

Chapter 8

Gender and headedness in nominal compounds in Somali

Nina Hagen Kaldhol

University of California, San Diego

This paper examines how gender is assigned to compounds in Somali, and how this relates to the notion of headedness. When two Somali noun roots of different genders are compounded, various types of mismatches in gender cues are found: subject-verb agreement is consistently predictable from the gender of the initial member of the compound (the semantic head), suggesting that this is the member which determines compound gender. In contrast, the definite article, which is phonologically bound to the final member, shows variable gender agreement: it is either in line with the compound gender or the gender of the final member. Somali furthermore exhibits 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 compound gender on the one hand, and the tone pattern of the compound on the other. I propose that the attested variation in definite article assignment is the result of choosing either a mismatch between definite article and tone pattern on the one hand, and on the other, a mismatch in agreement cues on the article and the verb.

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? Conflicting generalizations have previously been proposed regarding compound gender in Somali (Cushitic, Horn of Africa), and for particular compounds, different genders have been reported by different scholars. The aim of the present

Nina Hagen Kaldhol. 2021. Gender and headedness in nominal compounds in Somali. In Akinbiyi Akinlabi, Lee Bickmore, Michael Cahill, Michael Diercks, Laura J. Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected papers from the 50th Annual Conference on African Linguistics*, 175–202. Berlin: Language Science Press. DOI: [10.5281/zenodo.5578832](https://doi.org/10.5281/zenodo.5578832)



paper is to arrive at a robust empirical generalization of how gender is assigned to one particular type of nominal compound, namely the one in which both members are noun roots.

In order to sort out the patterns, I investigate agreement patterns from two different agreement targets, namely the definite article and the copula verb. The agreement variation presented here suggests that the reason for the discrepancies in previous literature is that researchers have been relying too heavily on the definite article as a cue to gender. In the case of non-compounds, the definite article agrees with the noun in gender. In compounds, on the other hand, there is variation with respect to definite article assignment, and it is therefore not a reliable gender cue. In contrast, subject-verb agreement is consistent: it is predictable from the gender of the initial member, which is the semantic head, suggesting that the semantic head also determines compound gender. The difference between the two agreement targets (the definite article is variable, subject-verb agreement is consistent) leads to occasional mismatches in the gender they express. This is illustrated in (1),¹ in which a feminine-masculine compound controls masculine agreement on the definite article, but feminine agreement on the copula verb.

- (1) Wadda+hálaq-u waa ficán 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)

Judging from subject-verb agreement, it is the initial member which determines compound gender. I propose that there are two competing generalizations available: one of linear adjacency, in which the article agrees with the closest member, and one in which the article agrees with the compound as a holistic unit. §2 provides background on gender, agreement, tone, compounds and headedness. §3 presents the main data for the present study. In §4, gender mismatches for Somali compounds are compared to other types of gender mismatches found cross-linguistically. §5 concludes, and explanations for the variable definite article assignment are discussed, appealing to properties of the Somali system as a whole.

¹Here and in the remainder of the paper, words will be transcribed following the Somali orthography, with High tones added (marked with acute accents: *á*). The deviations between the orthographic symbols and the IPA symbols are as follows: Long vowels are represented with two consecutive vowels (*aa*). *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 [h̥], *y* is a palatal glide [j], and ' is a glottal stop [ʔ].

2 Background

2.1 The gender system in Somali

2.1.1 Gender and agreement

Somali has a gender system with two genders: masculine (M) and feminine (F). Some terminological background will be provided before the Somali system is illustrated. First, the following definition of gender is provided by Hockett:

Genders are classes of nouns reflected in the behavior of associated words (Hockett 1958: 231)

The notion of *associated words* is crucial, and means that “the determining criterion is agreement” (Corbett 1991: 4). Nouns can be classified in other ways, e.g. based on their case marking or plural marking, but this is a matter of declension or nominal form-class, not gender.² This may or may not correlate with gender, but it is conceptually different from gender, and it is *agreement* which reveals the gender system (see also Güldemann & Fiedler 2019; Di Garbo 2014; 2016).³ Agreement can be defined as follows (see also Corbett 2006: 4):

The term *agreement* commonly refers to some systematic covariance between a semantic or formal property of one element and a formal property of another. (Steele 1978: 610)

This *systematic covariance* is illustrated with the Somali examples in (2–3): the gender of the noun (M or F) covaries with the form of the definite article (*ku* or *tu*)⁴ and the copula verb (*yahay* or *tahay*).

- (2) Agreement with a masculine noun
 Málab-ku waa fiicán yahay.
 honey(M.SG)-M.DEF.NOM DECL good 3SG.M.COP.PRS
 ‘The honey is good.’

²Comprehensive analyses of such classifications of nouns in Somali are provided in Andrzejewski 1964; Banti 1988b; Saeed 1999.

³Note that the agreement-based approach to gender adopted here is largely in line with previous works on Somali by Nilsson (2016) and Güldemann & Maniscalco (2016), but differs from the assumptions in other works, such as Godon (1998); Lampitelli (2013; 2019); Lecarme (2002).

⁴These are the nominative case forms of the definite articles. The unmarked forms are *ka* and *ta*. Somali is analyzed as having a case system of the type *marked nominative*, which is common in Cushitic languages (Mous 2012) and other languages of East Africa (König 2006). The nominative case marking is realized on the final element of the noun phrase that functions as a subject. For details about the form of case, see e.g. Banti (1986; 1988b).

(3) Agreement with a feminine noun

Shimbír-tu waa fiicán tahay.
 bird(F.SG)-F.DEF.NOM DECL good 3SG.F.COP.PRS
 ‘The bird is good.’

