Familiar Definite Marking in Magahi[1]

AIDAN SHARMA, *Rutgers University*

ABSTRACT

This paper investigates the nominal suffix *-waa* in Magahi, an Eastern Indo-Aryan language. Existing accounts of *-waa* vary from analyzing it semantically in terms of familiarity and non-honorificity (Alok 2022), diminutivity (Atreya & Sinha 2020), or definiteness (Kumar 2020) and syntactically in terms of whether it projects a head in the nominal spine (Kumar 2020) or not (Alok 2012, 2022). I argue that *-waa* is a familiar definite marker, similar to the German strong article (Schwarz 2009) and Akan familiar article (Arkoh & Matthewson 2013), with additional presuppositions of non-uniqueness (Owusu 2022) and non-honorificity. Additionally, I argue that *-waa* can either be generated as the definite allomorph of the general classifier (Kumar 2020) and undergo CLF to D movement, or be be base generated in D (Simpson 2005).

1 Introduction

This paper proposes a semantic and syntactic analysis of the nominal suffix *-waa* in Magahi[2] (Eastern Indo-Aryan) as a familiar definite marker. Earlier studies on *-waa* differ in their approaches to both its semantics and syntax. Alok (2012, 2022) analyzes *-waa* as a nominal particle that encodes familiarity as a presupposition and non-honorificity as expressive content (Potts 2007). In particular, Alok maintains that the main contribution of *-waa* is not definiteness and *-waa* is not projected as a head. Atreya & Sinha (2020) treat *-waa* as a diminutive marker that can convey endearment, derogation, and specificity. However, Kumar (2020) focuses on Magahi's status as a numeral classifier language and argues that *-waa* is the definite allomorph of the classifier used in "bare classifier" phrases similar to the bare classifier constructions in Cantonese (Cheng & Sybesma 1999). Lahiri (2021) also glosses *-waa* as a classifier.

This paper proposes that *-waa* is a familiar definite marker similar to the German strong article (Schwarz 2009), but with the relevant notion of familiarity including both strong and weak familiarity as defined by Roberts (2003). Syntactically, my analysis is most similar to the one in Kumar (2020). However, I argue that definites with *-waa* are not bare classifier phrases, but full DPs. In particular, I argue that *-waa* can originate as the head of a classifier projection and undergo CLF to D head movement or that *-waa* can be base generated in D, presumably through reanalysis of frequent CLF to D movement (Simpson 2005).

---

[1]All uncited data in this paper comes from elicitations done by the author with five native speakers of Magahi, three from the Nawada district of Bihar and two from the Jehanabad district.

[2]Magahi is primarily spoken in Bihar and is considered one of the three main Bihari languages, along with Maithili and Bhojpuri. The number of speakers is estimated to be between 9 and 12 million (Verma 2003).

2 Familiarity and Uniqueness

Whether definite descriptions require uniqueness (Frege 1892; Russell 1905) or familiarity (Heim 1982) has been a long-standing debate in linguistics and philosophy. The following examples (Schwarz 2013: 535 based on Hawkins 1978) seem to vary in whether they require uniqueness or familiarity.

(1)   a.   **Anaphoric**
           John bought *a book* and a magazine.
           The book was expensive.

      b.   **Immediate Situation Uniqueness**
           the desk   (**Context:** uttered in a room with exactly one desk)

      c.   **Larger Situation Uniqueness**
           the prime minister   (**Context:** uttered in the UK)

      d.   **Bridging**[3]
              i. **Producer-Product**
                 John bought *a book*.
                 The author is French.

             ii. **Part-Whole**
                 John's hands were freezing as he was *driving* down the street.
                 The steering wheel was bitterly cold and he had forgotten his gloves.

English does not distinguish between these uses of definites and *the* is used for all of them. However, Schwarz (2009) showed that some languages do differentiate between these types of definites and have dedicated articles for the concepts of uniqueness and familiarity.

2.1   German Weak and Strong Articles

Schwarz (2009) shows that there are two types of definite articles in German, which are differentiated by their their phonological status following a preposition. The *weak article* in (2) contracts after a preposition, but the *strong article* in (3) does not.

(2)   *Hans ging **zum** Haus*   (3)   *Hans ging **zu dem** Haus*
      Hans went **to_the**$_{weak}$ house   Hans went **to the**$_{strong}$ house

      'Hans went to the house'.   'Hans went to the house'.

Schwarz shows that these two articles not only have different phonological behavior, but also different semantics. The weak article is used for unique definites, and the strong article is used for familiar definites. This can be seen with the different definites described by

---

[3]Hawkins (1978) uses the term associative anaphora. The term bridging is due to Clark (1975).

Hawkins (1978). While anaphoric definites require the strong article (4), both immediate (5) and larger situation (6) uniqueness definites require the weak article.

