A challenge to this particular criterion is that there are many cases of heterosemous root pairs in Somali, that is, examples of roots that may function either as nouns or verbs. Examples include the final member of the compounds in (17)–(18) above. Another example is *xanuun*, which means both 'pain' and 'to be in pain'. In such cases, it is not clear whether the noun is derived from the verb or the verb is derived from the noun. This may even be the wrong question to ask–rather, roots of this kind are *heterosemous* (they can function as both nouns and verbs). Therefore, it is equally problematic to determine whether examples such as the ones in (21) involve noun-noun root compounds, or nominal noun-verb compounds. Examples like these are therefore left aside in the present study, although their importance for a broader understanding of compound gender will be taken up in section 5.

(21)   **Compounds with a heterosemous root**

    a.   madax + xanuun
       head + pain
       'headache'

    b.   calool + xanuun
       stomach + pain
       'stomachache'

Another challenge for studying compound gender in noun-noun root compounds in Somali, is that most of these compounds involve either two masculine nouns (22) or a feminine noun followed by a masculine one (23). There are also a few compounds in which both members are feminine (24), or the initial member is masculine while the final member is feminine (25) – however, there are few of these, for reasons to be outlined below. The gender values indicated in (22)–(25) refer to the gender of the members. Compound gender and agreement patterns will be presented in section 3.

(22)   **A Masculine + Masculine compound**

    gaari + fáras
    car.(M) + horse.(M)

    'horse cart'

(23)   **A Feminine + Masculine compound**

    laf + dhábar
    bone.(F) + back.(M)

    'spine'

(24) **A Feminine + Feminine compound**

kabsar + caleén
herb.(F) + leaf.(F)

'cilantro'

(25) **A Masculine + Feminine compound**

bur.(M) + saliíd.(F)
flour/cake + oil

'fritter'

For most feminine-feminine and masculine-feminine compounds, the final member is not a root, but has one of the suffixes *-eed/-ood/-aad* (of which *-aad* is lexically restricted, and *-ood* is used mostly with plural nouns ending in *-o*; only *-eed* will be illustrated here). Saeed (1999, p. 156) calls these suffixes *genitive case endings*, and therefore analyzes examples such as the ones in (26)–(27) as "lexicalized noun phrases" (the suffix is glossed as ATTR for *attributive* for reasons explained below). For comparison, a noun-noun genitive phrase is illustrated in (28) (from Saeed 1999, p. 64).

(26) **Feminine + Feminine-eed**

abeeso + bad-eed
snake.(F) + sea.(F)-ATTR

'moray eel (lit. snake of sea)'

(27) **Masculine + Feminine-eed**

libaax + bad-eed
lion.(M) + sea.(F)-ATTR

'shark (lit. lion of sea)'

(28) **The genitive construction**

áf   shimbir-eéd
mouth.(M) bird.(F)-GEN

'mouth of a bird, beak'

According to Saeed, most feminine nouns form their genitive this way, while masculine nouns usually show tonal genitive marking (e.g. *díbi* 'ox', *dibí* 'ox.GEN', Saeed 1999, p. 44). However, it has been argued by Banti (1986, 1988a) and Nilsson (2016a, 2019) that while these suffixes may have been genitive markers historically, they have been reanalyzed as derivational suffixes

forming attributive nouns. When used in compounds, they thus mark the final member as the non-head. Synchronically, they are used to form novel compounds, even when the final member is masculine (29)–(30).[8]

(29)   **Masculine-Masculine-eed**

sannad + dugsiyeed
year.(M) + school.(M).ATTR

'school year'

(30)   **Feminine-Masculine-eed**

hawl + qarmeed
work.(F) + nation.(M).ATTR

'manual labor'

The existence of compounds such as the ones in (29)–(30) along side the types in (26)–(27) above does not change the fact that the majority of noun-noun *root* compounds either are masculine-masculine or feminine-masculine, while feminine-feminine and masculine-feminine root compounds are few. This is challenging for a study of compound gender because it is difficult to generalize over the patterns found with any root compounds ending in a feminine noun, simply due to their small number. Therefore, a list of compounds of the type noun + noun-ATTR is also included in the present study for comparison. More details about the word list is provided in section 3.1.

2.4   Headedness in compounds

In many languages, the notion *head* is appealed to when describing gender in compounds. For example, compounds in Dutch are analyzed as right-headed because the rightmost member determines the semantic class, syntactic category, and the gender of the compound (Booij 2002, pp. 141–142). The examples in (31) illustrate that the compound gender is the same as the gender of the right member (the head), as indicated by the preceding definite article.

(31)   **Right-headed nominal compounds in Dutch** (Booij 2002, p. 142)

Common gender		Neuter gender	
de soep	'the soup'	het vlees	'the meat'
de vlees + soup	'the meat soup'	het soep + vlees	'the soup meat'

---

[8]There might be dialectal or generational differences in how productive this constructions is, especially since it is rather new. Banti (p.c.) informs me that the use with masculine nouns became increasingly common during Siyaad Barre's time (Siyaad Barre was the president of Somalia from 1969 to 1991).

Below, I will define properties of headedness that will be relevant to the Somali system when capturing the empirical generalizations regarding how gender functions in compounds. Before doing so, some background is in order, as the notion *head* has a long and somewhat messy history within linguistics. The concept and its defining properties as well as its status in different theories of both syntax, morphology, and phonology have been the topic of many studies: see e.g. Corbett, Fraser, and McGlashan (1993), Hudson (1987), and Zwicky (1985). A recent view that captures the basic intuition of what it means to be a *head* is illustrated in the quote in (32).

(32)   "Headedness is a pervasive phenomenon throughout different components of the grammar, which fundamentally encodes an asymmetry between two or more items, such that one is in some sense more important than the other(s)" (Moskal and Smith 2019)

In the Dutch example above, this asymmetry is manifested by several independent properties (here the determination of semantic class, syntactic category, and gender) which line up to point to a single member of a compound. The appeal with the notion thus seems to be that it can unify different properties, and the term *head* can be useful for referring to a descriptive category (in the sense of Haspelmath 2018) in a given language when this is the case. In the case of compounds, the criteria for headedness do not line up for all compounds or all languages (Guevara and Scalise 2009; Scalise and Fábregas 2010). More broadly, the idea that the notion "head" extends from syntax to morphology is controversial (e.g. Bauer 1990; Bauer 2017; Enger 1995; Fábregas and Masini 2015; Moskal and Smith 2019; Zwicky 1985), and part of the reason for this is that the independent criteria do not always converge, be it across languages or within one and the same language. Moskal and Smith (2019) argue the following:

(33)   "Whilst it seems clear for some phenomena such as category determination that words have heads, there is a lot of evidence to suggest that the properties of complex words are not all derived from one morpheme, but rather the features are gathered from potentially numerous morphemes within the same word" (Moskal and Smith 2019)

Furthermore, *head* is not a pre-theoretic concept: examples of authors that have argued against the notion and developed alternatives are Langacker (1987), working withing the framework of Cognitive Linguistics, and Croft (2001), working within Radical Construction Grammar.

For compounds in particular, there are reasons to treat headedness with caution: Guevara and Scalise (2009, p. 124) argue that most "false universals" (e.g. *the right-hand head rule* of Williams 1981, p. 248) in compounding are related to the notion *head,* and further add the following:

(34)   **Headedness and false universals**

"(W)hat we would call 'false universals' have come into being by the overwhelming amount of research on compounding in the Germanic languages, often driving researchers to apply automatically, and hastily, to every language the structural, semantic and pragmatic categories found in the most studied Germanic compound-types."
(Guevara and Scalise 2009, p. 122)

A similar frustration is expressed by Bauer (35).

(35)   "Discussion of headedness in compounds has tended to be restricted to the distinction between examples like *windmill* and examples like *egghead*. The first has been said to be headed, the second not to be. [...] We [...] need to look at more criteria, [and] we need to look at more compound-types. These must not only include types that are found in English." (Bauer 2017, p. 35)

The idea that some compounds lack heads, but others do not, has been captured by the distinction between *endocentric* (one of the members is the head) and *exocentric* (the head is something external – for example, a *red-head* is not a *head*, but a person with a read head/hair) (see Bloomfield 1933, pp. 235–237). However, there is an alternative view, as described by Bauer (36).

(36)   "[A]ll those compounds which have been called exocentric are regular compounds which are interpreted according to some figure of speech, so that *egghead* is a case of synecdoche, *fire-dog* is a case of metaphor and *phone-neck* ('a pain caused by over-use of the phone') is a case of metonymy" (Bauer 2017, p. 37)

See Bauer (2008, 2017) and Pepper (2010, 2020) for more on this view. Even in cases where the relationship between the two members is a clear case of hyponymy (for example, a *windmill* is a type of *mill*), the exact nature of the relation is *underspecified*, and many relations are possible. The set in (37), from Jackendoff (2016), illustrates this point. See also e.g. Eiesland 2015; Pepper 2020, ch. 6.

(37) **Semantic relations** (Jackendoff 2016, p. 16)
    chocolate cake   'a cake made with chocolate in it'
    birthday cake   'a cake to be eaten as part of celebrating a birthday'
    coffee cake   'a cake to be eaten along with coffee and the like'
    marble cake   'a cake that resembles marble'
    layer cake   'a cake formed in multiple layers'
    cupcake   'a little cake made in a cup'
    urinal cake   'a (nonedible) cake to be placed in a urinal'

For the purposes of the present paper, I will distinguish between three phenomena with different properties: *semantic heads, morphosyntactic heads,* and *morphosyntactic locus,* as defined in (38)–(40). These are intended to capture the Somali facts – see below for justification.

(38) **The semantic head**

    a. Definition: "the most contentful item that most closely denotes the same kind of thing that the whole constituent denotes" (Croft 2001, p. 259)

    b. Criterion: the 'IS A' criterion (Allen 1978) – for example, a *mailman* IS A *man*

    c. Additional assumptions:

      i. The semantic head does not need to be the *hyperonym* of the compound. That is, I do not analyze *egghead* as "exocentric", but assume the alternative view in (36) above.

      ii. The distinction between compounds in terms of their semantic relations cannot be reduced to one of *idiomatic* and *non-idiomatic*. The semantic relation is underspecified, and many relations are possible, as we saw in (37).

(39) **The morphosyntactic head**

    a. Definition: the member which determines the morphosyntactic properties of the compound; here, only gender is relevant.

    b. Criterion: the gender of the morphosyntactic head allows one to predict the form of agreement targets (target gender).

    c. Additional assumption: derivational affixes can also allow us to predict the behavior of agreement targets; they are therefore morphosyntactic heads in this sense.

(40) **The morphosyntactic locus**

    a. Definition: the member to which derivational and inflectional affixes attach.

    b. Additional assumption: this is an edge phenomenon (Bauer 2017, p. 31) rather than a type of head. It is a matter of location of potential affixes, and therefore a

property of a given system or construction rather than a fact about a member in a given compound.

There are some issues with these definitions which should be commented on, but it will be argued that these issues do not apply to the compounds included in the present study: First, the definition in (38) does not work for coordinative compounds such as *producer-director* (because this is *both* a producer and a director). Hence, this criterion do not apply to all compounds – but it does apply to all compounds in the present study (see the appendix). Second, the definition in (39) does not work for compounds in languages which do not have grammatical gender. Hence, this criterion does not apply to all languages and it is not a good comparative concept in the sense of Haspelmath (2018), but it does apply to Somali, which has a gender system. Third, the definition in (40) only applies to languages with derivational and inflection morphology. It furthermore excludes non-concatenative morphology and circumfixes. Somali is largely suffixing, which is what matters for the present purposes. These issues illustrate some of the problems for a general theory of headedness or cross-linguistic comparisons of compounds (see also Bauer 2017; Pepper 2020; Scalise and Fábregas 2010), but for the present purposes, these problems do not apply, and the definitions and criteria in (38)–(40) will allow us to capture the generalizations within the Somali system.