In order to describe the agreement relationship, it is useful to define two more concepts, namely *controller* and *target* (see Corbett 1991: 4): the nouns in (2–3) have inherent gender, and they determine the form of the definite article and the copula verb. Therefore, the nouns are the gender *controllers*: they are the elements that determine gender agreement. The definite article and the copula verb, on the other hand, do not have inherent gender, but rather, they are gender *targets*: they are the elements that have their respective forms determined by the controller.

The agreement-based approach to gender and Hockett’s definition of gender as being “reflected in the behavior of associated words” (1958: 231) open up the question of what “words” are. The definite articles in Somali are phonologically bound to nouns, as evidenced by the obligatory sandhi alternations they undergo, illustrated in (4–5).⁵ Each article has four different allomorphs, predictable from the final segment of the noun (see Saeed 1999: 28 for details).

(4) Masculine nouns

- | | | | |
|----|-----------------|-----------|---------------------|
| a. | <i>dáb-ka</i> | [dábka] | ‘the fire’ |
| b. | <i>qálbi-ga</i> | [gálbiga] | ‘the heart, spirit’ |
| c. | <i>aabbá-ha</i> | [aabbáha] | ‘the father’ |
| d. | <i>gó’a</i> | [góʔa] | ‘the portion’ |

(5) Feminine nouns

- | | | | |
|----|-----------------|------------------------|--------------|
| a. | <i>xeéb-ta</i> | [heébta] | ‘the shore’ |
| b. | <i>hooyá-da</i> | [hoojáda] | ‘the mother’ |
| c. | <i>gabádha</i> | /gabád-ta/ – [gabádɗa] | ‘the girl’ |
| d. | <i>úsha</i> | /úl-ta/ – [úfa] | ‘the stick’ |

The definite articles have variously been analyzed as clitics (e.g. Green & Morrison 2016) and suffixes (e.g. Nilsson 2016). I will remain agnostic to this issue here. However, the two analyses have different consequences for how the choice between *ka* and *ta* is described: if the articles are clitics, they might be argued to be independent elements, and thus count as “associated words” – that is, possible agreement targets. If they are suffixes, they form parts of the noun itself –

⁵There is variation between and within speakers in the realization of singleton and geminate stops – see Bendjaballah & Le Gac (2019) for a discussion and a recent production experiment.

that is, they are a part of the gender controller, and therefore not agreement targets. Under such an approach, the choice between *ka* and *ta* is a matter of form-class, not gender. In line with this approach, Güldemann & Maniscalco (2016) call the initial consonant of the articles *thematic consonants* rather than agreement markers. The fact that the shape of the definite article is predictable from gender, and thus covaries with gender like subject-verb agreement does, just means that form-class and gender happen to correlate. To my knowledge, this correlation is a perfect one in root nouns. Whether this is true for compounds as well is an empirical question, and therefore, evidence from uncontroversial agreement targets (such as verbs) is necessary when determining compound gender, and thus subject-verb agreement is also investigated in the present study (see §3).⁶

2.1.2 Gender and tone

Another aspect of the gender system in Somali is that there is a correlation between gender and tone (Armstrong 1934), and there are many pairs like the ones provided in (6–7) (from Hyman 1981: 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.⁷

- (6) Masculine nouns, penultimate H
 - a. *ínan-ka* ‘the boy’
 - b. *qaálin-ka* ‘the young he-camel’
 - c. *daméér-ka* ‘the he-donkey’
- (7) Feminine nouns, final H
 - a. *inán-ta* ‘the girl’
 - b. *qaalín-ta* ‘the young she-camel’
 - c. *dameér-ta* ‘the she-donkey’

Under the agreement-based approach adopted here, this is a matter of form-class, not gender. While tone and gender correlate, the correlation is not perfect, as the examples in (8–9) illustrate (examples from Andrzejewski 1964; Hyman 1981; Saeed 1999).

⁶A complication in the Somali case is that some nouns have different definite articles in their singular and plural forms, a notion that has been called *gender polarity* (see e.g. Lecarme 2002; Lampitelli 2013; Nilsson 2016). However, other agreement targets do not show gender distinctions for plural nouns (Nilsson 2016), and so only singular forms of compounds are investigated in the present study.

⁷The tone-bearing unit is the mora, and only vowels are moraic in Somali.

- (8) Masculine nouns with a final H
 - a. *maroodi-ga* ‘the elephant’
 - b. *abtí-ga* ‘the uncle’
 - c. *waá-ga* ‘the time, dawn’
- (9) Feminine nouns with a penultimate H
 - a. *sábtí-da* ‘the Saturday’
 - b. *úgax-da* ‘the egg’
 - c. *sáddex-da* ‘the three’⁸

These data illustrate that tone and gender may vary independently of another. Therefore, I assume that tone is lexical, rather than introduced via morphological rules assigning tones to nouns based on their gender (which was proposed by Hyman 1981).⁹

2.2 Compounds, gender, and headedness

2.2.1 Types of nominal compounds in Somali

Nominal compounds in Somali come in different shapes (see Banti & Jama 2016; Puglielli 1989; Saeed 1999: 154–160). In the simplest case, there are two noun roots, as in (10). Noun-noun compounds are distinguished from noun-noun genitive constructions in that the former has one High tone and a single definite article at the right edge, while the latter has two High tones and two articles (11).¹⁰

- | | |
|--|---|
| (10) Noun-noun root compound
<i>gaari+fáras-ka</i>
car+horse-M.DEF
‘the horse cart’ | (11) Genitive construction
<i>búug-ga macállin-ka</i>
book-M.DEF teacher-M.DEF
‘the book of the teacher’ |
|--|---|

The present study aims to uncover how gender is assigned to noun-noun root compounds of the type in (10). There seem to be few noun-noun root compounds in Somali. Other morphologically complex types are illustrated in (12) (examples from Zorc & Osman 1993) and include compounds in which one member ends in *-e* or *-o* (12a) (which function variously as nominalizers and thematic vowels;

⁸Numerals are nouns in Somali, and numerals from 2–8 are feminine (Saeed 1999: 69).