(4) **Anaphoric**

> *In der New York Bibliothek gibt es* **ein Buch** *über Topinambur. Neulich*
> in the New York library exists EXPL **a book** about topinambur recently
>
> *war ich dort und habe* **{#im / in dem}** **Buch** *nach einer Antwort*
> was I there and have **{#in_the$_{weak}$ / in the$_{strong}$}** **book** for an answer
>
> *auf die Frage gesucht, ob man Topinambur grillen kann.*
> to the question searched whether one topinambur grill can

'In the New York public library, there is **a book** about topinambur. Recently, I was there and searched **in the book** for an answer to the question of whether one can grill topinambur'. (Schwarz 2009: 30)

(5) **Immediate Situation Uniqueness**

> *Das Buch, das du suchst, steht* **{im / #in dem}** *Glasschrank.*
> the book that you look.for stands **{in_the$_{weak}$ / #in the$_{strong}$}** glass.cabinet

'The book that you are looking for is **in the glass cabinet**.' (Schwarz 2009: 39)

(6) **Larger Situation Uniqueness**

> *Armstrong flog als erster* **{zum / #zu dem}** *Mond.*
> Armstrong flew as first.one **{to_the$_{weak}$ / #to the$_{strong}$}** moon

'Armstrong was the first one to fly **to the moon**.' (Schwarz 2009: 40)

Schwarz claims the difference between the two is the presence of an anaphoric index, which blocks the contraction of the strong article. As for the semantics, the weak article takes a situation and a property as arguments. It presupposes there is a unique individual satisfying the property in the given situation and returns that individual. The strong article functions similarly to the weak article but also takes an index argument of type *e*, semantically equivalent to a pronoun. Again there is a uniqueness presupposition and the article returns a definite description but, for the strong article, the presupposition and definite description include a statement identifying the referent with the index argument. The structures and denotations for the two articles are given below (Schwarz 2019: 12).

(7) a. $[_{DP}[the_{weak} \, s]NP]$

b. $[\![the_{weak}]\!]^g = \lambda s_r \lambda P_{<e,st>} : \exists!x[P(x)(s_r)].\iota x[P(x)(s_r)]$

(8) a. $[_{DP} i[[the_{strong} s]NP]]$

b. $[\![the_{strong}]\!]^g = \lambda s_r \lambda P_{<e,st>} \lambda y : \exists!x[P(x)(s_r) \wedge x = y].\iota x[P(x)(s_r) \wedge x = y]$

Interestingly, the two types of bridging examined by Schwarz also make a distinction between the two articles.

(9) **Producer-Product Bridging**

Das	***Theaterstück***	*missfiel*	*dem*	*Kritiker*	*so*	*sehr,*	*dass*	*er*	*in*	*seiner*
the	**play**	displeased	the	critic	so	much	that	he	in	his

Besprechung	*kein*	*gutes*	*Haar*	***{#am / an dem}***		***Autor***	*ließ*
review	no	good	hair	**{#on_the**$_{weak}$ **/ on the**$_{strong}$**}**		**author**	left

'**The play** displeased the critic so much that he tore **the author** to pieces in his review.'   (Schwarz 2009: 53)

(10) **Part-Whole Bridging**

Der	***Kühlschrank***	*war*	*so*	*groß,*	*dass*	*der*	*Kürbis*	*problemlos*
the	**fridge**	was	so	big	that	the	pumpkin	without.a.problem

{im / #in dem}		***Gemüsefach***	*untergebracht*	*werden*	*konnte.*
{in_the$_{weak}$ **/ #in the**$_{strong}$**}**	**crisper**		stowed	be	could

'**The fridge** was so big that the pumpkin could easily be stowed **in the crisper**.' (Schwarz 2009: 52)

In producer-product bridging, the product (the play) behaves as if it introduces an antecedent for the producer (the author), and thus the strong article is used with *Autor* 'author'. Meanwhile in part-whole bridging, the part (the crisper) behaves as if it is unique relative to the whole (the fridge) and the weak article is used with *Gemüsefach* 'crisper'.

2.2   Akan Familiar Article

Arkoh & Matthewson (2013) investigate the semantics of the definite article *nʊ* in Akan (Kwa; Niger-Congo).[4] They argue it is a familiar definite article (glossed as FAM) and has the semantics of the German strong article.

(11) **Context**: beginning of conversation

mʊ̀-tɔ́-ɔ̀		*èkùtú*	*(*nʊ́).*	*èkùtú*	**(nʊ́)*	*yɛ̀*	*dɛ̀w*	*pápá.*
1SG.SUBJ-buy-PST	orange	FAM	orange	FAM	be	nice	good	

'I bought an orange. The orange was really tasty.'   (Arkoh & Matthewson 2013: 2)

---

[4]Arkoh & Matthewson also show that *nʊ* can be used as a third person pronoun and a dependent clause marker, but I will only mention its use as a familiar article here.

Example (11) above shows that *nʊ* cannot be used with indefinites, even if specific, but is required for anaphoric definites. However, *nʊ* does not strictly require linguistic antecedence. It is compatible with nouns that are familiar to all discourse participants even if not introduced in the discourse.

(12) **Context:** a parent talking to his/her spouse about their children

    mbofra nó   wɔ́ dan   nɔ́   mu
    children FAM be   room FAM in

    'The children are in the room.'   (Arkoh & Matthewson 2013: 7 from Saah 1994: 152)

While *nʊ* is compatible with these familiar definites, it is incompatible with larger situation uniqueness definites. These are expressed with bare nouns.