2.5   Gender and headedness in compounds in Somali

Some issues with the notion headedness were outlined above – there is a need for more in-depth research on compounds, both typologically and in particular languages (Guevara and Scalise 2009; Scalise and Fábregas 2010). The present study contributes to this body of literature, and one of the aspects is sorting out why conflicting generalizations have been proposed regarding both gender and headedness in Somali, as reflected in the quotes in (41).

(41)   **Headedness in Somali compounds**

    a. "From a syntactic standpoint, [NN-compounds] are [...] right-headed. The grammatical gender of the resulting compound is that of the final constituent."
(Green and Morrison 2016, p. 15)

    b. "Its gender is determined by the head of the compound. That means that we can easily find a sequence N_1 + N_2 + Art where N_1, the head of the compound, is masculine, N_2 is feminine and the article is masculine, or viceversa."
(Puglielli 1989, p. 7)

Note that it is not clear what type of "head" is referred to by these scholars. Furthermore, Banti and Jama (2016) say that NN_N compounds are left-headed, and NV_N compounds (in which the V

is a deverbal noun) are right-headed, and Saeed (1999, pp. 156–157) says that NN$_N$ compounds can be both left- and right-headed. One possible reason for the discrepancies, is that the term *head* is not defined explicitly in any of these sources. Section 2.4 illustrated that when doing descriptive, theoretical or typological work on compounds, there is a need to keep the criteria for headedness distinct from each other, and to provide clear definitions. When describing how gender is assigned to compounds, one may ask whether the compound gender consistently is determined by either the initial or the final member of a compound – that is, whether position in the structure is what allows a member to be the morphosyntactic head. The answer to this may or may not line up with what the semantic head of the compound is. Judging from the examples in (42)–(43), collected from a dictionary, it is not the case that gender in Somali compounds is determined consistently by the initial or the final member. In (42), the definite article matches the gender of the final member even though the semantic head is the initial member. In (43), the definite article matches the gender of the initial member.

(42) **Gender of compound = gender of final member**

laf-ta	F	+	garab-ka	**M**	→	lafgarab-ka		**M**
'bone'			'shoulder'			'shoulder bone'		
buug-ga	M	+	lacag-ta	**F**	→	buuglacag-ta		**F**
'book'			'money'			'cashbook, ledger'		
						(Zorc and Osman 1993)		

(43) **Gender of compound = gender of initial member** (Zorc and Osman 1993)

laf-ta	**F**	+	dhabar-ka	M	→	lafdhabar-ta		**F**
'bone'			'back'			'spine'		
bur-ka	**M**	+	saliid-da	F	→	bursaliid-ka		**M**
'flour/cake'			'oil'			'fritter'		
						(Zorc and Osman 1993)		

A further issue in the Somali case, is that different genders are reported by different scholars: For example, *birdanab* has variously been reported as feminine and masculine (44).

(44) **Different genders reported**
     bir (F) + *danab* (M) → *birdanab* 'magnet'
     F (Saeed 1999, p. 158) or **M** (Zorc and Osman 1993, p. 41)

Similar discrepancies are found for the compounds in which the final member has one of the suffixes *-eed/-aad-/-ood* (see section 2.3). The examples in (45)–(46), illustrating feminine + feminine-

eed compounds, are collected from a single dictionary, and while the first one is reported as masculine, the second is reported as feminine.

(45)   *hawl* (**F**) 'labor' + *gacan* (**F**) 'hand' + *eed* →   *hawl + gacmeed-ka* 'manual labor' (**M**)
       (Zorc and Osman 1993)

(46)   *abeeso* (**F**) 'snake sp.' + *bad* (**F**) 'sea' + *eed* →   *abeesa + badeed-da* 'moray, eel' (**F**)
       (Zorc and Osman 1993)

To summarize, the semantic head, the morphosyntactic head, and the morphosyntactic locus (as defined in section 2.4) in compounds in Somali are as listed in (47).

(47)   **Properties of noun-noun compounds in Somali**
       a. Semantic head: the left member
       b. Morphosyntactic head: unclear
       c. Morphosyntactic locus: right edge

The status of the morphosyntactic head is unclear given the previous literature on Somali. It seems to be variable, and one may ask what the reason for the discrepancies are. In the examples presented here, the only cue to compound gender is the definite article. In the present paper, I will argue that there is a need for independent evidence of compound gender from other agreement targets than the article. In the next section, patterns of agreement on the definite article is compared to those of subject-verb agreement. It will be argued that in cases of gender conflict between the semantics and the formal properties of compounds (their tone patterns, or the presence of a suffix such as *-eed*), Somali compounds act as *hybrid controllers*: first, they control different gender values on different agreement targets (article or verb). Second, a single target (most notably the definite article) may show variable agreement.

3   Gender agreement with noun-noun compounds

In order to examine the gender of compounds in more detail, noun-noun compounds were elicited from a native speaker of Somali living in San Diego. She is originally from Mustahil, Ethiopia, has lived in multiple places in the Horn of Africa, and speaks a variety of Somali best characterized as an instance of *Northern-Central Somali* (in the sense used by Tosco 2012).[9] All the remaining examples are provided by this speaker unless otherwise noted. The compounds

---

[9]This corresponds roughly to the dialect group referred to by Somalis as *maxaa-tiri*, and the spoken *koiné* that developed through dialect mixing. See Banti (2009) and Tosco (2012) for more on Somali dialectology.

and their members were produced in carrier phrases of the type *the X is good* or *the X is bad.* These phrases require both a choice of definite article (masculine *ku* and feminine *tu*)[10] and copula verb agreement (masculine *yahay*, feminine *tahay*). The importance of getting independent evidence of a noun's gender from an agreement target, such as the verb, was explained in section 2.1: there are reasons to believe that the definite article on its own is not a reliable cue to gender in Somali, and as we will see below, there are indeed occasional mismatches between the gender indicated by the definite article and the gender indicated through subject-verb agreement.

The present section is organized as follows. In section 3.1, details about the word list is provided. In sections 3.2–3.3, agreement patterns with noun-noun root compounds are presented. Elaboration on the definite article assignment is provided in section 3.4, followed by a discussion of tone patterns in section 3.5. Agreement patterns with compounds ending in one of the suffixes *-eed/-aad/-ood* are presented in section 3.6. Finally, the patterns are summarized in section 3.7.

3.1   The word list

A list of 86 noun-noun root compounds, as well as 38 compounds with an attributive suffix (*-eed/-aad/-ood*) were compiled based on a list of 1000 Somali compounds courtesy of Morgan Nilsson and another list of 1400 compounds courtesy of Christopher Green, as well as dictionaries (Mansuur and Puglielli 2012; Zorc and Osman 1993), Saeed's (1999) grammar, Caney's (1984) work on novel words in the Somali vocabulary, and personal field notes. For reasons explained in section 2.3, compounds with derivational suffixes, nouns ending in *-e* or *-o*, or any deverbal nouns or heterosemous roots were left aside for now and are not included in the list.

Certain other types of compounds were left aside in the present study because they have semantic or grammatical properties that potentially may override any structural gender assignment patterns in noun-noun compounds that the present study aims to uncover: first, compounds with a human referent, as the one in (48a), may show semantic agreement based on the gender (sex) of the referent.[11] Second, and for the same reason, compounds in which one of the members has a human referent were excluded. This include names of languages, such as the one in

---

[10]These are the nominative case forms of the definite articles, which are *ka* and *ta* in their citation forms – Somali has a case system of the type *marked nominative*, which is common in Cushitic languages (Mous 2012).

[11]An example of this in another language is the Italian compound *testa rasata* 'skinhead (lit. head shaved)': the initial member *testa* 'head' is feminine and controls feminine agreement on the second member *rasata* 'shaved'. However, the compound controls masculine agreement on the determiner, which in this case will be *il*, if the referent is male: *il testa rasata* (Scalise and Fábregas 2010, p. 122).

(48b) (*Soomaali* means 'a Somali person'). Third, compounds listed as collective nouns (48c) were excluded because such nouns have different agreement patterns (see e.g. Nilsson 2016b). Fourth, compounds in which one or both members are plural nouns, as in (48c)–(48d), were excluded because gender agreement with plural nouns functions differently than with singular nouns (see section 2.1), which makes it difficult to compare these patterns to the agreement patterns with the compound as a whole. For the same reason, compounds in which one of the members is a numeral, as in (48e), were left out: Numerals are analyzed as nouns in Somali by syntactic and morphological criteria (see Saeed 1999, p. 69) – however, numerals above one control plural agreement, and there are no gender distinctions in the plural. Finally, compounds in which one of the members is a color term (48f) were left out because while some color terms behave as nouns in Somali, it proved difficult to elicit them in a carrier phrase in which they control subject-verb agreement.

(48)   **Types of compounds that were left out of the present study**
    a.   *af-gaab* 'quiet, polite person (lit. language/mouth-shortness)'
    b.   *Af-Soomaali* 'Somali (lit. language/mouth-Somali)'
    c.   *askar + maroodi* 'poor young boys (lit. soldiers-elephant)'
    d.   *lafo + lafo* (*lafallafo*) 'skeleton, skinny animal/person (lit. bones-bones)'
    e.   *afar + gees* 'square (lit. four-side)'
    f.   *cir + guduud,* 'the times of the day in which the sky is red (lit. sky-red)'

The number of compounds that were recognized by and successfully elicited from the speaker consulted in the present study, was 35 noun-noun root compounds (out of 86), and 23 compounds with an attributive suffix (out of 38). The complete list of these compounds, including notes on their agreement patterns, is provided in the appendix. A note on productivity is in order here: Even when the strict selectional criteria described above are taken into account, the proportion of noun + noun root compounds in the sources listed above seem to be rather small. Furthermore, for some of the items that are listed as compounds in these sources, the speaker produced a phrasal construction (e.g. *búug-ga lacág-ta* 'the book of money' for *buug + lacag* 'cash-book'), and did not accept the corresponding compound form. Although Saeed (1999, p. 154) describes compounding in general in Somali as a productive process, it seems that noun-noun root compounding in particular may be less so.

3.2   Compounds in which the members have the same gender

The first question to ask is whether the gender of the members play any role at all in determining compound gender, or there is some other unrelated mechanism which assigns gender to compounds. Based on the following examples, the gender of the members do in fact seem to be the determining factor: When both members are masculine, the compound controls masculine agreement, as in (49), and when both members are feminine, the compound controls feminine agreement, as in (50).

(49)   **M+M compounds = M**

    a. Gaari+fáras-ku   waa   fiicán yahay.
        car+horse-**M**.DEF.NOM DECL good 3SG.**M**.COP.PRS

        'The horse cart is good.'   cf. *gaarígu (M), fárasku (M)*

    b. Fool+maroodí-gu   waa   fiicán yahay.
        tooth+elephant-**M**.DEF.NOM DECL good 3SG.**M**.COP.PRS

        'The ivory/elephant's tusk is good.'   cf. *fóolku (M), maroodígu (M)*

(50)   **F+F compounds = F**

    a. Laf+aráx-du   waa   fiicán tahay.
        bone+vertebra-**F**.DEF.NOM DECL good 3SG.**F**.COP.PRS

        'The spine is good.'   cf. *láftu (F), aráxdu (F)*

    b. Kabsar+caleén-tu   waa   fiicán tahay.
        herb+leaf-**F**.DEF.NOM DECL good 3SG.**F**.COP.PRS

        'The parsley/cilantro is good.'   cf. *kabsártu (F), caleéntu (F)*

Compound gender is thus not independent of the gender of its members. However, when the two members have the same gender, one cannot determine which one determines compound gender. The next section addresses such compounds.

3.3   Compounds in which the members have different gender

When the members of a compound have different genders, the agreement patterns are usually predictable from the gender of the *initial* member. That is, for this speaker, masculine-feminine compounds control masculine agreement (51), and most feminine-masculine compounds control feminine agreement (52). In these examples, both the definite article and the verb agreement match the gender of the initial member, even though the definite article is linearly adjacent to the second member.