⁹In line with this, loanword adaptations suggest that the tone-gender correlation is not fully productive (see Kaldhol 2017; Kaldhol & Stausland Johnsen forthcoming).

¹⁰The second noun has a genitive function here although there is no overt exponent of the genitive when the noun is definite.

see e.g. Le Gac 2016: 295) or is a derived noun (12b). Some nominal constructions are nominalized forms of adjectival compounds (12c). Many nominal compounds include a heterosemous root (a root that can function as either a noun or a verb) (12d). Sometimes the final member is marked with one of the suffixes *-eed/-aad/-ood* (12e), an attributive suffix, according to Banti (1986; 1988a).

(12) Other types of nominal compounds

- | | | | |
|----|--------------------------|----------------------|--------------------------|
| a. | <i>il+tuk-e</i> | eye+crow | ‘crow’s eye (plant sp.)’ |
| b. | <i>hadal+qor-aal</i> | talk+[write-NMLZ] | ‘protocol’ |
| c. | <i>madax+bannaan-i</i> | [head+free]-NMLZ | ‘independence’ |
| | cf. <i>madax+bannaan</i> | head+free | ‘independent’ |
| d. | <i>madax+xanuun</i> | head+pain/be in pain | ‘headache’ |
| e. | <i>libaax+bad-eed</i> | lion+[sea-ATTR] | ‘shark’ |

Compounds of these types are left aside for now because they are likely to function differently with respect to gender. For example, derivational suffixes and the vowels *-e/-o* assign their own gender and form-class (Saeed 1999). Similarly, heterosemous roots have been analyzed as nouns derived from verbs, with gender determined by the derivation process (see e.g. Saeed 1999: 149).

2.2.2 Compounds and headedness

The notion of head is often 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: 141–142). The examples in (13) illustrate that the compound gender is the same as the gender of the right member (the head), as indicated by the preceding definite article.

(13) Right-headed nominal compounds in Dutch (Booij 2002: 142)

Common gender		Neuter gender	
de soep	‘the soup’	het vlees	‘the meat’
de vlees+soep	‘the meat soup’	het soep+vlees	‘the soup meat’

In the case of Somali, conflicting generalizations have been proposed regarding gender and headedness, as reflected in the following quotes:

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

Its gender is determined by the head of the compound. That means that we can easily find a sequence $N_1 + N_2 + \text{Art}$ where N_1 , the head of the compound, is masculine, N_2 is feminine and the article is masculine, or vice versa. (Puglielli 1989: 7)

The notion of head is not a pre-theoretic concept: examples of authors that have argued against the notion and developed alternatives are Langacker (1987), working within the framework of Cognitive Linguistics, and Croft (2001), working within Radical Construction Grammar. Thus invoking the concept of head is not an explanation on its own, but only if it is tied to a particular framework or a set of criteria.¹¹ Bauer (2017: 40) argues that in the case of compounds, the notion splits into several subcriteria which do not apply to all compounds and do not always agree. The term of head may be useful for referring to a language-particular category (in the sense of Haspelmath 2018) in a given language if it can unify different phenomena, as is the case for compounds in Dutch. However, this is not the case for all languages, and there is a need for more in-depth research on compounds, both typologically and in particular languages (Scalise & Fábregas 2010; Guevara & Scalise 2009).

For the purposes of the present paper, I will distinguish between three phenomena with different properties: first, the LOCUS OF INFLECTION AND DERIVATION (the member to which derivational and inflectional affixes attach).¹² Second, the SEMANTIC HEAD is “the most contentful item that most closely denotes the same kind of thing that the whole constituent denotes” (Croft 2001: 259), and identified by the “IS A” criterion (Allen 1978) – for example, a *mailman* is A *man*.¹³ Third, the MORPHOSYNTACTIC HEAD is the member which determines the morphosyntactic properties of the compound (here, only gender is relevant). The morphosyntactic head is a useful notion for compounds if it allows one to predict the form of agreement targets (and thus compound gender). Note that derivational suffixes can be morphosyntactic heads under this criterion.

¹¹The defining properties of “heads” as well as its status in different theories of both syntax, morphology, and phonology have been debated, see Corbett et al. (1993); Zwicky (1985); Hudson (1987); and Moskal & Smith (2019).

¹²I assume that this is an edge phenomenon (Bauer 2017: 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.

¹³The semantic relation between the members in a compound is unstated. Many types of relations are possible, and I assume here that it cannot be reduced to one of *idiomatic* and *non-idiomatic*, or one of *endocentric* and *exocentric*. See Bauer (2008; 2017); Pepper (2020); Eiesland (2015); Jackendoff (2016) for more on this view.

Somali is largely a suffixing language, and the locus of inflection and derivation is the right edge, also in compounds. This was illustrated in (10) with the definite article: *gaari+fāras-ka* ‘the horse cart’, and in (12c) with a nominalizer: *madax+bannaan-i* ‘independence’. The position of the semantic head in compounds varies depending on the word class of the members involved. Noun-noun root compounds are semantically left-headed – for example, *gaari+fāras* ‘horse-cart’ is a type of *gaari* ‘cart’. Adjectival noun-adjective compounds are semantically right-headed: for example, *madax-bannaan* ‘independent (lit. head-free)’ is a type of *bannaan* ‘free’.¹⁴ Compounds in which the initial member is a noun and the second member is a heterosemous root are also semantically right-headed: for example, *madax+xanuun* ‘headache’ in (12d) above is a type of *xanuun* ‘pain’. See Banti & Jama (2016) for more examples and types.