(13) *kwámì nyá-à   kràtàá fí-ì   ègyá krɔ́nkrɔ́n póp hɔ́*
    Kwame get-PST letter from-PST father holy   pope there

    'Kwame got a letter from the holy father Pope'.   (Arkoh & Matthewson 2013: 11)

(14) *amstɔ́ŋ   nyí nyímpá áà   ó-dzí-ì   kán tú-ù   kɔ́-ɔ̀   ɔsìrán dʊ̀*
    Armstrong is   person REL 3SG.SUBJ-eat-PST first fly-PST go-PST moon top

    'Armstrong was the first person to fly to the moon'.   (Arkoh & Matthewson 2013: 11)

Because *nʊ* seems to require familiarity rather than uniqueness, Arkoh & Matthewson (2013) propose that it is the Akan equivalent of the German strong article. However, Owusu (2022) notes that adopting the semantics associated with the German strong article for *nʊ* does not rule out its use with larger situation uniqueness definites. While DPs such as *the Pope* or *the first man to fly to the moon* are necessarily unique, they can also be familiar. For example, in a conversation where the Pope has already been mentioned, *nʊ* would seem to be licensed with *ègyá krɔ́nkrɔ́n póp* 'holy father Pope' based on the other Akan data. Additionally, Akan has no weak article that might rule out the strong article on the basis of competition. Thus, Owusu proposes that *nʊ* requires an additional presupposition of non-uniqueness, usually associated with demonstratives (Dayal & Jiang 2023).

(15) **Non-uniqueness**
$\exists s' s \leq s' \land |\{x \mid P(x)(s')\}| > 1$

Recall that the strong article takes a situation argument s_r and a property argument $P_{<e,st>}$. Then, (15) says that there must be a larger situation containing the situation argument of the determiner, in which there is more than one individual that satisfies *P*. In other words,

while the familiar determiner picks out an individual that is unique/maximal relative to *P* in the given situation, there do exist other individuals that satisfy *P*; they are just not in the given situation. Larger situation uniqueness definites cannot satisfy this presupposition and are therefore correctly predicted to be incompatible with *nʊ*.

3 Magahi *-waa*

Before discussing the semantics and syntax of *-waa*, a basic overview of its phonology is necessary. The suffix *-waa* attaches only to nouns, and the sequence forms a single prosodic unit. This can have effects on stress and vowel length, such as in the example below.

(16)

	book	book-*waa*
Romanization	*kitaab*	*kitabwaa*
IPA	[kɪ.t̪ɑːb]	[kɪ.t̪əb.wɑː]

Besides this interaction with stress and vowel length, *-waa* also has five allomorphs: *-waa, -aa, -(i)yaa, -(i)yãã, -maa* (Atreya & Sinha 2020). While there are general patterns that determine the use of each allomorph depending on the phonology of the root noun (e.g., *-yaa* occurs after nouns ending in *ii* and *-maa* occurs after nouns ending in a nasal consonant), several of them can be used in identical environments and there is a large amount of inter- and intra-speaker variation. For example, my consultants used three allomorphs for *-waa* with the word for 'book': *kitab-waa, kitab-aa, kitab-iyaa*. One speaker even used all three at different times.[5] Despite this slightly blurry set of facts, Atreya & Sinha and Alok agree that *-waa* is the general form, and I will gloss all these allomorphs as -WAA in the Magahi examples in this paper.

3.1 Semantics of *-waa*

I am taking definiteness to be the primary contribution of *-waa* since its use forces a definite interpretation. In particular, nouns suffixed with *-waa* cannot be interpreted as indefinites (17), generics (18), or kinds (19) (cf. Alok 2012: 46). Even if the indefinite in (17) is specific, *-waa* is disallowed.

(17) *(ek tho) bilai-(#yaa)*
     one CLF cat-(#WAA)

     'a/one cat'

(18) *chiṛai-(#waa) uṛa hai*
     bird-(#WAA) fly AUX

     'Birds fly.'

(19) *dainasor-(#waa) bilupt ho gelai*
     dinosaur-(#WAA) extinct be went

     'Dinosaurs are extinct.'

---

[5]See Atreya & Sinha (2020) for more information on the allomorphs of *-waa* and Alok (2022) for the possibilities of different allomorphs on the same noun.

Additionally, Löbner's (1985) diagnostic shows that *-waa* is not a demonstrative.

(20) #*laik-waa sutiit hai auu laik-waa na sutiit hai*
     boy-WAA sleeping is and boy-WAA not sleeping is

   'The boy is sleeping and the boy is not sleeping.'   (Alok 2012: 25)

Nouns suffixed with *-waa* are number neutral, but a plural suffix *-(a)n* can be suffixed to nouns as well, in which case they must be interpreted as plural.

(21) *kutt-waa*   (22) *kutt-waa-n*
     dog-WAA   dog-WAA-PL

   'the dog(s)'   'the dogs'

However, I will not discuss number marking and will focus on singular definites.

3.1.1 Anaphoricity

Magahi *-waa*, like the Akan article and German strong article, is obligatory on anaphoric definites. Note that a demonstrative can optionally co-occur with *-waa*.

(23) *kal   ham ek tho kutta dekhaliai. (uu) kutt-#(waa) bari sundar   halai*
     yesterday 1SG one CLF dog   saw   DEM dog-WAA   very beautiful was

   'Yesterday I saw a dog. The/that dog was very beautiful.'

The narrative sequence below, modeled after the example in Jenks (2018: 510), shows that *-waa* is obligatory for anaphoric definites regardless of syntactic position.[6]

(24) *class me e-go   laṛaka auu laṛakii hai*
     class in one-CLF boy   and girl   is

   'There is a boy and a girl in class.'

   a. *ham (uu) laṛak-#(waa) ke   kal   milaliai*
      1SG DEM boy-WAA   ACC yesterday met

      'I met the/that boy yesterday.'

   b. *ham (uu) laṛak-#(waa) ke   khatir e-go   upahaar le   liye hai*
      1SG DEM boy-WAA   GEN for   one-CLF gift   take bring AUX

      'I'm bringing a gift for the/that boy.'