(51) **M + F compounds = M**

    a. Bur + saliíd-ku   waa   fiicán yahay.
        cake/flour + oil-**M.DEF.NOM** DECL good 3SG.**M**.COP.PRS

        'The fritter is good.'   cf. *búrku (M)*, *saliíddu (F)*

    b. Daacuun + caloól-ku   waa xún yahay.
        cholera + stomach-**M.DEF.NOM** DECL bad 3SG.**M**.COP.PRS

        'The cholera/stomach disease is bad.'   cf. *daacúunku (M)*, *caloóshu (F)*

(52) **F + M compounds = F**

    a. Shimbir + málab-tu   waa   fiicán tahay.
        bird + honey-**F.DEF.NOM** DECL good 3SG.**F**.COP.PRS

        'The honey-bird is good.'   cf. *shimbírtu (F)*, *málabku (M)*

    b. Laf + gárab-tu   waa   fiicán tahay.
         bone + shoulder-**F.DEF.NOM** DECL good 3SG.**F**.COP.PRS

        'The shoulder bone is good.'   cf. *láftu (F)*, *gárabku (M)*

In other cases, the choice of definite article and the verb agreement are in conflict (53). Subject-verb agreement matches the gender of the initial member, but unlike the examples in (51) and (52), the definite article does not. Instead, it matches the gender of the second member.

(53) **F + M compounds with conflicting gender agreement on definite article and verb**

    a. Wadda + hálaq-u   waa   fiicán tahay
        road + snake-**M.DEF.NOM** DECL good 3SG.**F**.COP.PRS

        'The way snakes go/move is good.'   cf. *waddádu (F)*, *hálaqu (M)*

    b. Fool + díin-ku   waa xún tahay
        labor.pain + turtle.**M.DEF.NOM** DECL bad 3SG.**F**-COP.PRS

        'The false labor pains are bad.'   cf. *foóshu(F)*, *díinku (M)*

One might have expected *wadda + halaq-du* (F) and *fool + diin-tu* (F) here, because of the feminine agreement on the verb. In the case of *fool + diin-ku*, there may be a separate reason for why the masculine article is assigned instead of the feminine: the final member of the compound, the masculine noun *díin-ka* 'the turtle', forms a minimal pair with a feminine noun: *diín-ta* 'the religion'. In this particular case, the form *fool + diin-ta* may be blocked to avoid the potential interpretation *labor.pain + religion*. However, no such explanation is available for *wadda + halaq* to my knowledge.

The speaker sometimes chooses different definite articles on different instances, even for the same compound (54). This variation is not attested in subject-verb agreement, which instead is consistently predictable from the gender of the initial member.

23

(54) **Variable definite article, but consistent subject-verb agreement**

    a. Laf + sakáar -ku   ∼ -tu   waa fiicán tahay
           bone + chest -**M**.DEF.NOM ∼ -**F**.DEF.NOM DECL good 3SG.**F**.COP.PRS
           'The breast bone/sternum is good.'   cf. *láftu (F), sakáarku (M)*

    b. Bir + jíir   -ku   ∼ -tu   waa fiicán tahay
           iron + mouse -**M**.DEF.NOM ∼ -**F**.DEF.NOM DECL good 3SG.**F**.COP.PRS
           'The mouse trap is good.'   cf. *bírtu (F), jíirku (M)*

Variable definite article assignment is only attested with F + M compounds in the present study, not with M + F compounds. However, there are few M + F root compounds to begin with, for reasons that were explained in section 2.3. For the present purposes, I assume that agreement may potentially vary for M + F compounds too, and that it is unattested in the present study simply due to the small sample size of M + F compounds. This point will be taken up again in section 3.6.

Since subject-verb agreement is consistent, but definite article assignment is variable, there are occasional mismatches in agreement on the article and the verb. Notice also that although the definite article is added on the *right* edge (recall from section 2.5 that the final member is the morphosyntactic locus), it is the *initial* member which is the semantic head (either metaphorically or hyperonymically – see section 2.4) in all of these cases. For example, *lafsakaar* 'breast bone' (54a) is a type of *laf* 'bone', hence *laf* (F) is the semantic head. *Fooldiin* 'false labor pains (turtle-labor pains)' (53b) is metaphorically a type of *fool* 'labor pains'.[12]

The generalization that holds regardless of the type of semantic relationship, is that subject-verb agreement is predictable from the gender of the *initial* member of the compound, while the definite article is free to vary. Mismatches in agreement patterns are found for compounds across different semantic relations.

3.4　Definite article assignment

While subject-verb agreement suggests that it is the *initial* member of noun-noun compounds that determines compound gender, compounds control different gender values on the definite article. There seems to be two competing generalizations for definite article assignment: one based on linear adjacency, in which the definite article reflects the gender of the closest member (55a), and another based on left-headedness, in which it reflects the gender of the semantic head (55b).

---

[12]The explanation I was provided for *fool-diin*, is that turtles are slow. The idea is that after a false labor, it can still be a long time until the baby comes.

(55)   **Competing generalizations**

    a.   LINEAR ADJACENCY

       Wadda + **hungúri-gu**   waa   fiicán tahay.
       road + throat-**M**.DEF.NOM DECL good 3SG.**F**.COP.PRS

       'The esophagus (the throat-road) is good.'   cf. *waddádu (F), hungúrigu (M)*

    b.   LEFT-HEADEDNESS

       Wadda + hungúri-**du**   waa   fiicán tahay.
       road + throat-**F**.DEF.NOM DECL good 3SG.**F**.COP.PRS

       'The esophagus (the throat-road) is good.'   cf. *waddádu (F), hungúrigu (M)*

A possible analysis for these data is that the linear adjacency strategy involves the *internal* assignment of a definite article to the rightmost noun, as illustrated with the structure in (56). Another way of putting this, is that the assignment of *-gu* here reflects a relationship between the compound construction and the construction *hungúri-ga,* which exists outside of the compound. This strategy contrasts with the left-headedness strategy, which can be illustrated as in (57).

(56)   LINEAR ADJACENCY

    [F + M-gu]F

    [wadda + hunguri-**gu**]F

(57)   LEFT-HEADEDNESS

    [F + M]F -du

    [wadda + hunguri]F -**du**

The mismatches in gender cues found in examples such as (53) and (54) above are reminiscent of various other types of gender mismatches, such as those found with *hybrid nouns, conjunct agreement,* and *agreement attraction* or *proximity concord.* The Somali data will be discussed in light of these phenomena in section 4. Before turning to this, another type of gender mismatch in Somali is presented, namely one following from the relationship between definite article assignment and tonal gender marking.

3.5   Tone patterns

Another type of mismatch involving gender in compounds is found in their tone patterns. In *non*-compounds, there is a correlation between tone and gender (though not a perfect one), such that masculine nouns typically have a penultimate High tone, and feminine nouns typically have

a final High tone. This was explained in section 2.1, and the examples provided in (8)–(9) are repeated here in (58)–(59) (from Hyman 1981, p. 172).

(58) **Masculine nouns, penultimate H**

    a. *ínan* 'boy'

    b. *qaálin* 'young he-camel'

    c. *daméer* 'he-donkey'

(59) **Feminine nouns, final H**

    a. *inán* 'girl'

    b. *qaalín* 'young she-camel'

    c. *dameér* 'she-donkey'

When two nouns are compounded, only the final member retains its High tone, and the location of the tone does not shift, no matter what the gender of the compound is. That is, while the gender of the initial member determines subject-verb agreement, the final member determines the tone pattern, and in M + F and F + M compounds, one therefore gets mismatches between tone and gender, as illustrated in (60)–(61).

(60) **Masculine semantic head, but feminine tone pattern**
    búr (M) + saliíd (F) → *bursaliíd* 'fritter' (**M**)

(61) **Feminine semantic head, but masculine tone pattern**
    láf (F) + gárab (M) → *lafgárab* 'shoulder bone' (**F**)

What could have happened in this case, is that a new High tone were assigned to the compound based on the gender of the semantic head to the left (rendering *bursalíid* and *lafgaráb*), but this is not the case.[13] As explained in section 2.2, I view tone as a lexical property of nouns and a property of their inflectional class. *Laf + gárab* is an example of a compound in which the final member is a declension 1a noun: *gárab* 'shoulder'. In (49b) above, we saw an example of a compound (*fool + maroodí* 'elephant's tusk') in which the final member is a declension 3 noun: *maroodí* 'elephant'. In both cases, the final member is masculine, but the former has a penultimate High tone, and the latter has a final High tone. The tone pattern of the compound thus reflects the declension of the final member, not the gender. The generalizations are thus as illustrated in (62).

(62) **Tone in noun-noun compounds**

    a. In nominal compounds, only the final member bears a High tone.

       Examples: *laf + gárab* 'shoulder bone', *fool + maroodí* 'elephant's tusk'

---

[13]This has been reported by Banti (2016) for nominal noun-adjective compounds such as *bád* 'sea' (F) + *wéyn* 'big' → *badweýn* 'ocean' (F): while *wéyn* has a penultimate High tone, *badweýn* has a final High tone, in line with its gender (feminine). The role that tone plays in marking gender on compounds thus seems to depend on the grammatical properties of the compound and its members (such as word class). Note that there seems to be variation between speakers with respect to the tone pattern in this particular compound: *badweýn* ∼ *badwéyn* (personal field notes).

  b. In noun-noun root compounds, the location of that High tone is determined by the declension of the final member.
   Examples: *gárab* 'shoulder' (D1a), *maroodí* 'elephant' (D3)

The goal of the present paper is to capture the empirical generalizations relevant for how gender functions in compounds in Somali, and not to provide a formal analysis of the tone patterns (see Downing and Nilsson 2019; Green and Morrison 2016; Hyman 1981; Lampitelli 2013; Le Gac 2016 for a variety of proposals couched within different frameworks). However, some thoughts are in order before I get back to the relationship between the tone patterns and the agreement variation outlined above.

For the present purposes, I will assume that the location of the High tone in a noun-noun root compound is motivated by the relationship between the compound on the one hand, and the second member as it occurs outside of the compound context on the other. This is not just a property of a particular compound, but also a part of the more abstract compounding schema in Somali. Such a relationship would be in line with the idea of constructional schemas as envisioned by e.g. Bybee (2003) or Booij (2010a,b) (taking a usage-based perspective), rather than the result of an operation or process of tone assignment. The fact that there is only one High tone on compounds is under this view not the result of tone deletion on the initial member (or the result of the properties of proposed universal prosodic domains – see Green and Morrison 2016 for such a formalization of Somali compounds using recursive Prosodic Words, and Downing and Nilsson 2019 for a formalization appealing to a Complex Word Group, a hypothesized domain between the Prosodic Word and the Phonological Phrase. For arguments against this approach to tone in Somali, see Kaldhol and Stausland Johnsen forthcoming). Rather, it is a property of the morphological construction, and a property which creates a contrast with this construction and the noun-noun genitive construction, in which both nouns have a High tone (as explained in section 2.3). Downing and Nilsson (2019) report that at least for some Somali speakers, the second noun in noun-noun genitive constructions is toneless, contrary to what is expected based on previous literature (see e.g. Saeed 1999). They formalize this by appealing to "prosodic restructuring" – however, the tone patterns are still predictable from construction type, as the examples in (63) illustrate.[14]

(63) **Noun-noun root compounds vs. noun-noun genitive constructions**
  a. guri + márti
   house + guest
   'guest house'

---

[14]Note that *márti* 'guest', like other masculine nouns, exhibits tonal genitive marking: *martí*. An alternative analysis is that this in fact is the collective noun *martí* 'guests'. See Downing and Nilsson (2019).

b. gúri martí ∼ gúri marti
   house guest.GEN house guest

   'house of guest'   (Downing and Nilsson 2019, p. 139)

As for the previous point that the location of the High tone in a noun-noun root compound is determined by the declension of the second noun: this results in a tension between the form of the compound (the inflection class indicated by the second member) and the gender of the semantic head (the initial member). One could have expected variable tone patterns as result of this tension, but this is not attested in the present study. The tone patterns of noun-noun compounds always correspond to the tone pattern of the final noun (as in *gárab – laf + gárab*). Instead, there is variable definite article assignment, such that the article either matches the form of the final member (its declension, and therefore its gender), as in (64), or the gender of the semantic head, as in (65).