Finally, the notion *morphosyntactic head* is useful if it allows us to predict the gender of the compound from a structural property (for example, you can predict the gender of compounds in Dutch from the gender of the final member). Judging from the examples in (14–15), collected from a dictionary, this does not apply to Somali. In (14), the definite article matches the gender of the final member. In (15), the definite article matches the gender of the initial member. Since all logical possibilities are attested, it is not the case that compound gender is consistently determined by either the initial or the final member, and therefore, one cannot predict the gender of the compound (indicated by the definite article) based on position.

- (14) Gender of compound = gender of final member (Zorc & Osman 1993)

<i>laf-ta</i>	F	+	<i>garab-ka</i>	M	→	<i>laf+garab-ka</i>	M
‘bone’			‘shoulder’			‘shoulder bone’	
<i>buug-ga</i>	M	+	<i>lacag-ta</i>	F	→	<i>buug+lacag-ta</i>	F
‘book’			‘money’			‘cashbook, ledger’	

- (15) Gender of compound = gender of initial member (Zorc & Osman 1993)

<i>laf-ta</i>	F	+	<i>dhabar-ka</i>	M	→	<i>laf+dhabar-ta</i>	F
‘bone’			‘back’			‘spine’	
<i>bur-ka</i>	M	+	<i>saliid-da</i>	F	→	<i>bur+saliid-ka</i>	M
‘flour’			‘oil’			‘fritter’	

A further issue in the Somali case is that for some compounds, different genders are reported by different scholars (16).

¹⁴Whether or not derivational suffixes, such as *-i* in *madax+bannaan-i* ‘independence’ can be semantic heads is less clear. It is not relevant for the present purposes and will not be discussed further here.

- (16) Different genders reported
bir ‘iron’ (F) + *danab* ‘electricity’ (M) → *birdanab* ‘magnet’
F (Saeed 1999: 158) or M (Zorc & Osman 1993: 41)

The question is how compound gender was determined by these authors. In the next section, I present novel data suggesting that the discrepancies arise from relying on the definite article as a cue to gender. A clearer pattern emerges when we look at subject-verb agreement.

3 Gender and tone assignment to noun-noun root compounds

In order to examine the gender of compounds, noun-noun compounds were elicited from a native speaker of Somali living in San Diego. She is originally from Mustahil, Ethiopia, and thus speaks a Central variety of what is typically called Northern Somali or Maxaatiri, but she has lived in multiple places in the Horn of Africa. All the remaining examples are provided by this speaker unless otherwise noted. The compounds and their members were produced in carrier phrases of the type *the X is good* or *the X is bad*. These phrases include both the definite article (masculine *ku* and feminine *tu*) and copula verb agreement (masculine *yahay*, feminine *tahay*). Under the assumption that gender is revealed by agreement (see §2.1.1), this carrier phrase provides evidence of compound gender. Information about the word list is provided in the appendix.

3.1 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 whether there is some other unrelated mechanism which assigns gender to compounds. Based on the agreement patterns in the following examples, the gender of the members does in fact seem to be the determining factor: when both members are masculine, the compound gender is also masculine, as in (17), and when both members are feminine, the compound gender is also feminine, as in (18).

- (17) 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)

(18) 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)

Unattested in the present study are feminine-feminine compounds with masculine agreement, and masculine-masculine compounds with feminine agreement. Compound gender is thus not independent of the gender of its members.

3.2 Compounds in which the members have different gender

When the members of a compound have different genders, different outcomes are attested. As we will see, what all of these outcomes have in common, is that subject-verb agreement is consistently predictable from the gender of the initial member of the compound. That is, masculine-feminine compounds control masculine subject-verb agreement (19), and feminine-masculine compounds control feminine subject-verb agreement (20). These examples also illustrate the first type of outcome, in which the definite article matches the gender of the initial member, even though it is linearly adjacent to the second member.

(19) 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), *saliídu* (F)
- b. Daacuun+calóol-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)

(20) 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. *shimbirtu* (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 (21). Subject-verb agreement matches the gender of the initial member, but unlike in (19) and (20), the definite article does not. Instead, it matches the gender of the second member.

(21) 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+diin-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), *diinku* (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 *diin-ka* ‘the turtle’, forms a minimal pair with a feminine noun: *diin-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 consulted speaker sometimes chooses different definite articles on different instances, even for the same compound (22). This variation is not attested in subject-verb agreement, which instead is consistently predictable from the gender of the initial member.

(22) Variable definite article, 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. *birtu* (F), *jírku* (M)

The generalization is thus that for this speaker, subject-verb agreement is consistently predictable from the gender of the initial member, while the definite article assignment is variable. Under the assumption that there is such a thing as compound gender, it is the initial member of noun-noun root compounds which determines compound gender, which in turn controls subject-verb agreement.¹⁵ The initial member of noun-noun root compounds does not only determine compound gender and thus functions as a morphosyntactic head, but it is also the semantic head. Following the criteria set up in §2.2.2, *laf+sakáar* ‘breast bone’ (22a) is a type of *laf* ‘bone’, hence *laf* (F) is the semantic head. *Fool+diin* ‘false labor pains (labor.pain-turtle)’ (21b) is metaphorically a type of *fool* ‘labor pains’, hence *fool* (F) is the semantic head.¹⁶

This leaves the question of why the definite article may vary. There does not seem to be any systematic patterns determining when the article reflects the gender of the initial member, and when it reflects the gender of the final member. That is, it could have been the case that the mismatches in gender cues from the verb and the article were predictable, e.g. from the type of semantic relation between the two members. However, this does not seem to be the case, as mismatches in gender cues are found across different semantic relations, e.g. both hyponymy, as in *laf+sakáar* ‘breast bone’ (22a), and metaphors, as in *fool+diin* ‘false labor pains (labor.pain-turtle)’ (21b). The next section discusses definite article assignment in more detail.