---

[6]Additionally, see Alok (2012: 29-32) for arguments that *-waa* is not a topic marker.

c. *(uu) laṛak-#(waa) biis saal ke lago hai*
   DEM boy-WAA   twenty year GEN seem AUX

   'The/that boy looks 20 years old.'

d. *hamra nā̃ĩ lago hai ki (uu) laṛak-#(waa) bahoot interesting hai*
   1SG.OBL NEG seem AUX COMP DEM boy-WAA   very interesting is

   'I don't think that the/that boy is very interesting.'

Finally, donkey definites also require *-waa* on the definite NP.

(25) a. *sabhe kisaan jekra paas gadha hai, uu gadha-#(waa) ke maaro*
   every farmer REL.OBL near donkey is 3SG donkey-WAA ACC beat
   hai
   AUX

   'Every farmer who has a donkey beats the donkey.'

b. *agar kisaan ke paas gadha hai, to uu gadha-#(waa) ke maaro*
   if farmer GEN near donkey is then 3SG donkey-WAA ACC beat
   hai
   AUX

   'If a farmer has a donkey, he beats the donkey.'

This further highlights the connection between *-waa* and familiarity (Jenks 2015b).

3.1.2 Non-uniqueness

In addition to familiarity, the incompatibility of *-waa* with larger situation uniqueness definites shows that it encodes non-uniqueness.[7]

(26) *chaand-(#waa) uuglai*
   moon-(#WAA) rose

   'The moon rose.'

(27) *suuraj-(#waa) puurab me uugo*
   sun-(#WAA) east in rise
   hai
   AUX

   'The sun rises in the east.'

---

[7]Simpson & Biswas (2016: 11) and Ushasi Banerjee (p.c.) report that definiteness marking with the classifier *Ta* in Bangla is possible for larger situation uniqueness definites such as 'moon' in episodic sentences but not generic ones. Similar facts are reported for Akan (Comfort Ahenkorah p.c.) and Cantonese (Ka-Fai Yip & Margaret Chui Yi Lee p.c.). I have not explored this fully in Magahi, but one of the three speakers I checked with, despite his initial judgement that it was infelicitous, said that *-waa* was acceptable with *chaand* 'moon' in (26). However, that speaker did not find *-waa* acceptable with *suuraj* 'sun' in (27) at all. More research is needed, but it is interesting that (26) has an episodic interpretation while (27) has a generic one.

(28) *amerika ke raaspati-(#yaa) paagal hai*
America GEN president-(#WAA) crazy is

'America's president is crazy.'

Superlatives, which are necessarily unique, are also incompatible with *-waa*.

(29) a. *pahala aadamii-(#yaa) chaand par*
first man-(#WAA) moon on

'the first man on the moon'

b. *duniya ke sab se tej aadamii-(#yaa) se milai ke mun hai*
world GEN all from smart man-(#WAA) from find PRT want AUX

'I want to meet the world's smartest man.'

Recently, however, it has been argued that certain classifier languages do not display a true unique/familiar dichotomy like the one in German. Instead, Yip et al. (2023) argue on the basis of Cantonese and Bangla that the lack of definite marking on larger situation uniqueness definites is not because the definites require non-uniqueness. Rather, Yip et al. argue that the bare nouns used in these instances behave like quasi-names, such as *Mom* in English. While Yip et al. present a convincing account of the Cantonese and Bangla data, I argue that this cannot extend to Magahi. In fact, Magahi *-waa* can also occur with proper and quasi-names, but only for people familiar to you of equal/lower social status.

(30) *ratan-maa*
Ratan-WAA

'Ratan' (referring to a friend or younger known person named Ratan)

(31) *bhai-waa*
brother-WAA

'brother' (referring to a younger brother) (Alok 2022: 1)

This behavior is not unusual. The use of definite elements with names is attested cross-linguistically e.g., in Greek, Maori (Anderson 2004). Additionally, Magahi has allocutive agreement that encodes honorificity in the clausal domain (Alok 2020, 2021; Alok & Baker 2022), so it is not entirely surprising that honorificity could be encoded in the nominal domain as well. However, while this shows that *-waa* is compatible with quasi-names, it could still be the case that larger situation uniqueness definites are treated as honorific/quasi-names of high status. To see that this is not the case, it is useful to note that *-waa* can be added to (quasi-)names of social superiors to show disrespect/contempt (Alok 2022). Nevertheless, in such cases, my consultants still consider *-waa* to be infelicitous on larger situation uniqueness definites.[8]

---

[8]I would like to thank an anonymous reviewer for suggesting to test this.

(32) *obaama ke baad ke halai raaspati? uu raaspati-(#yaa) buṛabak hai*
Obama GEN after who was president DEM president-(#WAA) idiot is

'Who was the president after Obama? That president is an idiot."

Yet, in a hypothetical scenario where our solar system has multiple suns or the Earth has multiple moons, my consultants say it would be felicitous to use *-waa* with *suuraj* 'sun' and *chaand* 'moon'. Thus, I maintain that non-uniqueness is a presupposition of *-waa*. I take the restrictions on the use of *-waa* with (quasi-)names to indicate that it also has a presupposition of non-honorificity (but see Alok 2022 who treats the non-honorificity of *-waa* as expressive content). It is this presupposition of non-honorificity that gives rise to the semantic effects described by Atreya & Sinha (2020); Alok (2022).