(64) **Match between tonal gender marking and article** (masculine *-ku*, penult H)

  [F + M-ku]F

  laf + gárab-ku

(65) **Mismatch between tonal gender marking and article** (feminine *-tu*, penult H)

  [F + M]F -tu

  laf + gárab-tu

Like the tone pattern, the linear adjacency strategy for definite article assignment (64) can be captured by a relationship between the compound on the one hand, and the final noun outside of the compound context on the other (*gárab-ku*). While the compound gender strategy for definite article assignment (65) results in a mismatch between the definite article and the tonal gender marking, it will, as we have seen, result in a match between the article and subject-verb agreement, which does not vary (66).

(66) **Matching agreement, mismatching tone**

  Laf + gárab-tu   waa fiicán tahay.
  bone + shoulder-**F.DEF.NOM** **DECL** good 3SG.**F**.COP.PRS

  'The shoulder bone is good.'   cf. *láftu (F), gárabku (M)*

That is, in M + F and F + M compounds, there will either be a mismatch between definite article assignment and tonal gender marking, or a mismatch between definite article assignment and the gender of the semantic head (and the subject-verb agreement), and the attested variation potentially reflects choosing between the two types of mismatches. One could in theory have

expected all the gender cues to line up, e.g. by assigning a High tone to the compound that reflects the gender of the semantic head, and have both the definite article and the subject-verb agreement be controlled by this gender. Instead, there is a restriction against changing the location of the High tone within a root, so that this does not happen. This restriction applies to noun roots in general, not just within compounds: tonal shift within a root is a marginal phenomenon in Somali, and restricted to certain forms of declension 2 nouns (the ones ending in *-e/-o*, see section 2.2), which were left out of this study. These could furthermore be analyzed as morphologically complex, as the vowels *-e/-o* sometimes act as derivational suffixes, sometimes as thematic vowels – see Le Gac (2016, p. 295) for discussion. One could alternatively expect all agreement targets to reflect the gender indicated by the phonological form of the compound (that is, its tone pattern). Instead, the situation of having the semantic head to the left, and tone pattern determined by the final member, leads to mismatches in gender cues, and variable agreement patterns.

3.6   Compounds with the suffixes -eed/-aad/-ood

The mismatches in gender agreement that have been discussed so far involve feminine-masculine compounds. Masculine-feminine compounds produced by this speaker consistently get a masculine definite article, resulting in a mismatch between article and tone pattern (as in *bur + saliíd-ka* 'the fritter'). No mismatch in article and subject-verb agreement is attested with such compounds in the present data set. It is possible that the reason for this simply is that there are very few masculine-feminine compounds to begin with (in fact, only two were successfully elicited, namely *bur + saliíd* 'fritter', and *daacuun + caloól* 'cholera'), and hence the likelihood of variation is lower. In fact, masculine-feminine compounds that are assigned a feminine article are attested in other sources, e.g. with *buug + lacag-ta* 'the cashbook' (Zorc and Osman 1993) (the speaker in the present study used the construction *búug-ga lacág-ta* instead of a compound for this particular item).

There is a reason for the small number of masculine-feminine compounds (compared to feminine-masculine compounds) in which both members are roots: as explained in section 2.3, compounds in which the final member is feminine typically has one of the suffixes *-eed/-aad/-ood*. At least historically, these were genitive suffixes used with feminine nouns (masculine nouns show tonal genitive marking), but synchronically, they rather seem to derive attributive forms of nouns (see Banti 1986, 1988a; Nilsson 2016a, 2019). A list of compounds with this suffix was included in the present study for comparison to root compounds. As the examples in (67) illustrate, agreement with masculine-feminine-*eed* compounds is masculine across the board.

However, agreement with feminine-feminine-*eed* compounds can be either feminine or masculine, and both the definite article and the verb shows variation (68).

(67) **M+F-eed compounds – masculine agreement**

    a. Dayax+gacméed-ku   waa wéyn yahay.
        moon+hand.ATTR-**M**.DEF.NOM DECL big   3SG.**M**.COP.PRS

        'The satellite is big.'   cf. *dáyaxu (M), gacántu (F)*

    b. Libaax+bad-éed-ku   waa xún yahay.
        lion+sea-ATTR-**M**.DEF.NOM DECL bad 3SG.**M**.COP.PRS

        'The shark is bad.'   cf. *libáaxu (M), báddu (F)*

(68) **F+F-eed compounds – variable agreement**

    a. Shax+taariikh-éed-ku   waa wéyn tahay.
        checkers+date-ATTR-**M**.DEF.NOM DECL big   3SG.**F**.COP.PRS

        'The grid calendar is big.'   cf. *sháxdu (F), taariíkhdu (F)*

    b. Jirrid+maskax-éed-ku   waa wéyn yahay.
        stem+brain-ATTR-**M**.DEF.NOM DECL big   3SG.**M**.COP.PRS

        'The brain stem is big.'   cf. *jirríddu (F), maskáxdu (F)*

    c. Hees+carruur-éed-du   waa fiicán tahay.
        song+children-ATTR-**F**.DEF.NOM DECL good 3SG.**F**.COP.PRS

        'The children's song is nice.'   cf. *heéstu (F), carruúrtu (F)*

To unify the agreement patterns illustrated here with the ones found with root compounds, I propose the following analysis: First, these suffixes assign gender, like derivational suffixes in general in Somali (see section 2.1), and they are therefore morphosyntactic heads (in the sense defined in section 2.4). More specifically, they assign masculine gender, since agreement can be masculine even in cases in which both noun roots are feminine (68). They also have a penultimate High tone, and most nouns with a penultimate High tone are masculine nouns of declension 1a.

Second, as illustrated with root compounds in section 3.3, the definite article will either agree with the semantic head of a compound (the initial member), or the element to which it is linearly adjacent. In (67)–(68), that element is not a root noun, but a derived noun: [[Root]$_F$ *eed*]$_M$. In (67), both the semantic head and the suffix (or the derived noun)[15] are masculine. There is no gender conflict, and no variation in agreement. In (68), the semantic head is feminine (in fact, both noun roots are feminine), while the suffix is masculine.

---

[15]Note that because nouns derived with *-eed* always have an attributive function, and therefore never function as the head of a subject noun phrase, one cannot independently establish their gender through subject-verb agreement, or, to my knowledge, in any other way.

There is thus a gender conflict, and the definite article shows variable agreement, just like with feminine-masculine root compounds.

Third, as argued in section 3.3, the verb agrees with the *head* of the compound. In noun-noun root compounds, there is only one candidate for a head, namely the semantic head. In compounds with an *-eed* suffix, the properties of headedness are split between the initial member (the semantic head) and the suffix, which acts like a morphosyntactic head in that is assigns gender. The verb can agree with either head-like element, and variation is attested when there is a gender mismatch between the two types of heads (68).

For the present data set, this variable agreement pattern is unattested with root compounds, which under this analysis only have one candidate for a head. There is a parallel to be drawn between these patterns and the variable agreement patterns found in other cases of gender mismatches across languages, and this point will be elaborated on in section 4. Note also that the possibility of the verb agreement to be controlled by a suffix at the right edge of a morphological construction does not come out of nowhere: the [*Noun* + [[*Noun*] -suffix]] construction discussed here coexists with a [[*Noun* + *Noun*] -suffix] construction, which has allowed for a competing generalization. This will be discussed further in section 5, in which I argue that a complete understanding of the variable agreement patterns found in the present study cannot come from looking at a single construction, but properties of the system as a whole.

3.7   Summary of patterns

The agreement patterns with noun-noun root compounds presented in this section can be summarized as follows: first, when the two members of a compound have the same gender, that gender determines the form of the agreement targets, which in the present study is the definite article and the copula verb. Since such compounds have a consistent agreement pattern, they can be considered to be consistent controllers. Second, when the two members have different genders (M + F or F + M), the compound controls different feature values on different targets: while the subject-verb agreement is predictable from the gender of the *initial* member of such a compound (which also is the semantic head), the choice of definite article varies, and can either reflect the gender of the initial member, or the final member. Such compounds can therefore be considered to be *hybrid controllers* (Corbett 2006, pp. 11–12), a point that will be taken up again below.

We also saw that there may be mismatches between the gender indicated by the tone pattern of compounds, and the gender indicated by the agreement: in a noun-noun root compound, only

the final member bears a tone (as in *laf* + *gárab* 'shoulder bone'), and the location of this tone reflects the tone pattern of that noun outside of the compound context (as in *gárab* 'shoulder'). There seems to be a restriction against shifting the location of a High tone within the root, so instead of reassigning a High tone to reflect the gender of the semantic head (the initial member), the tone pattern instead reflects the gender of the non-head. In cases where the gender of the semantic head determines the form of the agreement targets, there will thus be a mismatch between the gender indicated through agreement, and the gender indicated by the tone. Note that since the choice of definite article may vary, it can either match the gender indicated by the tone pattern (which in turn is determined by the gender of the final member), or the gender indicated by the semantic head. Conversely, there will either be a *mismatch* between the two agreement targets (article and verb), or between the article and the tone pattern. In table 4, the 16 logically possible combinations of these factors are listed, and the attested ones are indicated in the final column. The complete list of compounds is provided in the appendix.

Gender of members	Tone	Definite article	Subject/verb agreement	Attested?
M + F	Fin	M	M	**yes**
	Pen	M	M	no
	Fin	F	M	no
	Pen	F	M	no
	Fin	M	F	no
	Pen	M	F	no
	Fin	F	F	no
	Pen	F	F	no
F + M	Fin	M	M	no
	Pen	M	M	no
	Fin	F	M	no
	Pen	F	M	no
	Pen	M	F	**yes**
	Fin[16]	M	F	**yes**
	Pen	F	F	**yes**
	Fin[16]	F	F	**yes**

Table 4: Noun-noun root compounds

The agreement patterns found with *noun* + *noun-eed* compounds (section 3.6) are slightly different. In table 5, their logically possible agreement patterns are listed. This table is half the size of table 4, because none of the *noun* + *noun-eed* compounds have a final High tone: the logically possible combinations in which there is a final High tone are left out of the table for

---

[16]The final High tone in these examples reflect the fact that the final member belongs to declension 3, a class of nouns that have a final High tone in their definite forms despite being masculine (e.g. *maroodí-ga* 'the elephant').

simplicity. As indicated in the table, *M + F-eed* compounds consistently take masculine agreement, while considerable variation is attested for *F + F-eed* compounds: there is variation even in subject-verb agreement, unlike what is attested in the present study for *root* compounds, for which only the definite article varies.

Gender of members	Tone	Definite article	Subject/verb agreement	Attested?
M + F-ATTR.(M)	Pen	M	M	**yes**
	Pen	F	M	no
	Pen	M	F	no
	Pen	F	F	no
F + F-ATTR.(M)	Pen	M	M	**yes**
	Pen	F	M	no
	Pen	M	F	**yes**
	Pen	F	F	**yes**

Table 5: Noun + noun-eed compounds

Based on these patterns, I argued that the *-eed/-aad/-ood* suffixes assign masculine gender. The definite article may either agree with the semantic head (the initial member), or the element to which it is linearly adjacent (the suffix). The verb agrees with the head, and in the case of *noun-noun-eed* compounds, the headedness properties are split between two elements: the initial member (the semantic head) and the final element, which is a suffix (the morphosyntactic head). Normally, suffixes are dominant in Somali in the sense that they determine properties such as word class, declension, gender and tone. They could arguably even be semantic heads, as in *edeb-dárro* 'lack of good manners (lit. good manners-lacking)'. However, the *-eed/-aad/-ood* suffixes only take scope over the final member of a compound, turning that noun into an attributive. The suffix has head-like properties, but the semantic head of the compound will still be the initial member. This split of headedness properties between different members may thus be the cause of the variation seen in its agreement patterns, both on the article and the verb.