A challenge to the study of gender in noun-noun root compounds is that the number of F+F and M+F root compounds is less than the number of M+M and F+M compounds. The reasons for this is that compounds in which the final member is feminine usually have an attributive suffix (see §2.2.1), and these compounds are left aside for now since they are not root compounds. No agreement variation is found with M+F compounds in the present study (they are always masculine), but this may just be unattested because there are few such compounds to begin with.

3.3 Definite article assignment

While subject-verb agreement suggests that it is the initial member of noun-noun compounds that determines compound gender, the choice of definite article *ku* or

¹⁵The alternative would be that it is the initial member of the compound which controls agreement directly. In this case, agreement would need to be able to see the internal structure of the compound.

¹⁶The explanation I was provided for *fool+diin*, is that turtles are slow. The idea is that after a false labor, there are still months until the baby comes.

tu varies and is not a reliable gender cue. Instead, there seem 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 (23a), and another based on compound gender, in which it agrees with the compound as a holistic unit (23b).

(23) 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. Compound gender

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)

The linear adjacency strategy reflects the relationship between the compound construction and the construction *hungúri-ga*, which co-exists with the compound. When this strategy is used, the compound is thus treated as belonging to the same form-class as its final member. Another way of putting this is that the assignment of *-gu* here involves the *internal* assignment of a definite article to the rightmost noun, as illustrated with the structure in (24). This strategy contrasts with the compound gender strategy, which can be illustrated as in (25). Note that these formalizations are meant as illustrations only (the aim of this paper is not to provide a formal analysis).

(24) Linear adjacency

[F + M-*gu*]_F
[*wadda* + *hungúri-gu*]_F

(25) Compound gender

[F + M]_F -*du*
[*wadda* + *hungúri*]_F -*du*

One may ask how and why the speaker has formed a linear adjacency generalization for definite article assignment. Unlike the copula verb, the definite article is phonologically bound to the compound (see §2.1.1). This means that even when the choice between *ku* and *tu* reflects the compound gender, determined by the initial member, the choice of allomorph is phonologically determined by the final segment of the final member, which is the one it attaches to. This is illustrated in (26).

- (26) Bir+xárbi-du waa xún tahay.
 iron+war-F.DEF.NOM DECL bad 3SG.F.COP.PRS
 ‘The barrel cleaning rod is bad’ (cf. *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 lenition. 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).

3.4 Tone and gender assignment

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 §2.1.2, and illustrated in (6–7).

In noun-noun compounds, only the final member has a High tone, and the location of the tone is determined by the form-class of that member, no matter what the gender of the compound is. That is, while the initial member determines compound gender, 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 (27–28).

- (27) Masculine compound, but feminine tone pattern
búr (M) + *saliíd* (F) → *bur+saliíd* ‘fritter’ (M)
 (28) Feminine compound, but masculine tone pattern
láf (F) + *gárab* (M) → *laf+gárab* ‘shoulder bone’ (F)

What could have happened in this case, is that a new High tone is assigned to the compound based on compound gender, determined by the semantic head to the left (rendering *bursaliíd* and *lafgaráb*), but this is not the case.¹⁷

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

¹⁷Such shifts have 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).

formal analysis of the tone patterns (see Hyman 1981; Lampitelli 2013; Le Gac 2016; Green & Morrison 2016; Downing & Nilsson 2019 for a variety of proposals couched within different frameworks). As explained in §2.1.2, I view tone as a lexical property of nouns which is independent of, but correlating with gender. For example, *maroodí-gu* ‘the elephant’ is a masculine noun, but has a High tone on the final mora (Andrzejewski’s (1964) Declension 6b). In (17b) above, we saw an example of this word occurring as the final member of a compound (*fool+maroodí-gu* ‘the elephant’s tusk’). The compound also has a final High tone – that is, the tone pattern of the compound reflects the form-class of the final member, not its gender. These generalizations are summarized in (29).

(29) 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’
- b. In noun-noun root compounds, the location of that High tone is determined by the form class of the final member.
Examples: *gárab-ka* ‘the shoulder’, *maroodí-ga* ‘the elephant’

There is no attested variation in tone assignment in the present data set; the tone patterns of noun-noun compounds always correspond to the tone pattern of the final noun (as in *gárab* – *laf+gárab*). But as we have seen, there is variation in definite article assignment. I propose that this variation in definite article assignment reflects the article’s intermediate status as an instance of an agreement target (whose form is determined by the compound gender), and an instance of a phonologically bound element whose form is determined by the nominal form class (sensitive to the properties of the final member). If the article is assigned based on the form class of the closest member, as illustrated in (30), the choice of definite article matches the tone pattern. If the article is assigned based on compound gender, as illustrated in (31), there is a mismatch between the form class indicated by the definite article and the form class indicated by the tone pattern.

(30) Match between tone pattern and article (masculine *-ku*, penultimate H)

[F + M-*ku*]F

laf+gárab-ku

(31) Mismatch between tone pattern and article (feminine *-tu*, penultimate H)

[F + M]F *-tu*

laf+gárab-tu

cf. *gárab-ku*

While the compound gender strategy for definite article assignment (31) 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 (32).

- (32) Matching agreement, mismatching tone

Laf+gárab-tu *waa fíicán 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 tone pattern (a mismatch in form class), or a mismatch between definite article assignment and subject-verb agreement (a mismatch in gender). I propose that the attested variation reflects choosing between the two types of mismatches.