I will not attempt to give an account of the relationship here, but definiteness marking and non-honorificity appear to be connected in many other classifier languages as well. Jenks (2015a: 5) reports that in Thai high animate referents are expressed as bare nouns rather than bare classifier constructions, even when anaphoric. Simpson & Biswas (2016: 6) report similar data in Bangla and note that including the classifier indicates disrespect/contempt toward the referent. Finally, regarding names, Saul & Wilson (1980: 26) report that in Nùng classifiers may be used with proper names, but only for children. A possible explanation for these facts might come from the inverse relationship between marked semantic features and honorificity, explored by Wang (2023) for pronouns.

3.1.3 Weak Familiarity

So far, it appears that whenever German uses the strong article, Magahi uses *-waa* and whenever German uses the weak article, Magahi uses a bare noun. But, like the Akan article, *-waa* does not strictly require linguistic antecedence.

(33) **Context:** Ram and Rakesh are standing on the side of the road and see a dog on the other side. Ram says to Rakesh:

(uu) kutt-#(waa) saṛak paar karailai chaho hai
DEM dog-waa road across cross want AUX

'The/that dog wants to cross the road.'

Example (33) shows there is a contrast between Magahi *-waa* and the German strong article. Rather than separating the semantics of *-waa* from familiarity, I propose that the relevant notion of familiarity is slightly broader than linguistic antecedence. In particular, I adopt the definitions of *strong familiarity* and *weak familiarity* from Roberts (2003).

(34) **Strong Familiarity** (Roberts 2003: 304)
The NP has as antecedent a discourse referent introduced via the utterance of a (usually) preceding NP.

(35) **Weak Familiarity** (Roberts 2003: 304)

  i. The entity referred to is perceptually accessible to the interlocutors.

 ii. The entity referred to is globally familiar in the general culture or at least among the participants in the discourse, although not mentioned in the immediate discourse.

iii. Introduction of the NP's discourse referent is licensed solely by contextual existence entailments.

iv. Weak familiarity is guaranteed by giving a functional interpretation to the definite description (which function may have to be accommodated, with the intended argument(s) both familiar and highly salient (Bridging)).

The example in (33) is a case described by (35i.). Except for bridging, which will be mentioned in Section 3.1.4, the other examples of weak familiarity also require *-waa* in Magahi.

(36) **Context:** Ram and John are from the same town which has a single hospital that everyone knows about. Ram is not feeling well, so John tells him:

jaa aspatal-iyaa me dekhwaala
go hospital-WAA in examine

'Go and get check up in the hospital.'   (Alok 2022: 5)

(37) a. *sabhe hotel ke kamar me e-go kitaab rakhal raho hai*
      every hotel GEN room in one-CLF book kept PROG AUX

     'A book is kept in every hotel room.'

  b. *ii kamar-waa me kitab-aa tiivi-yaa ke niiche rakhal hai*
     DEM room-WAA in book-WAA TV-WAA GEN below kept is

     'In this room, the book is kept below the TV.'

Thus, *-waa* is a familiar definite marker that encompasses both strong and weak familiarity. The semantics I propose for *-waa* is largely the same as that of the German strong article.

(38) $[\![\text{-}waa]\!] = \lambda s_r \lambda P_{<e,st>} \lambda y : \exists!x\,[P(x)(s) \wedge x = y] \wedge \underline{\exists s'[s \leq s' \wedge |\{x \mid P(x)(s')\}| > 1]} \wedge \underline{\underline{\text{NHON}(x)}}.\iota x[P(x)(s) \wedge x = y]$

It has the additional presuppositions of non-uniqueness (single-underlined), as argued for in Akan by Owusu (2022), and non-honorificity (double-underlined) to capture the incompatibility of *-waa* with both larger situation uniqueness definites and (quasi-)names of people of higher social status. I assume that any referent that satisfies strong or weak familiarity (Roberts 2003) can introduce an anaphoric index as an argument for *-waa*.

232

3.1.4 Bridging

Magahi, unlike German, does not appear to distinguish between producer-product and part-whole bridging. I have not performed a full exploration of bridging in Magahi but, for the following two examples, speakers report *-waa* to be optional.

(39) *kamar-waa me dekhaliai. chat(-waa) bari ũũcha hai*
    room-WAA in looked   roof-WAA very high is

    'I looked into the room. The roof is very high.'

(40) *kal   e-go   kitaab paṛhaliai. okar   lekhak(-waa) samajhdaar hai*
    yesterday one-CLF book   read   3SG.GEN author-WAA smart   is

    'Yesterday I read a book. The author is very smart.'

The optionality of *-waa* here is a puzzle that I leave open for future research.

3.2 Syntax of *-waa*

This section argues that *-waa* can be a classifier that undergoes CLF to D movement or be base generated in D (Simpson 2005). In order to see how this analysis works, first we need to consider the usual use of classifiers in Magahi.

3.2.1 Magahi Classifiers

The majority of nouns in Magahi require a classifier to combine with numerals, though the classifier system is not very rich. There are two general classifiers *-go* and *tho* in free variation.