For the *F + F-eed compounds* in table 5, there is one unattested pattern: One in which the definite article is feminine, but the subject-verb agreement is masculine. Given the current analysis, this is an accidental gap: if both the definite article and the subject-verb agreement is variable in this particular type of compound due to the split of headedness properties, then this type of pattern should be attested too. The fact that it is not is captured by the following generalization, which holds for all of the attested patterns in tables 4–5: If the definite article reflects the gender of the semantic head, then so does the subject-verb agreement.

To summarize, the distinction between noun-noun root compounds and noun + noun-*eed* compounds is as follows: noun-noun root compounds have their semantic head to the *left*. This is the only head in such compounds, and subject-verb agreement is consistently predictable from the gender of that member. One could propose that the semantic head is also the morphosyntactic head in these cases; alternatively, agreement follows from semantics. The definite article can vary since it either agrees with the semantic head, or the final element of the compound, to which it is linearly adjacent.

Noun + noun-eed compounds also have their semantic head to the left. However, the suffix -*eed* acts as a morphosyntactic head in that it can control gender agreement. Once again, the definite article can vary since it either agrees with the semantic head, or the final element of the compound, to which it is linearly adjacent. The difference is that the subject-verb agreement also varies, since it can agree with the semantic head (the initial member), or the morphosyntactic head (the suffix). One can propose the following structure for these compounds: [Noun [Noun-eed]]. The question, then, is how the gender of the suffix is accessed by the agreement controllers. In section 5, I will argue that this variation results from competing generalizations allowed for by the system as a whole, such as the competing schema based on morphological constructions of the following type: [[Noun [Noun]]suffix]. Such constructions, in which there is a derivational suffix on the right edge, which unlike -*eed* takes scope over the whole compound, will be explored further in section 5. The main point for now is that the pattern in which the subject-verb agreement is controlled by a suffix on the right edge (a fact about linear position, not hierarchical structure) does not come out of nowhere, but is motivated by a coexisting type of morphological construction in Somali.

4   Gender mismatches

The gender mismatches found with compounds in Somali with respect to both agreement patterns and tonal gender marking is reminiscent of other types of gender mismatches reported in the literature, such as those found with hybrid nouns, conjunct agreement, and agreement attraction. In the present section, these phenomena will be discussed in turn.

4.1   Hybrid nouns

The data presented here are reminiscent of *hybrid nouns*, that is, nouns whose agreement value varies according to agreement target (Corbett 1991, 2006; Enger and Corbett 2012). This

phenomenon is attested with different agreement features, most notably gender and number. An example of a number hybrid is the English noun *committee,* which refers to a group of people, and can show either singular agreement (*the committee has decided...*) or plural agreement (*the committee have decided...*). Often, there is a mismatch between the formal properties of the noun and the properties of its referent. For example, the lexical gender of a noun can sometimes be in conflict with the semantic gender of its referent. A well-cited example is the German word *mädchen* 'girl', which controls both neuter and feminine gender. Consider the examples in (69).

(69) **A German hybrid noun** (from Corbett 1991, p. 228)

    a. Das   Mädchen, das   ich gesehen habe...
        DEF.**N** girl   that.**N** I   seen   have
        'the girl I saw...'

    b. Schau dir dieses Mädchen an, wie gut   sie/es   Tennis spielt
        look you this.**N** girl   at how good 3SG.**F/N** tennis plays
        'Do look at this girl, see how well she plays tennis.'

Notice that all the agreement targets are neuter, except the pronoun, which shows variation: both feminine (*sie*) and neuter (*es*) are possible. Similar patterns are found with selected nouns in many languages, such as Czech *děvče* 'girl', Dutch *jongetje* 'little boy' and *vrouwtje* 'little woman', French *sentinelle* 'sentry, guard', Polish *ofiara* 'victim', Russian *vrač* '(female) doctor' (Corbett 1991, pp. 228–232; Corbett 2006, p. 218) and Norwegian *mamma* 'mother' (Enger and Corbett 2012). What these examples have in common, is that some agreement targets indicate the semantic gender of the referent (*semantic* or *referential* agreement), while others indicate the gender assigned by some other phonological or morphological property of the noun (*syntactic* or *formal* agreement): for example, while German words ending in the suffix *-chen* are formally Neuter, the referent of *mädchen* 'girl' is female, and this is arguably what allows for the variation in agreement patterns. Not only morphological form (such as the suffix *-chen* in German) matters, but also phonological form. This is illustrated by the Russian word *vrač* 'doctor' (70).

(70) **A Russian hybrid noun** (from Corbett 2006, p. 158)

    a. Ivanov-a   xoroš-ij   / xoroš-aja   vrač
        Ivanov-SG.NOM good-**M.SG.NOM** / good-**F.SG.NOM** doctor.SG.NOM
        'Ivanova is a good doctor.'

    b. Vrač   prišel   / prišla
        doctor.SG.NOM come.**PST.SG.M** / come.**PST.F.SG**
        'The (woman) doctor came.'

In Russian, nouns that end in a consonant are always masculine. Therefore, the phonological form of *vrač* would suggest that this is a masculine noun. However, its referent can be either a female or a male doctor. When the referent is female, as in (70), but the formal gender is masculine, there is a gender mismatch, and this results in variation in the gender indicated by different agreement targets. The target gender may either reflect the formal gender indicated by the phonological form (masculine), or the semantic gender of the referent (feminine). Of these two options, the first type of agreement has been called *syntactic* or *formal* agreement, while the second has been called *semantic* or *notional* agreement (Corbett 2006, p. 155).

Although none of the compounds in the present study have a human referent, there is a parallel to be drawn between hybrid nouns and Somali compounds with variable agreement patterns. The data presented in section 3.3 showed that subject-verb agreement is predictable from the gender of the initial member of the compound. This member is also the semantic head. Arguably, this is a type of semantic agreement based on referential gender, even though there is a conceptual difference between agreement with the semantic head, and agreement based on the gender of a human referent (or plural agreement with a singular noun denoting a group of people). The point is that the referent of *laf-gárab* 'shoulder bone' is a type of *láf* 'bone' (F), not a type of *gárab* 'shoulder' (M). The definite article may also show this type of semantic agreement and reflect the gender of the semantic head of the compound. However, as we have seen, it can also reflect the gender of the final member (the linear adjacency strategy), in which case it is in line with the inflectional class (and as a result, the gender) suggested by the tone pattern of the compound. One can think of the linear adjacency strategy as a type of formal agreement based on the gender cue provided by a phonological property of the compound.

This is comparable to what we saw with Russian *vrač* 'doctor' in (70) above, but while it is segmental phonology which is relevant in Russian (nouns ending in a consonant belong to declension 1, which consists of nouns that typically control masculine agreement), it is the tone pattern that is relevant in Somali: most nouns with a penultimate High tone belong to declension 1a, and control masculine agreement. In compounds, the final member keeps its lexical tone rather than shifting it to reflect the gender of the semantic head (the initial member). So the compound *laf + gárab* 'shoulder bone' has a penultimate High tone (rather than a final: **lafgaráb*) because the final member has a penultimate High tone (*gárab*). While the correlation between tone and gender in Somali is far from perfect, it is the case that very few feminine nouns ending in consonants have a penultimate High tone (see section 2.2). Although the initial member of a compound like *laf + gárab* belongs to declension 1b and is feminine, the form of the compound suggests that the compound belongs to declension 1a, like its final member, which controls masculine agreement.

There is of course a conceptual difference between agreement with a noun like *vrač*, in which lexical gender and referential gender differs, and agreement with a noun-noun compound, in which both members have their own lexical gender. In this respect, conjunct agreement may be enlightening (note that conjunct agreement is different from compound agreement in another respect, namely that there are not just two nouns, but also two referents). This is the topic of the next section.

4.2 Conjunct agreement

Another type of gender mismatch occurs when two nouns of different genders are conjoined. This type of mismatch is resolved in different ways in different languages: agreement targets may agree with one of the conjoined nouns (typically the closest one), or all of them (for example with plural agreement or default agreement) (see e.g. Corbett 1991, 2006 for overviews of cross-linguistic data). The example in (71) illustrates both of these options within the same sentence.

(71) **Conjunct agreement in Russian** (Corbett 2006, p. 220)

Èt-a   vzyskatel'nost',   samokritičnost'   tože
This-F.SG.NOM exactingness.(F.SG.NOM) self-criticalness.(F.SG.NOM) also
raspolagal-i k nemu.
disposed-PL to 3SG.M.DAT

'This exactingness and self-criticalness also disposed me favorably towards him.'

The demonstrative, which has scope over both nouns, is singular, and agrees with the nearer conjunct. The predicate is plural, and Corbett (2006, p. 220) analyzes this as number resolution, a type of semantic agreement in which the target reflects the fact that there is a plurality of referents, rather than the agreement features of any individual conjunct (here feminine singular).

Based on the examples illustrated in (72)–(73), we can conclude that subject-verb agreement with conjuncts in Somali results in semantic agreement resulting from number resolution. The verb does not reflect the agreement features of any of the conjuncts (feminine or masculine), but rather occurs in its plural form, which does not express gender distinctions (see section 2.1). This holds true regardless of the order of the two noun phrases.[17]

---

[17]Note that direct agreement with disjoint noun phrases (*the X or the Y*) seems to be avoided, and that it is difficult to elicit natural-sounding sentences of the type in (73).

(72) **Conjunct agreement in Somali**

    a. Gabádh-dha iyo wíil-ku=ba   waa baahán yihiin.
        girl-F.DEF   and boy-M.DEF.NOM=each DECL hungry **3PL**.COP.PRS

        'The girl and the boy are hungry.'

    b. Wíil-ka   iyo gabádh-dhu=ba   waa baahán yihiin.
        boy-M.DEF and girl-F.DEF.NOM=each DECL hungry **3PL**.COP.PRS

        'The boy and the girl are hungry.'

(73) **Disjunct agreement in Somali**

    a. Hálkáas báa=ay   joog-een   libáax-a   ama abeesá-da.
        there   FOC=**3PL** stay-PST.**3PL** lion-M.DEF or   python-F.DEF

        'The lion or the python was there.'

    b. Hálkáas báa=ay   joog-een   abeesá-da   ama libáax-a.
        there   FOC=**3PL** stay-PST.**3PL** python-F.DEF or   lion-M.DEF

        'The python or the lion was there.'

This contrasts with noun-noun phrases (genitive constructions), in which the verb consistently agrees with the initial noun.[18]

(74) **Noun-noun phrases (left-headed)**

    a. Búug-ga   macállin-ku   waa adág yahay.
        book-**M**.DEF teacher-**M**.DEF.NOM DECL hard 3SG.**M**.COP.PRS

        'The book of the (male) teacher is hard/difficult.'

    b. Búug-ga   macallimád-du   waa adág yahay.
        book-**M**.DEF teacher-**F**.DEF.NOM DECL hard 3SG.**M**.COP.PRS

        'The book of the (female) teacher is hard/difficult.'

    c. Warqád-da macállin-ku   waa adág tahay.
        paper-**F**.DEF teacher-**M**.DEF.NOM DECL hard 3SG.**F**.COP.PRS

        'The paper of the (male) teacher is hard/difficult.'

    d. Warqád-da macallimád-du   waa adág tahay.
        paper-**F**.DEF teacher-**F**.DEF.NOM DECL hard 3SG.**F**.COP.PRS

        'The paper of the (female) teacher is hard/difficult.'

Conjunct agreement could also be compared to agreement with noun-noun compounds if it were the case that the definite article could take scope over two conjoined nouns (as in *the mother and father*). This could potentially have formed a parallel to noun-noun compounds, as

---

[18]It is the final definite article which gets the case marking. Nominative case marking is phrasal and targets the final morpheme of a noun phrase – see Banti (1984); Saeed (1999, p. 66); Lampitelli and Le Gac (2016) for details. Note that there is no overt genitive marking when both nouns are definite – see Nilsson (2016a).

it would have the structure *Noun1 and Noun2-DEF*, a structure similar to that of definite noun-noun compounds, *Noun1 + Noun2-DEF*. To my knowledge, this is marginal if not impossible in Somali, as coordinated noun phrases need to either both be indefinite (in which case there is no article), or both be definite (in which case each noun has its own article) (75).