3.5 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 are the definite article and the copula verb. Second, when the two members have different genders (M+F or F+M), the initial member determines compound gender (as evidenced by the subject-verb agreement), but the final member determines form class (as evidenced by the tone pattern). The definite article assignment is variable, and is either in line with the compound gender (determined by the initial member, which also is the semantic head), or in line with the form-class of the final member.

4 Gender mismatches

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

4.1 Hybrid nouns

The compounds presented in §3.2 are reminiscent of hybrid nouns, that is, nouns whose agreement value varies according to agreement target (Corbett 1991; 2006; Enger & 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 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 (33).

- (33) A German hybrid noun (from Corbett 1991: 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 (see e.g. Corbett 1991: 228–232 for an overview). 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 formal gender of the noun (*syntactic* or *formal* agreement), often determined by a phonological or morphological property of the noun: for example, while German words ending in the suffix *-chen* are formally neuter, the referent of *Mädchen* ‘girl’ is female, and this mismatch is argued to be what allows for the variation in agreement patterns.

Although none of the compounds in the present study have a human referent precisely to avoid effects like this, there is a parallel to be drawn between hybrid nouns and Somali compounds with variable agreement patterns. The data presented in §3.2 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 semantic agreement and reflect the gender of the semantic head of the compound. However, as we have seen, it can also reflect the form-class of the final member (the linear adjacency strategy), and as a result, its gender. One can therefore think of the linear adjacency strategy as a type of formal agreement. This is comparable to the German example *Mädchen* ‘girl’ in (33), in which there is a conflict between referential gender and formal gender. There is, of course, a conceptual difference between agreement with a hybrid noun, in which the formal gender and the gender of the referent 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, and this is the topic of the next section. Note that conjunct agreement is different from compound agreement in another respect, namely that there are not just two nouns, but also two referents.

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 (34) illustrates both of these options within the same sentence.

- (34) Conjunct agreement in Russian (Corbett 2006: 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 one of the conjuncts. The predicate is plural, and Corbett (2006: 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).

The properties of conjunct agreement in Somali could in theory inform the agreement variation found with noun-noun compounds; however, this does not

seem to be the case. Based on the examples given in (35–36), subject-verb agreement with conjuncts shows 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. This holds true regardless of the order of the two noun phrases.¹⁸

(35) 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.’

(36) 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 the genitive construction, in which the verb consistently agrees with the initial noun, which is the head of the noun phrase (37). This construction thus forms a clearer parallel to noun-noun compounds – however, no variation in definite article assignment is attested.

(37) Genitive constructions (left-headed NPs)

- 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.’

¹⁸Note 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 (36).

- 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 more easily 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 it would have the structure Noun1 *and* Noun2-DEF, 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) (38).

- (38) 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: 69–72). For example, *labá-da* ‘the two’ is feminine, and *tobán-ka* ‘the ten’

is masculine (39). 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 (39).

(39) 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 women.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 (40), 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 (35). Unlike the example in (35), but like noun-noun compounds, the examples in (40) crucially have a single definite article. Notice that the choice of article consistently reflects the gender of the final member in these cases.¹⁹

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

¹⁹The choice between the two orders in (40) reflects regional variation (Nilsson 2018). However, some speakers, including the one I worked with, produce both orders.

5 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 is consistently predictable from the gender of the initial member (the semantic head) 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 compound gender, which is determined by the initial member (F+M-ta). The combination of consistent subject-verb agreement, but variable definite article assignment, results in mismatches in gender cues on these two agreement targets when the definite article is assigned in line with the linear adjacency strategy.

Since the choice of definite article varies, I argue that the gender indicated by the article is not a reliable cue to compound gender. This explains the discrepancies in previous literature (see §2). Since subject-verb agreement is consistent, and predictable from the gender of the initial member of noun-noun root compounds, I argue that it is this member which determines compound gender. In noun-noun root compounds, the initial member is also the semantic head (for example, *laf+dhábar* ‘spine, back-bone’ is a type of *láf* ‘bone’).

The agreement patterns found with noun-noun compounds in Somali when the gender of the two members do not match are reminiscent of those found with other types of gender mismatches across languages, e.g. with hybrid nouns and conjunct agreement (§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 argue that the agreement variation is the result of at least two factors operating in conjunction: first, while the semantic head is to the left in noun-noun compounds, the locus of inflection is on the right edge, such that the definite article is phonologically bound to the final member. This means that even when the article reflects compound gender, linear adjacency matters: its allomorph is determined by the final segment of the final member (§3.3).

Future research should investigate agreement patterns with compounds in which there is a heterosemous root, such as *madax+xanuun* ‘headache’ (§2.2.1). The analysis proposed here predicts no agreement variation with such compounds: since they are semantically right-headed (§2.2.2), and since the semantic head is also the morphosyntactic head and determines compound gender, the two strategies for definite article assignment are not in conflict in these cases. Rather, the linear adjacency strategy and the compound gender strategy both target the final member when the compound is right-headed.

Second, the mismatch between subject-verb agreement and definite article assignment is paralleled by a mismatch in gender and tone pattern frequently found in compounds (§3.4). In noun roots, there is a correlation between gender and tone, such that most masculine nouns have a penultimate High tone (*gárab* ‘shoulder’), and most feminine Nouns have a final High tone (*salííd* ‘oil’). As explained in §2.1.2, the tone pattern is a matter of nominal form-class, not gender. In noun-noun compounds, it is the initial member which determines compound gender, but the final member which determines the tone pattern. If the two members have different genders, the result is either a masculine compound with a final rather than a penultimate High tone (M+F *bur+salííd* ‘fritter’ (M)), or a feminine compound with a penultimate rather than final High tone (F+M *laf+gárab* ‘shoulder bone’ (F)). That is, there will either be a mismatch between definite article assignment and the tone pattern, or a mismatch between definite article assignment and subject-verb agreement, and as argued in §3.4, the attested variation in definite article assignment seems to reflect a choice between the two types of mismatches.