(41) {*e-go / ek tho*}   *aam*   (42) {*chaar-go / chaar tho*} *aadamii*
    {one-CLF / one CLF} mango   {four-CLF / four CLF}   man

    'one dog'   'four men'

The difference between them is that *-go* forms a closer prosodic unit with the numeral, as seen in the form of the numeral *ek* 'one' in (41). While *-go* and *tho* occur with the vast majority of nouns, some nouns require more contentful measure phrases (*massifiers* in the terminology of Cheng & Sybesma 1998).

(43) *ek muṭṭhi bhaat*   (44) *du gilas paani*
    one handful rice   two glass water

    'a/one handful of rice'   'two glasses of water'

Finally, there are two other classifiers which are only used with certain quantifiers.

(45) *baṛi menii aadamii*
    lot CLF man

    'lots of men'   (Alok 2012: 47)

(46) *tanii sun aadamii*
    little CLF man

    'a few men'   (Alok 2012: 47)

The few exceptions to the Magahi classifier system are related to time, such as *din* 'day', which do not combine with classifiers/measure phrases at all.

3.2.2 Classifiers in Definites

The question to be answered is how we know that *-waa* is related to the classifier position. For Kumar (2020), the evidence is that Magahi definites with *-waa* do not have the classifiers *-go* or *tho* and are very similar to bare classifier definites in other languages, especially closely related Bangla. Consider the following indefinite/definite pair in Bangla.[9]

(47) *ɛk ṭa boi*
    one CLF book

    'a/one book'   (Dayal 2012: 204)

(48) *boi ṭa*
    book CLF

    'the book'   (Dayal 2012: 204)

The Bangla definite in (48) looks very similar to a Magahi *-waa* definite, with the classifier seemingly occupying the same position as *-waa*. Note also that the classifier occurs prenominally in the indefinite but post-nominally, like *-waa*, in the definite.

    Despite the similarities in a closely related language, so far there has been no evidence internal to Magahi that *-waa* behaves like the definite allomorph of the classifier. I argue that the best evidence comes from adjectives. Magahi adjectives in noun phrases with *-waa* must be suffixed with *-kaa* (masc.) or *-kii* (fem.), both of which I gloss as -KAA. This suffix is generally thought of as an allomorph of *-waa* that displays definiteness agreement on adjectives (Sinha 1966: 114). However, Kumar (2020, 2022) shows that this is actually a case of determiner spreading involving multiple DPs (Alexiadou 2014).[10] Additionally, despite its usual description as being definite, *-kaa* can occur in indefinite noun phrases.

(49) *paapaa, e-go   baṛa-kaa baet lete aiba kaa*
    papa   one-CLF big-KAA bat bring come Q

    'Father, bring me a big bat, won't you?'   (Kumar 2022: 39)

Interestingly, after *-kaa* in indefinites, you can also get the classifier *-go*.

---

[9]The Bangla facts have been discussed much more, and the empirical picture is a lot more complicated than presented here. For accounts of classifiers, definites, and the Bangla noun phrase more generally, please see Bhattacharya (1999a,b); Dayal (2012, 2014); Chacón (2012); Biswas (2012, 2016); Syed (2015, 2016, 2017); Simpson & Biswas (2016); Syed & Simpson (2017).

[10]A determiner spreading analysis for these phrases also gives credence to the analysis of *-waa* as a determiner.

(50) *hamra e-go   baṛa-kaa-{go / #waa}   kutta chahi*
       1SG   one-CLF big-KAA-{CLF / #WAA} dog   want

      'I want a big dog.'

But in definite DPs only -*waa*, not -*go*, can appear after -*kaa*.

(51) a. *baṛa-ka-{waa / #go}   kutt-waa hamra   par bhãũk gelai*
          big-KAA-{WAA / #CLF}   dog-WAA 1SG.GEN on   bark   went

          'The big dog barked at me.'

I take this as evidence that -*waa* is indeed a definite allomorph of the classifier. However, -*waa* can also combine with nouns that are incompatible with -*go* and *tho*.

(52) *baabujii thoṛe sun bhaat banalkai. bhat-waa ṭebal par hai*
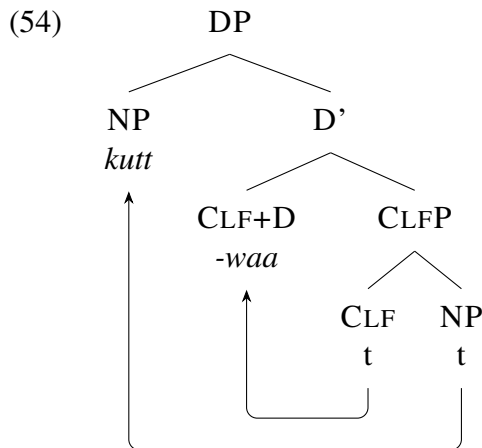   dad   little CLF made rice-WAA table   on is

      'Dad made a little rice. The rice is on the table.'

Din 'day', which does not occur with classifiers, can also occur with -*waa* (Prasad 2008: 189). The compatibility of -*waa* with such nouns means it serves as a general type of definiteness marker, which I take to be associated with the higher functional head D. Earlier, it was shown that -*waa* can also occur in CLF. To account for how it can be associated with both positions, I propose that for most definite nouns, -*waa* begins in CLF and moves to D. This eventually allowed for its reanalysis as a definite marker of category D leading to its compatibility with nouns such as *bhaat* 'rice' and *din* 'day' (cf. the analysis of Vietnamese definites in Simpson 2005). As evidence for CLF to D movement, consider the three forms of Magahi nouns reported by Grierson & Hoernle (1885): bare nouns or "the short form", nouns with -*waa* or "the long form", and nouns where -*waa* occurs twice or "the redundant form". The redundant forms of nouns provide support for CLF to D movement as they arguably involve pronunciation of both copies of -*waa*.[11] The forms for *baat* 'word' are given below.