(75) **The scope of the definite article**

a. hooyó iyo aabbé
mother and father
'a mother and a father'

b. hooyá-da   iyo aabbá-ha
mother-F.DEF and father-M.DEF
'the mother and the father'

c. *hooyó iyo aabbá-ha
mother and father-M.DEF
Intended: 'the mother and father'

d. *hooyá-da   iyo aabbé
mother-F.DEF and father
Intended: 'the mother and father'

There is one particular type of construction that allows for this, though, which is described below, and which pertains to numerals. Numerals in Somali are considered to be nouns by syntactic and morphological criteria: for example, they can head a noun phrase and are modified by determiners, and the choice of determiner is typically considered to indicate their grammatical gender (see Saeed 1999, pp. 69–72). For example, *labá-da* 'the two' is feminine, and *tobán-ka* 'the ten' is masculine (76). The complication with numerals is that one cannot independently establish their gender by considering subject-verb agreement: As these examples illustrate, numerals (above one) lead to plural agreement on the verb, and as already mentioned above, there are no gender distinctions in such cases. Note that the numerals are the heads of the noun phrases in (76).

(76) **Agreement with numeral nouns**

a. Labá-da   nín = ba   waa dheerdhéer yihiin.
two-**F**.DEF man.GEN = each DECL tall.PL   **3PL**.COP.PRS
'The two men (lit. the two of man) are tall.'

b. Tobán-ka naagoód = ba   waa dheerdhéer yihiin.
ten-**M**.DEF woman.GEN = each DECL tall.PL   **3PL**.COP.PRS
'The ten women (lit. the ten of women) are tall.'

The construction of interest for the present purposes has to do with the scope of the definite article over complex numerals. Complex numerals can have either of the structures in (77), and can be analyzed as phrasal compounds: they have two High tones rather than one, and the numerals are intermediated by the conjunction *iyo* 'and', which otherwise is used for coordinating noun phrases, as in (72) above. Unlike the example in (72), but like noun-noun compounds,

the examples in (77) crucially have a single definite article. Notice that the choice of article consistently reflects the gender of the *final* member in these cases.[19]

(77) **Complex numerals**

   a. labá-iyo-tobán-ka   →   *labíyotobán-ka*
       two.(F)-and-ten.(M)-**M**.DEF
       'the twelve, twelve o'clock'

   b. tobán-iyo-labá-da
       ten.(M)-and-two.(F)-**F**.DEF
       'the twelve, twelve o'clock'

The examples in (77) constitutes a parallel to noun-noun compounds and illustrates another type of construction in which linear adjacency matters for definite article assignment.

4.3   Agreement attraction

The linear adjacency strategy for definite article assignment, illustrated by a form like *laf+gárab-ka* (a feminine-masculine compound with a masculine definite article) is at first sight reminiscent of what has been called *agreement attraction* or *proximity concord* (see Bock and Miller 1991): For example, in the English sentence *The key to the cabinets are on the table*, the verb (*are*) agrees with an intervening noun (*cabinets* rather than *key*), and the same might be said for the definite article in Somali in *laf+gárab-ka* above. That is, by the time the speaker has uttered the second member of the compound and chooses a definite article, the gender of the first noun may have been "switched off". However, since subject-verb agreement does not vary this way with noun-noun root compounds, and the verb in fact occurs *after* the definite article, this does not seem to explain the mismatch in article and verb agreement. It would require the speaker to "switch on" the gender of the first noun again to produce the gender agreement patterns seen in section 3.3. Furthermore, an analysis along these lines would involve analyzing the agreement mismatches as a processing phenomenon. This does not seem to be what is going on here, since the agreement mismatches occur even in a controlled elicitation setting, and the speaker does not indicate that there is anything "wrong" with sentences in which different agreement targets show different agreement values.

---

[19]The choice between the two orders in (77) reflects regional variation (Nilsson 2018). However, some speakers, including this particular one, produce both orders.

5   Discussion

The agreement patterns found with noun-noun compounds in Somali when the gender of the two members do not match (section 3.3), are reminiscent of those found with other types of gender mismatches across languages, e.g. with hybrid nouns and conjunct agreement (section 4). A question that arises is what it is about Somali that allows for the variation in agreement patterns in noun-noun compounds in particular. I will argue that the agreement variation is the result of a number of factors operating in conjunction: the location of the semantic head relative to the definite article, the tonal gender marking, the fact that the article is phonologically bound to the final member of the compound (the non-head), and the coexistence of "left-headed" and "right-headed" constructions within the same system. Taking a usage-based perspective, I will argue that the variation we see in cases of gender conflict results from a system which allows for competing generalizations to be made by speakers when the language is acquired.

5.1   The location of the semantic head and the definite article

The data presented in sections 3.2–3.3 illustrated that noun-noun compounds in Somali have their semantic head to the left, and their definite article, as one of the possible agreement targets, to the right. Table 6 illustrates an overview of logically possible combination of headedness and the location of agreement targets in noun-noun compounds. Somali is an example of the pattern in (i).

		semantic head to the left		semantic head to the right
postposed agreement target	i.	**head** + nonhead **target**	iii.	nonhead + **head target**
preposed agreement target	ii.	**target head** + nonhead	iv.	**target** nonhead + **head**

Table 6: Possible locations of heads and agreement targets

The mirror image of the Somali pattern is the one in (iv) – we already saw an example of this in section 2.4, which illustrated compounds in Dutch. Another example of a language like this is Norwegian. The Somali (i) and the Norwegian (iv) patterns are illustrated in (78)–(79). While Norwegian has preposed indefinite articles, as in this example, it has postposed definite articles. Therefore, Norwegian will also serve to illustrate the pattern in (iii), given in (80). Finally, the pattern in (ii), in which the agreement target precedes the compound, and the semantic head is to the left, is illustrated with Spanish in (81).[20]

---

[20]The Norwegian examples reflect Urban East Norwegian (the Oslo dialect) and are provided by the author, while the Spanish examples were provided by José Armando Fernández Guerrero.

(78)   **head** + nonhead **target** (Somali)

    a.   bur + saliíd-ka
       flour/cake.(M) + oil.(F)-M.DEF
       'the fritter'

    b.   laf + dhábar-ta
       bone.(F) + back.(M)-F.DEF
       'the spine'

(79)   **target** nonhead + **head** (Norwegian)

    a.   et   nisse + hus
       N.INDF nisse.(M) + hus.(N)
       'a house for nisses'
       (a *nisse* is a hob-like creature from Norwegian folklore)

    b.   en   hus + nisse
       M.INDF hus.(N) + nisse.(M)
       'a house-nisse'

(80)   nonhead + **head target** (Norwegian)

    a.   nisse + hus-et
       nisse.(M) + hus.(N)-N.DEF
       'the house for nisses'

    b.   hus + nisse-n
       hus.(N) + nisse.(M)-M.DEF
       'the house-nisse'

(81)   **target head** + nonhead (Spanish)

    a.   el   arco + iris
       M.DEF bow.(M) + iris.(F)
       'the rainbow'

    b.   la   casa + cuartel
       F.DEF house.(F) + barracks.(M)
       'the barracks house, military building'

While the Somali construction in (78) and the Norwegian construction in (79) share the fact that the agreement target occurs on the opposite edge of the semantic head, any variation in agreement is, to my knowledge, impossible in the Norwegian case. That is, the forms *en nisse + hus* and *et hus + nisse* are ungrammatical. Therefore, some properties of the Norwegian and Somali systems will be compared in order to shed some light over this difference between them. The first thing to note is that while Norwegian has a tone system, there is no tonal gender marking comparable to what is found in Somali. Hence there are no cases of mismatches between the

gender indicated by the semantic head, and the gender indicated by the phonological form of a compound (as no such cues exist). Further comparisons will be made in what follows.

5.2   The boundness of the article

While the indefinite article in Norwegian is an independent, unbound form, and adjectives may intervene between the article and the noun (e.g. *en liten hus + nisse* 'a small house-nisse'), the definite article in Somali is phonologically bound (it undergoes the obligatory sandhi alternations described in section 2.1), and no modifiers can intervene between the noun and the article.[21] The Somali definite article has this in common with the Norwegian definite article illustrated in (80), which is postposed. The Norwegian definite suffixes have developed from independent forms which showed gender agreement historically (Faarlund 2009), but scholars disagree about whether they synchronically should be analyzed as showing gender agreement or not. If one uses Hockett's definition of gender (gender is "reflected in the behavior of associated words" Hockett 1958, p. 231) demarcatively, then definite suffixes do not show gender agreement, since they are not "associated words". Rather, their choice of form is a matter of allomorphy, a property of the paradigms of different declensions. Declension and gender may or may not line up in a given language, but they are at least conceptually different under this view. The question is tricky and partially depends on one's view of syntax and morphology (see Enger (2004), Enger and Corbett (2012), Lødrup (2011), Rødvand (2017), and Svenonius (2016)). A similar issue has been raised with the Somali definite article, which variously has been analyzed as a clitic (thus a possible agreement target) and a suffix (whose form depends on declension) (see e.g. Nilsson 2016b). If it is analyzed as a suffix, and Hockett's definition of gender is used demarcatively, then the variation found in definite article assignment to compounds reflects variation in inflectional class assignment rather than agreement, and we know cross-linguistically that gender and inflectional class do not necessarily line up.[22]

Regardless of one's take on the question of whether bound articles show agreement, there is another point to be made about the phonological boundness of the Somali definite article: the choice of allomorph depends on phonological properties of the non-head of a noun-noun compound, even when its gender value is determined by the head. This is illustrated in (82).

---

[21]There is one exception to this, namely possessive determiners, as in *gúri-gaá-ga*, house-POSS-DEF, 'your house'. However, the definite article shows some idiosyncratic phonological properties in this particular construction (see Saeed 1999, p. 115).

[22]A famous example is Latin nouns of the type *agricola* 'farmer', which declines as a feminine noun, but controls masculine agreement.

(82)   bir + xárbi-du
       iron + war-F.DEF.NOM
       'barrel cleaning rod'   (c.f. *bír-tu* (F), *xárbi-gu* (M))

Here, the form of the feminine article is *-du*, and not *-tu* because the final member (which is masculine) ends in *-i*, and conditions a voiced stop (realized by some as a fricative). The form of the feminine article is *-tu* in *bír-tu* 'the iron'. This illustrates that linear adjacency matters for the phonological form of the definite article even when it does not matter for agreement: these are structural generalizations paying attention to form (locally). However, linear adjacency matters in other ways too, as in the morphological constructions discussed in the next section.

5.3   When left-headed and right-headed structures coexist

Recall from section 2.1 that derivational affixes are suffixal in Somali, and that they usually determine gender. In this sense, they are morphosyntactic heads, and form right-headed structures. This is illustrated in (83), in which the derived forms are feminine regardless of the gender of the noun they are derived from.[23]

(83)   **Gender in derived words**
       a.   *nasíib* (**M**)   'fortune'   *-dárro*   →   *nasiibdárro*   'misfortune' (**F**)
       b.   *edéb* (**F**)   'good manners'   *-dárro*   →   *edebdárro*   'lack of good manners' (**F**)

Noun-noun compounds can be derived too. In this case, the derivational suffix is added at the right edge, and assigns gender, as illustrated in (84). It therefore functions as a morphosyntactic head.

(84)   **Nouns derived from compounds**
       a.   *mádax* (**M**) 'head' + *bannáan* (**M**) 'emptiness, freedom' + *-í* NMLZ (**F**)
            →   *madaxbannaaní* 'independence' (**F**)
       b.   *qáb* (**M**) 'pride' + *wéyn* 'big' + *-í* NMLZ (**F**)
            →   *qabweyní* 'arrogance' (**F**)

Note that the structures here are [[Noun [Noun]] -suffix] and [[[Noun] Adj] -suffix] – these are thus right-headed morphological schema coexisting with the one involving the suffix *-eed*

---

[23]While *-dárro* is a bound form which is not used on its own (hence the label "suffix" here), it can be analyzed as a derived form of *daran* '(to be) bad, serious, tough'. Therefore, a parallel can be drawn between this and the types of compounds that Banti and Jama (2016) analyze as right-headed, mentioned in section 2.4, namely noun-noun compounds in which the rightmost member is a deverbal noun. I am grateful to Giorgio Banti for pointing this out to me.