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. Furthermore, to achieve a full understanding of the relationship between gender and form-class and how this relates to headedness, other aspects of form-class in compounds should be investigated as well, most notably case marking and plural marking.

Abbreviations

All abbreviations (except H) in the paper follow the Leipzig Glossing Rules. H refers to high tone.

Acknowledgements

I would like to thank my consultants and language teachers in San Diego and Oslo for sharing their time, culture and language with me, with a special thanks to the speaker who contributed the data to the present study. I am grateful to Morgan Nilsson for sharing his list of Somali compounds. For useful comments and suggestions on previous versions of this work, I would like to thank (in alphabetical order) Farrell Ackerman, Georgio Banti, Gabriela Caballero, Laura Downing, Christopher Green, Morgan Nilsson, Martin Orwin, Sharon Rose, and two

anonymous reviewers. Furthermore, the work has benefited from feedback from my colleagues at UC San Diego, and the audiences at *Somali workshop: linguistics and geography 2018*, the *50th Annual Conference on African Linguistics*, and the poster session at the *LSA Linguistic Institute 2020*. All mistakes are my own.

Appendix

The word list used for elicitation consists of 86 noun-noun root compounds which were compiled based on a list of 1000 Somali compounds courtesy of Morgan Nilsson, as well as dictionaries (Mansuur & Puglielli 2012; Zorc & Osman 1993), Saeed's (1999) grammar, Caney's (1984) work on novel words in the Somali vocabulary, and personal field notes. Only noun-noun compounds in which both members are roots were included in the present study. Certain types of compounds were left aside for the present purposes because they have semantic or grammatical properties that potentially override any structural gender assignment patterns in noun-noun compounds that the present study aims to uncover: for example, compounds that are not root compounds are left aside because derivational suffixes assign their own gender (see §2.2.1). Further examples include proper nouns (such as names of languages, e.g. *Af+Soomaali* 'the Somali language') and collective nouns (*askar+maroodi* 'poor young boys (lit. soldiers-elephant)'), and compounds in which one of the members is not a prototypical noun, such as a numeral (e.g. *afar+gees* 'square (lit. four-side)'), or a color term (e.g. *cir+guduud* 'the times of the day in which the sky is red (lit. sky-red)'). Compounds in which one of the members is a plural or collective noun were also left aside (e.g. *xoolo-madax* 'capital, money assets (lit. livestock/property-head)'). Finally, because animacy and semantic gender are likely to play a role, compounds with a human referent were also left out of the present study (e.g. *af-gaab* 'quiet, polite person (lit. language/mouth-shortness)').

References

- Allen, Margaret Rees. 1978. *Morphological investigations* (Ph. D. Dissertation). University of Connecticut.
- Andrzejewski, Bogumil W. 1964. *The declension of Somali nouns*. London: School of Oriental & African Studies, University of London.
- Armstrong, Lilius E. 1934. The Phonetic Structure of Somali. *Mitteilungen des Seminars für Orientalische Sprachen zu Berlin* 37(3). 116–161.

- Banti, Giorgio. 1986. Review of serzisko, ausdrück der possessivität im somali. *Sonderdruck aus Kratylós* 31. 161–168.
- Banti, Giorgio. 1988a. “Adjectives” in East Cushitic. In Marianne Beckhaus-Gerst & Fritz Serzisko (eds.), *Papers from the International Symposium on Cushitic and Omotic Languages, Cologne 1986*, 204–259. Hamburg: Helmut Buske Verlag.
- Banti, Giorgio. 1988b. Two Cushitic systems: Somali and Oromo nouns. In Harry van der Hulst & Norval Smith (eds.), *Autosegmental studies on pitch accent*, 11–50. Dordrecht: Foris Publications.
- Banti, Giorgio. 2016. Tonal sandhi in Northern Somali. (Paper presented at the Colloquium on African Languages and Linguistics 2016, Leiden University).
- Banti, Giorgio & Jama Musse Jama. 2016. Compounds and Multi-word Expressions (MWEs) in Somali. Ms.
- Bauer, Laurie. 2008. Exocentric compounds. *Morphology* 18. 51–74.
- Bauer, Laurie. 2017. *Compounds and compounding*. Cambridge: Cambridge University Press.
- Bendjaballah, Sabrina & David Le Gac. 2019. The acoustics of Somali voiced stops. Workshop on Somali Grammar, University of Gothenburg.
- Booij, Geert. 2002. *The morphology of Dutch*. Oxford: Oxford University Press.
- Caney, John Charles. 1984. *The modernisation of Somali vocabulary, with particular reference to the period from 1972 to the present*. Hamburg: Helmut Buske Verlag.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Corbett, Greville G., Norman M. Fraser & Scott McGlashan. 1993. *Heads in grammatical theory*. Cambridge: Cambridge University Press.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Di Garbo, Francesca. 2014. *Gender and its interaction with number and evaluative morphology: An intra- and intergenealogical typological survey of Africa*. Stockholm University. (Doctoral dissertation).
- Di Garbo, Francesca. 2016. Exploring grammatical complexity crosslinguistically: The case of gender. *Linguistic Discovery* 14. 46–85.
- Downing, Laura J. & Morgan Nilsson. 2019. Prosodic restructuring in Somali nominals. In Samson Lotven, Silvina Bongiovanni, Phillip Weirich, Robert Botne & Samuel Gyasi Obeng (eds.), *African linguistics across the disciplines: Selected papers from the 48th Annual Conference on African Linguistics*, 125–142. Berlin.
- Eiesland, Eli-Anne. 2015. *The semantics of Norwegian noun-noun compounds*. University of Oslo. (Doctoral dissertation).