(53)   **short form   long form   redundant form**
       baat   *bat-iyaa*   *bat-iya-waa*
       word   word-WAA   word-WAA-WAA   (Grierson & Hoernle 1885: 16)

Thus, we have the following forms for -*waa* definites.

---

[11]I only consulted one Magahi speaker regarding redundant forms, but he did not accept them for any of the nouns I checked. Given this, it is possible that CLF to D movement of -*waa* is not synchronically active and -*waa* has been totally reanalyzed as category D. However, then the alternation between -*waa* and -*go* on adjectives would need further explanation. For this reason, I will keep the CLF to D movement analysis and assume that pronunciation of both copies of -*waa* is no longer allowed. However, I believe it is possible that the classifier use of -*waa* is restricted to just the adjectival examples, and an analysis where -*waa* is otherwise base-generated in D is a feasible alternative that would differ minimally from the present account.

(54)
```
            DP
          /    \
        NP      D'
       kutt    /  \
          CLF+D    CLFP
          -waa     /  \
                 CLF   NP
                  t     t
```

(55)
```
            DP
          /    \
        NP      D'
       bhat    /  \
              D    NP
            -waa    t
```
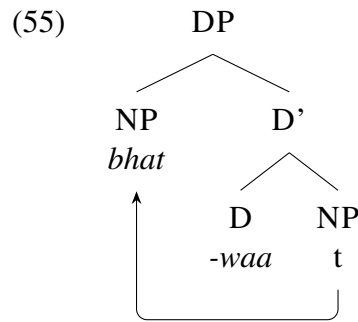
A few features of the analysis in Schwarz (2009) are omitted in these structures for simplicity. For the structures in (54) and (55) there must also be a situation argument introduced with D and an index argument in a higher specifier of DP, above the SpecDP position that the NP occupies. In both DPs, the NP moves to SpecDP to check a [DEF] feature (cf. the NP-raising analysis of Bangla in Bhattacharya 1999a,b and its reformulation in terms of definiteness rather than specificity in Dayal 2012). The difference between the two is that for *kutt-waa* 'the dog', *-waa* is generated in CLF and moves to D, but for *bhat-waa* 'the rice', it is generated in D.

Interestingly, this process of a classifier becoming more determiner-like might also be happening for singular definites in Bangla. Consider the following example, where the indefinite expressions use the human classifier *jon*, but the definite expressions use the general classifier *Ta*.

(56) *laibreri-te Ek jon   notun mEthor ar Ek jon   gard rakha*
      library-LOC one CLF*human* new   janitor and one CLF*human* guard keep

     hoechhe. mEthor *(Ta) porisromi,   kintu gard *(Ta) besh kuMRe.
      was   janitor CLF hard-working but   guard CLF quite lazy

     'The library hired a new janitor and a new guard. The janitor is hard-working, but the guard is quite lazy.'   (Simpson & Biswas 2016: 5)

Furthermore, use of the human classifier *jon* in the definite expressions is ungrammatical (Saurov Syed, p.c.), and *Ta* can also be used as a definite marker for nouns like 'rice', which it cannot combine with in indefinite constructions.

(57) *ma   olpo bhaat ranna koreche. bhaat-Ta Tebil-er opor rakha ache.*
      Mother some rice cook did   Rice-CLF table-LOC on   kept is

     'Mom cooked some rice. The rice is on the table.'   (Indira Das, p.c.)

236

This indicates that *Ta* can be used as a general marker of definiteness, like *-waa* in Magahi, and might not always be (exclusively) associated with the CLF position. However, there are other classifiers besides *jon* which can be used in both indefinites and bare classifier definites,[12] so the situation is not exactly the same as in Magahi and needs further research.

3.2.3 Arguments Against *-waa* as a Determiner

Alok (2012, 2022) presents three main arguments against the analysis of *-waa* as a determiner, based on the following: blocking, linear order, and numerals. Starting with blocking, Alok (2022) states if *-waa* is a determiner, the existence of definite bare nouns in Magahi is a Blocking Principle violation.

(58) **Blocking Principle** (Chierchia 1998: 360)
      For any type shifting operation τ and any X:
      $*\tau(X)$ if there is a determiner D such that for any X in its domain $D(X) = \tau(X)$.

However, even if bare noun definites require type-shifting via *iota*, this does not necessarily indicate that the blocking principle is being violated. As argued in Section 3.1, *-waa* has presuppositions of familiarity, non-uniqueness, and non-honorificity. Thus, the Blocking Principle will not be violated if type-shifting is used for the reference of an individual that fails to meet any of the presuppositional requirements of *-waa*, which is indeed the case for definite bare nouns in Magahi.

Turning to linear order, Alok (2012, 2022) notes that *-waa* occurs after the noun and adjectives are pre-nominal. Alok argues this is an issue if *-waa* heads a DP because because the way to resolve the linear order of *-waa* and the noun would be N to D movement. This would then mean that adjectives in noun phrases with *-waa* would have to occur after the N-*waa* sequence. However, the analysis in this paper derives the position of *-waa* relative to the noun by NP-raising. Thus, adjectives would still be expected to occur to the left of the noun, since they would be inside the NP during NP-raising.[13]

Finally, Alok argues that *-waa* is not a definite determiner because it occurs in indefinite numeral/quantifier expressions.