([Noun [Noun-eed]]), in which the properties of headedness are split between the initial noun and the suffix (see section 3.6). These examples illustrate how generalizations involving linear adjacency and agreement with the final element may come about.

There are several other types of morphological constructions whose agreement patterns should be investigated in the future to shed light on compound gender in Somali. A small selection was elicited by the speaker in the present study, and will be provided here for a preliminary comparison. First, compounds in which one of the members is a heterosemous root (discussed in section 2.3) have their semantic head to the right, and the gender agreement is predictable from the gender of this member, as illustrated with the examples in (85). These compounds are both masculine, regardless of the gender of the initial member.

(85) **Gender in compounds with a heterosemous root**

    a. *mádax* (**M**) 'head' + *xanúun* (**M**) 'pain'/'be in pain'
    → *madax + xanúun* 'headache' (**M**)

    b. *caoól* (**F**) 'stomach' + *xanúun* (**M**) 'pain'/'be in pain'
    → *calool + xanúun* 'stomachache' (**M**)

Second, when adjectival noun-adjective compounds are derived to form nominal compounds, the resulting structure is semantically right-headed: compare (86)[24] and (87). Using the 'IS A' criterion from section 2.4, I analyze both *madax + adáyg* 'stubbornness' and *calool + adáyg* 'courage' as a type of *adáyg* 'hardness' (itself a derived noun, and masculine).[25] Once again, the gender is determined by the final member.

(86) **Adjectival compounds**

    a. mádax (M) 'head' + adág 'hard' → mádaxadág 'stubborn'
    b. caoól (F) 'stomach' + adág 'hard' → caoóladág 'courageous'

(87) **Nominalized adjectival compounds**

    a. Madax + adáyg-gu   waa xún yahay.
    head + hardness-**M**.DEF.NOM DECL bad 3SG.**M**.COP.PRS
    'The stubbornness is bad.'

    b. Calool + adáyg-gu   waa fiicán yahay.
    stomach + hardness-**M**.DEF.NOM DECL good 3SG.**M**.COP.PRS
    'The courage is good.'

---

[24]The tone patterns of adjectival compounds are different from the tone patterns of nominal compounds - more specifically, they have two High tones.

[25]This is a type of derivation restricted to a few adjectives in Somali. Another example is *jecel* 'loving' – *jecayl* 'love'.

The examples discussed here illustrate that the linear adjacency strategy for definite article assignment to compounds does not come out of nowhere. The morphological constructions discussed in the present section have their semantic head to the *right*, and as usual, the definite article is added at the right edge. While the consistency of these agreement patterns needs to be determined in future research, it is possible that these right-headed structures have allowed for a structural generalization in which the choice of definite article is determined by the gender of the element it attaches to.

Since left-headed and right-headed constructions coexist in Somali, it is not the case that the definite article consistently is on the opposite side of the head of a morphological construction, like we saw in noun-noun root compounds – rather, the definite article is consistently added at the right edge of nouns. Since the location of the definite article is consistent, but the location of the head varies, it follows that the location of the agreement controller varies relative to the location of the agreement target. This property of the definite article in Somali is not something that becomes clear from looking at a single word or a single type of construction – rather, it is a property of the system as a whole.

To summarize, I argue that the facts discussed in the present section taken together has allowed for competing generalizations or subschemas for definite article assignment, resulting in variation in cases of conflicting gender cues. This analysis would predict variation (for a given speaker, but also between speakers) in cases where the semantic head is to the left and there is a conflict between the gender of the head and the gender of the rightmost element, but not in cases of no conflict. It also captures why this variation is attested in Somali, but not in a language like Norwegian: Norwegian is also largely suffixing, so derived forms are right-headed, but so are noun-noun compounds are in this language. There may be additional factors within the Norwegian system which pushes *against* variable agreement patterns. However, a further exploration of this possibility goes beyond the scope of the present study.

6   Concluding remarks

When two Somali noun roots of different genders are compounded, various types of mismatches in gender cues are found. While subject-verb agreement consistently is predictable from the gender of the initial member of noun-noun root compounds, the choice of definite article varies. Two competing generalizations are available with respect to definite article assignment: one based on linear adjacency, in which the article agrees with the closest and final member (F + **M-ka**), and one based on the gender of the semantic head, which is the initial member (**F** + **M-ta**). When the linear adjacency strategy for definite article assignment (F + **M-ka**) is

combined with subject-verb agreement based on the gender of the initial member, the result is a mismatch between gender values on the two agreement targets. Noun-noun root compounds in Somali thus shows variable agreement patterns, typical of hybrid agreement controllers.

The mismatch between subject-verb agreement and definite article assignment is paralleled by a mismatch in gender and tone pattern frequently found in compounds (section 3.5). Non-compounds in Somali show tonal gender marking: most singular masculine nouns have a penultimate High tone (*gárab* 'shoulder' (M)), and most feminine nouns have a final High tone (*saliíd* 'oil' (F)). In noun-noun compounds, it is the *final* member which determines the tone pattern. If this is in conflict with the gender of the semantic head (to the left), the result is either a masculine noun with a final rather than a penultimate High tone (M + F *bur* + *saliíd* 'fritter' (M)), or a feminine noun with a penultimate rather than final High tone (F + M *laf* + *gárab* 'shoulder bone' (F)). That is, there will either be a match between definite article and subject-verb agreement, but a mismatch between definite article assignment and tonal gender marking, or a mismatch between definite article assignment and subject-verb agreement, but a match between definite article and tone pattern. As argued in section 3.5, the attested variation in definite article assignment potentially reflects choosing between the two types of (mis-)matches.

Noun-noun root compounds constitutes just one out of many constructions and word-formation strategies in Somali. I argue that one needs to take a birds-eye view of the system as a whole to explain the variation found in definite article assignment: First, the Somali grammar and lexicon beyond noun-noun compounds is a system in which left-headed and right-headed structures coexist, allowing for competing subpatterns of generalizations to be formed by speakers: for example, while noun-noun root compounds are left-headed, derived forms and certain other types of compounds are right-headed (section 5). Second, the properties of headedness do not always line up: In compounds in which the final member has an attributive suffix, such as the *-eed* in *jirrid* + *maskax-éed* 'brain stem (lit. stem-brain-ATTR)', the properties are split between the initial member (the semantic head) and the final element (which may control masculine agreement although both members of the compound are feminine) (section 3.6). Speakers are therefore left with a choice when assigning an article, and the variable outcome of that choice reflects the competing generalizations that has been allowed for by the system as a whole.

In a broader typological perspective, the Somali data illustrate the following point: "Does language X have left-headed or right-headed compounds?" is simply the wrong question to ask, as the answer will depend on which criteria one uses to define *head*. Because most "false universals" in compounding are related to the notion "head" (Guevara and Scalise 2009, p. 124), both typological and descriptive studies of compounds are better off describing the *properties* of headedness directly (as I have aimed to do in this paper), than by forcing the data into pre-

established categories such as "heads". See e.g. Hyman (2009) for a property-driven approach to typology, and Haspelmath (2007) and Haspelmath (2018) on *language-particular categories* and *comparative concepts* as alternatives to pre-established categories.

Compounds in Somali are understudied, and there are many open questions to pursue. The present study is based on data from a single speaker, and future research should include data from more speakers to investigate potential individual differences. Additionally, one should move beyond noun-noun compounds to other types of structures, such as nominal noun-adjective and noun-verb compounds. Furthermore, Somali has many different ways of forming the plural. Therefore, another question is how compounds are pluralized, and how this relates to form-class and headedness.

References

Allen, M. R. (1978). *Morphological investigations (Ph.D. dissertation)*. University of Connecticut.

Andrzejewski, B. W. (1964). *The declension of Somali nouns*. London: School of Oriental and African Studies, University of London.

Armstrong, L. E. (1934). "The Phonetic Structure of Somali". In: *Mitteilungen des Seminars für Orientalische Sprachen zu Berlin* 37.3, pp. 116–161.

Banti, G. (1984). "The morphology of the nominative case in Somali". In: *Discussion papers for the Fifth International Phonology Meeting*. Ed. by W. U. Dressler, O. E. Pfeiffer, and J. R. Rennison. Wiener Linguistische Gazette, Beiheft 3.

– (1986). "Review of Serzisko, Ausdruck der Possessivitat im Somali". In: *Sonderdruck aus Kratylos* 31, pp. 161–168.

– (1988a). "'Adjectives' in East Cushitic". In: *Papers from the International Symposium on Cushitic and Omotic Languages, Cologne 1986*. Hamburg: Helmut Buske Verlag, pp. 204–259.

– (1988b). "Two Cushitic Systems: Somali and Oromo Nouns". In: *Autosegmental Studies on Pitch Accent*. Ed. by H. van der Hulst and N. Smith. Dordrecht: Foris Publications, pp. 11–50.

– (2009). "Somali variation across space and society: state of the art". In: *HH Polylektalität 27-29 Nov. 2009*.

– (2016). *Tonal Sandhi in Northern Somali*. Paper presented at Colloquium on African Languages and Linguistics 2016, Leiden University.

Banti, G. and J. M. Jama (2016). *Compounds and Multi Word Expressions (MWEs) in Somali (ms.)*

Barillot, X. (2002). *Morphophonologie gabaritique et information consonantique latente en somali et dans les langues est-couchitiques (PhD thesis)*. Université Paris VII Denis Diderot.

Bauer, L. (1990). "Be-heading the word". In: *Journal of linguistics* 26 (1), pp. 1–31.

Bauer, L. (2008). "Exocentric compounds". In: *Morphology* 18 (1), pp. 51–74.

– (2017). *Compounds and Compounding*. Cambridge: Cambridge University Press.

Bendjaballah, S. and D. Le Gac (2019). *The acoustics of Somali voiced stops*. Workshop on Somali Grammar, University of Gothenburg.

Bloomfield, L. (1933). *Language*. New York: Henry Holt and Company.

Bock, K. and C. A. Miller (1991). "Broken Agreement". In: *Cognitive Psychology* 23 (1), pp. 45–93.

Booij, G. (2002). *The Morphology of Dutch*. Oxford: Oxford University Press.

– (2010a). "Compound construction: Schemas or analogy? A construction morphology perspective". In: *Cross-disciplinary issues in compounding*. Ed. by S. Scalise and I. Vogel. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 98–108.

– (2010b). "Construction morphology". In: *Language and linguistics compass* 4 (7), pp. 543–555.

Bybee, J. (2003). *Phonology and language use*. Cambridge: Cambridge University Press.

Caney, J. C. (1984). *The Modernisation of Somali Vocabulary, with Particular Reference to the Period from 1972 to the Present*. Hamburg: Helmut Buske Verlag.

Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press.

– (2006). *Agreement*. Cambridge: Cambridge University Press.

Corbett, G. G., N. M. Fraser, and S. McGlashan (1993). *Heads in grammatical theory*. Cambridge: Cambridge University Press.

Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Downing, L. and M. Nilsson (2019). "Prosodic restructuring in Somali nominals". In: *African linguistics across the disciplines: Selected papers from the 48th Annual Conference on African Linguistics*. Ed. by S. Lotven et al. Language Science Press, pp. 125–142.

Eiesland, E.-A. (2015). *The Semantics of Norwegian Noun-Noun Compounds (Ph.D. dissertation)*. University of Oslo.

Enger, H.-O. (1995). "Har ord hoder? Litt om sammensetninger og deres bøyning". In: *LexicoNordica* 2, pp. 33–42.

– (2004). "On the relation between gender and declension: A diachronic perspective from Norwegian". In: *Studies in Language* 28 (1), pp. 51–82.

Enger, H.-O. and G. G. Corbett (2012). "Definiteness, gender, and hybrids: Evidence from Norwegian dialects". In: *Journal of Germanic Linguistics* 24 (4), pp. 287–324.