- Enger, Hans-Olav & Greville G. Corbett. 2012. Definiteness, gender, and hybrids: Evidence from Norwegian dialects. *Journal of Germanic Linguistics* 24. 287–324.
- Godon, Elsa. 1998. *Aspects de la morphologie nominale du somali: La formation du pluriel*. Université de Paris VII. (MA thesis).
- Green, Christopher R. & Michelle E. Morrison. 2016. Somali wordhood and its relationship to prosodic structure. *Morphology* 26(1). 3–32.
- Guevara, Emiliano & Sergio Scalise. 2009. Searching for universals in compounding. In Sergio Scalise, Elisabetta Magni & Antonietta Bisetto (eds.), *Universals of language today*, 101–128. Dordrecht: Springer.
- Güldemann, Tom & Ines Fiedler. 2019. Niger-Congo “noun classes” conflate gender with deriflection. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: General issues and specific studies*, vol. 1, chap. 5, 95–146. Berlin: Language Science Press.
- Güldemann, Tom & Samuele Maniscalco. 2016. A new approach to gender in Somali. Paper presented at the 22. Afrikanistentag, Humboldt Universität zu Berlin.
- Haspelmath, Martin. 2018. How comparative concepts and descriptive linguistic categories are different. In Daniël Olmen, Tanja Mortelmans & Frank Brisard (eds.), *Aspects of linguistic variation*, 83–114. Berlin: De Gruyter.
- Hockett, Charles F. 1958. *A course in modern linguistics*. New York: Macmillan.
- Hudson, Richard A. 1987. Zwicky on heads. *Journal of Linguistics* 23. 109–132.
- Hyman, Larry M. 1981. Tonal accent in Somali. *Studies in African Linguistics* 12(2). 169–203.
- Jackendoff, Ray. 2016. English noun-noun compounds in Conceptual Semantics. In Pius ten Hacken (ed.), *The semantics of compounding*, 15–37. Cambridge: Cambridge University Press.
- Kaldhol, Nina Hagen. 2017. *The Oslo dialect of Somali: Tonal adaptations of Norwegian loanwords*. University of Oslo. (MA thesis).
- Kaldhol, Nina Hagen & Sverre Stausland Johnsen. Forthcoming. There is no “one High tone per word” rule in Somali. In Björn Köhnlein & Pavel Iosad (eds.), *Synchronic and diachronic aspects of tonal accent*. Oxford: Oxford University Press.
- König, Christa. 2006. Marked nominative in Africa. *Studies in Language* 30. 655–732.
- Lampitelli, Nicola. 2013. The decomposition of Somali nouns. *Brill’s Journal of Afroasiatic Languages and Linguistics* 5. 117–158.
- Lampitelli, Nicola. 2019. On nominal declensions and syntactic case. Paper presented at the Workshop on Somali Grammar, University of Gothenburg.

- Langacker, Ronald. 1987. *Foundations of cognitive grammar: Theoretical prerequisites*. Vol. 1. Stanford, CA: Stanford University Press.
- Le Gac, David. 2016. Somali as a tone language. In Jon Barnes, Alejna Brugos, Stefanie Shattuck-Hufnagel & Nanette Veilleux (eds.), *Proceedings of Speech Prosody 2016*, vol. 8, 292–296. Boston University.
- Lecarme, Jacqueline. 2002. Gender “polarity”: Theoretical aspects of Somali nominal morphology. In Paul Boucher & Marc Plénat (eds.), *Many morphologies*, 109–141. Somerville, MA: Cascadilla Press.
- Mansuur, Cabdalla Cumar & Annarita Puglielli. 2012. *Qaamuuska Af-Soomaaliga*. Rome: Rome TrE-Press.
- Moskal, Beata & Peter W. Smith. 2019. The status of heads in morphology. In *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. DOI: 10.1093/acrefore/9780199384655.013.583.
- Mous, Maarten. 2012. Cushitic. In Zygmunt Frajzyngier & Erin Shay (eds.), *The Afroasiatic languages*, 342–422. Cambridge: Cambridge University Press.
- Nilsson, Morgan. 2016. Somali gender polarity revisited. In Doris Payne, Sara Pacchiarotti & Mokaya Bosire (eds.), *Diversity in African languages: Selected papers from the 46th Annual Conference on African Linguistics*, 451–466. Berlin: Language Science Press.
- Nilsson, Morgan. 2018. *Somaliska skriftspråket: En hundraåring som sprudlar av energi*. Bokmässan i Göteborg.
- Pepper, Steve. 2020. *The typology and semantics of binominal lexemes: Noun-noun compounds and their functional equivalents*. University of Oslo. (Doctoral dissertation).
- Puglielli, Annarita. 1989. Compounds in Somali. Manuscript.
- Saeed, John I. 1999. *Somali*. Vol. 10. Amsterdam: John Benjamins.
- Scalise, Sergio & Antonio Fábregas. 2010. The head in compounding. In Sergio Scalise & Irene Vogel (eds.), *Cross-disciplinary issues in compounding*, 109–126. Amsterdam: John Benjamins.
- Steele, Susan. 1978. Word order variation: A typological study. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language: Syntax*, 585–623. Stanford: Stanford University Press.
- Zorc, R. David & Madina M. Osman. 1993. *Somali - English dictionary with English index*. 3rd edn. Kensington: Dunwoody Press.
- Zwicky, Arnold M. 1985. Heads. *Journal of linguistics* 21. 1–29.