Faarlund, J. T. (2009). "On the history of definiteness marking in Scandinavian". In: *Journal of Linguistics* 45 (3), pp. 617–639.

Fábregas, A. and F. Masini (2015). *Prominence in morphology: the notion of head*.

Green, C. R. and M. E. Morrison (2016). "Somali wordhood and its relationship to prosodic structure". In: *Morphology* 26.1, pp. 3–32.

– (2018). "On the morphophonology of domains in Somali verbs and nouns". In: *Brill's Journal of Afroasiatic Languages and Linguistics* 10 (2), pp. 200–237.

Guevara, E. and S. Scalise (2009). "Searching for universals in compounding". In: *Universals of language today*. Ed. by S. Scalise, E. Magni, and A. Bisetto. Dordecht: Springer, pp. 101–128.

Harris, J. W. (1991). "The exponence of gender in Spanish". In: *Linguistic Inquiry* 22 (1), pp. 27–62.

Haspelmath, M. (2007). "Pre-established categories don't exist: Consequences for language description and typology". In: *Linguistic typology* 11 (1), pp. 119–132.

Haspelmath, M. (2018). "How comparative concepts and descriptive linguistic categories are different". In: *Aspects of Linguistic Variation*. Ed. by D. Olmen, T. Mortelmans, and F. Brisard. Berlin, Boston: De Gruyter, pp. 83–114.

Hetzron, R. (1967). "Agaw numerals and incongruence in Semitic". In: *Journal of Semitic studies* 12 (2), pp. 169–197.

– (1972). "Phonology in syntax". In: *Journal of linguistics* 8 (2), pp. 251–265.

Hockett, C. F. (1958). *A Course in Modern Linguistics*. New York: Macmillan.

Hudson, R. A. (1987). "Zwicky on heads". In: *Journal of linguistics* 23 (1), pp. 109–132.

Hyman, L. M. (1981). "Tonal accent in Somali". In: *Studies in African linguistics* 12.2, pp. 169–203.

– (2009). "How (not) to do phonological typology: the case of pitch-accent". In: *Language Sciences* 31.2, pp. 213–238.

Jackendoff, R. (2016). "English noun-noun compounds in Conceptual Semantics". In: *The semantics of compounding*. Ed. by P. ten Hacken. Cambridge: Cambridge University Press, pp. 15–37.

Kaldhol, N. H. and S. Stausland Johnsen (forthcoming). "There is no 'one High tone per word' rule in Somali". In: *Synchronic and diachronic aspects of tonal accent*. Ed. by B. Köhnlein and P. Iosad. Oxford: Oxford University Press.

Lampitelli, N. (2013). "The decomposition of Somali nouns". In: *Brill's Journal of Afroasiatic Languages and Linguistics* 5 (1).

Lampitelli, N. and D. Le Gac (2016). *Remarks on the Nominative Case in Somali*. Colloquium on African Languages and Linguistics 2016, Leiden University.

Langacker, R. (1987). *Foundations of cognitive grammar, Volume I: theoretical prerequisites*. Stanford, CA: Stanford University Press.

Le Gac, D. (2002). "Tonal Alternations and Prosodic Structure in Somali". In: *Speech Prosody*.

– (2003). "Tonal Alternations in Somali". In: *Research in Afroasiatic Grammar II*. Ed. by J. Lecarme. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 287–304.

– (2016). "Somali as a Tone Language". In: *Speech Prosody*. Boston, pp. 292–296.

Lødrup, H. (2011). "Hvor mange genus er det i Oslo-dialekten?" In: *Maal og minne* 103 (2).

Mansuur, C. C. and A. Puglielli (2012). *Qaamuuska Af-Soomaaliga*. Roma TrE-Press.

Moskal, B. and P. W. Smith (2019). "The Status of Heads in Morphology". In: *Oxford Research Encyclopedia of Linguistics*.

Mous, M. (2012). "Cushitic". In: *The Afroasiatic Languages*. Ed. by Z. Frajzyngier and E. Shay. Cambridge: Cambridge University Press, pp. 342–422.

Nilsson, M. (2016a). *Does Somali have a genitive case?* Afrikanistentag, Berlin.

– (2016b). "Somali Gender Polarity Revisited". In: *Diversity in African languages: Selected papers from the 46th Annual Conference on African Linguistics*.

– (2018). *Somaliska skriftspråket: En hundraåring som sprudlar av energi*. Bokmässan i Göteborg.

– (2019). *On the Non-Canonicity of case in Somali*. Paper presented at Workshop on Somali Grammar, University of Gothenburg.

Pepper, S. (2010). *Nominal compounding in Nizaa – a cognitive perspective. MA Thesis.* SOAS, University of London.

– (2020). *The typology and semantics of binominal lexemes: Noun-noun compounds and their functional equivalents (Ph.D. dissertation).* University of Oslo.

Puglielli, A. (1989). *Compounds in Somali (ms.)*

Rødvand, L. I. S. (2017). *Empirical investigations of grammatical gender in American Heritage Norwegian (Master's thesis).* University of Oslo.

Saeed, J. I. (1999). *Somali (London Oriental and African Language Library Vol. 10).* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Scalise, S. and A. Fábregas (2010). "The head in compounding". In: *Cross-disciplinary issues in compounding.* Ed. by S. Scalise and I. Vogel. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 109–126.

Svenonius, P. (2016). *Declension class and the Norwegian definite suffix.* The Morphosyntax-Phonology Connection: Locality and Directionality at the Interface.

Tosco, M. (2012). "The unity and diversity of Somali dialectal variants". In: *The harmonization and standardization of Kenyan languages: Orthography and other aspects.* Ed. by N. O. Ogechi, J. A. N. Oduor, and P. Iribemwangi. Cape Town: The Centre for Advanced Studies of African Society (CASAS), pp. 263–280.

Williams, E. (1981). "On the Notions "Lexically Related" and "Head of a Word"". In: *Linguistic Inquiry* 12 (2), pp. 245–274.

Zorc, R. D. and M. M. Osman (1993). *Somali - English Dictionary with English Index.* 3rd ed. Kensington: Dunwoody Press.

Zwicky, A. M. (1985). "Heads". In: *Journal of linguistics* 21 (1), pp. 1–29.

Acknowledgments

Appendix

Noun-Noun root compounds

Compound	Translation	Literal translation	Tone	Definite article	Subject/verb-agreement
kabsar + caleen	cilantro	herb + leaf	Fin	F	F
laf + arax	spine	bone + vertebra	Fin	F	F
dabo + dhilif	tailbone	tail + tail.tip	Fin	F	F

Table 7: F + F compounds

Compound	Translation	Literal translation	Tone	Definite article	Subject/verb-agreement
gaadhi + dameer	cart pulled by a donkey	car + donkey	Pen	M	M
gaadhi + faras	horse cart	car + horse	Pen	M	M
gabbal + gaab	evening, time just before sunset	daylight + shortness	Pen	M	M
kul + jiir	lukewarm temperature	heat + boneless meat	Pen	M	M
nasiib + wanaag	good luck, fortunately	luck + good	Pen	M	M
roob + garas	light rain	rain + tree.species	Pen	M	M
waa + beri	dawn, early morning	dawn + time	Pen	M	M
dhagax + aboodi	medicinal stone	stone + vulture	**Fin**[a]	M	M
fool + maroodi	ivory, elephant's tusk	front.tooth + elephant	**Fin**[a]	M	M
dhul + bar	equator	earth + half	**Fin**[b]	M	M

Table 8: M + M compounds

[a]The final High tone in these examples reflect the fact that the final member belongs to the class of nouns that have a final High tone despite being masculine (*aboodí-ga, maroodí-ga*).

[b]The final High tone in this example reflect the fact that the final member is monomoraic. That is, given that only the final member in a compound carries a High tone, this is the only option for where that High tone can go. Hence the compound as a whole have a High tone on the final mora despite being masculine.

Compound	Translation	Literal translation	Tone	Definite article	Subject/verb-agreement
bur + saliid	fritter	flour/cake + oil	Fin[c]	M	M
daacuun + calool	cholera	cholera + stomach	Fin[c]	M	M

Table 9: M + F compounds

---

[c]The location of the High tone in these examples needs to be verified. The surface pitch on the final syllable is ambiguous between level (a simplified rising tone, that is, a High tone on the final mora) and falling (a High tone on the penultimate mora), and there is a chance that the tone (perhaps optionally) shifts to the penultimate mora to reflect the masculine compound gender.

[d]The final High tone in these examples reflect the fact that the final member belongs to declension 3, a class of nouns that have a final High tone in their definite forms despite being masculine (*doollí-ga, maroodí-ga*).

Compound	Translation	Literal translation	Tone	Definite article	Subject/verb-agreement
bir + danab	magnet	iron + thunder	Pen	F	F
bir + jiir	mouse trap	iron + mouse	Pen	F/M	F
bir + xarbi	barrel cleaning rod	iron + war	Pen	F	F
fool + diin	false labor pains	labor.pain + turtle	Pen	M	F
kaadhi + dhiig	disease, bloody urine	urine + blood	Pen	F/M	F
laf + dhabar	spine	bone + back	Pen	F	F
laf + duud	spine	bone + back	Pen	F	F
laf + garab	shoulder bone	bone + shoulder	Pen	F	F
laf + guri	a medical condition	bone + house	Pen	F	F
laf + madax	skull	bone + head	Pen	F/M	F
laf + sakaar	breast bone	bone + center.of.chest	Pen	F/M	F
liin + ugunji	blood orange	citrus.fruit + blood.orange	Pen	F/M	F
magaalo + madax	capital city	city + head	Pen	F	F
magaalo + qaran	city state	city + nation	Pen	M	F
shimbir + malab	honey guide bird	bird + honey	Pen	F	F
shimbir + yaxaas	bird sp.	bird + crocodile	Pen	F	F
waddo + halaq	the way snakes go	road + snake	Pen	M	F
waddo + hunguri	esophagus	road + throat	Pen	F/M	F
bir + doolli	mouse trap	iron + mouse	**Fin**[d]	F	F
ciddi + maroodi	elephant's claw	nail/claw + elephant	**Fin**[d]	M	F

Table 10: F + M compounds

The suffixes -eed/-aad/-ood

Compound	Translation	Literal translation	Tone	Definite article	Subject/verb-agreement
fool + beenaad	false labor pains	labor.pain + falsity.ATTR	Pen	F	F
dawo + dhacameed	natural medicine	medicine + tree.shelter.ATTR	Pen	M	F
hawl + gacmeed	manual labor	work + hand.ATTR	Pen	M	F
hees + carruureed	children's song	song + children.ATTR	Pen	F	F
jirrid + maskaxeed	brain stem	stem + brain.ATTR	Pen	M	M
khariidad + gacmeed	sketch map	map + hand.ATTR	Pen	M	F
shaqo + gacmeed	manual labor	work + hand.ATTR	Pen	M	F
shax + taariikheed	grid calendar	Somali.checkers + calendar.ATTR	Pen	M	F
taariikh + nololeed	biography	history + life.ATTR	Pen	M	F/M

Table 11: F + F-ATTR compounds

Compound	Translation	Literal translation	Tone	Definite article	Subject/verb-agreement
udub + dhexaad	center post of hut	pole + center.ATTR	Pen	M	M
dayax + gacmeed	satellite	moon + hand.ATTR	Pen	M	M
dhagax + ciideed	sand stone	stone + sand.ATTR	Pen	M	M
dhagax + dixeed	pebble	stone + dry.streambed.ATTR	Pen	M	M
dhagax + nuuradeed	lime stone	stone + lime.ATTR	Pen	M	M
dhul + beereed	farmland	earth + farm.ATTR	Pen	M	M
gaari + gacmeed	wheel barrow	car + hand.ATTR	Pen	M	M
geed + abaareed	desert plant	plant + drought.ATTR	Pen	M	M
libaax + badeed	shark	lion + sea.ATTR	Pen	M	M
maroodi + badeed	walrus	elephant + sea.ATTR	Pen	M	M
masar + gacmeed	handkerchief	scarf + hand.ATTR	Pen	M	M

Table 12: M + F-ATTR compounds