(59) *chaar go   kitab-waa*
      four   CLF book-WAA

      'four of the books'   (Alok 2022: 16)

(60) *kuch   kitab-waa*
      some book-WAA

      'some of the books' (Alok 2022: 16)

However the translations, taken directly from Alok (2022), show these are partitives and likely involve a more complex structure with two DPs (cf. the indefinites in 17).

To conclude, I will give one more piece of evidence for analyzing *-waa* as a determiner. In vocatives, *-waa* is disallowed.

---

[12]I would like to thank Andrew Simpson for pointing this out.

[13]This is consistent with Alok's (2012) analysis of adjectives as phrasal adjuncts inside NP rather than heads.

(61) **Context:** You see a boy across the street and want to call him.

laṛaka-(#waa)!
boy-#WAA

'Boy!'

This is consistent with observed differences between DPs and NPs (Longobardi 1994).

4 Conclusion

This paper has provided an overview of the Magahi nominal suffix *-waa*, with emphasis on the familiar/unique distinction proposed by Schwarz (2009). I have proposed, contrary to Alok (2012, 2022) and Atreya & Sinha (2020) but in line with Kumar (2020), that the primary contribution of *-waa* is definiteness.

Semantically, *-waa* is similar to the German strong article, but with additional presuppositions of non-uniqueness (Owusu 2022) and non-honorificity and a weaker requirement for familiarity. It is used for both strong and weak familiarity, as defined by Roberts (2003).

Syntactically, I have argued that *-waa* is the definite allomorph of the general classifier in Magahi. The primary evidence for this came from the alternation between *-waa* and *-go* on definite/indefinite adjectives. I have also proposed that besides undergoing CLF to D movement, *-waa* can be base generated in D (Simpson 2005) based on the redundant forms in Grierson & Hoernle (1885) and the compatibility of *-waa* with nouns that do not take the general classifiers. Crucially, I showed that *-waa* occurring in D does not violate the Blocking Principle (Chierchia 1998) or the order of elements in the Magahi noun phrase.

Acknowledgements

References

Alexiadou, Artemis. 2014. *Multiple determiners and the structure of DPs*. John Benjamins Publishing Company.

Alok, Deepak. 2012. *A language without articles: the case of Magahi*: Jawaharlal Nehru University, New Delhi MA thesis.

Alok, Deepak. 2020. *Speaker and addressee in natural language: Honorificity, indexicality and their interaction in Magahi*: Rutgers, The State University of New Jersey dissertation.

Alok, Deepak. 2021. The morphosyntax of Magahi addressee agreement. *Syntax* 24(3). 263–296.

Alok, Deepak. 2022. The discourse particle -waa in Magahi nominal domain. *Discourse* .

Alok, Deepak & Mark Baker. 2022. Person and honorification: Features and interactions in Magahi. *Glossa: a journal of general linguistics* 7(1).

Anderson, John M. 2004. On the grammatical status of names. *Language* 80(3). 435–474.

Arkoh, Ruby & Lisa Matthewson. 2013. A familiar definite article in Akan. *Lingua* 123. 1–30.

Atreya, Lata & Sweta Sinha. 2020. Phonological and functional analysis of diminutive marker -waa in Magahi. *Dialectologia: revista electrònica* 43–57.

Bhattacharya, Tanmoy. 1999a. Specificity in the Bangla DP. *Yearbook of South Asian Languages and Linguistics* 2. 71–99.

Bhattacharya, Tanmoy. 1999b. *The structure of the Bangla DP*: University College London dissertation.

Biswas, Priyanka. 2012. Reanalyzing definiteness in Bangla. In *Annual Meeting of the Berkeley Linguistics Society*, 19–29.

Biswas, Priyanka. 2016. *Number marking and definiteness in Bangla*: University of Southern California dissertation.

Chacón, Dustin. 2012. Head movement in the Bangla DP. *Journal of South Asian Linguistics* 4(1). 3–24.

Cheng, Lisa Lai-Shen & Rint Sybesma. 1998. Yi-wan tang, yi-ge tang: Classifiers and massifiers. *Tsing Hua journal of Chinese studies* 28(3). 385–412.

Cheng, Lisa Lai-Shen & Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30(4). 509–542.

Chierchia, Gennaro. 1998. Reference to kinds across language. *Natural Language Semantics* 6(4). 339–405.

Clark, Herbert H. 1975. Bridging. In R.C. Schank & B.L. Nash-Webber (eds.), *Theoretical issues in natural language processing*, New York: Association for Computing Machinery.

Coppock, Elizabeth & David Beaver. 2015. Definiteness and determinacy. *Linguistics and Philosophy* 38. 377–435.

Dayal, Veneeta. 2012. Bangla classifiers: Mediating between kinds and objects. *Rivista di Linguistica* 24(2). 195–226.

Dayal, Veneeta. 2014. Bangla plural classifiers. *Language and Linguistics* 15(1). 47–87.

Dayal, Veneeta & Li Julie Jiang. 2023. The puzzle of anaphoric bare nouns in Mandarin: A counterpoint to Index! *Linguistic Inquiry* 54(1). 147–167.

Etxeberria, Urtzi & Anastasia Giannakidou. 2010. Contextual domain restriction and the definite determiner. In Isidora Stojanovic Francois Recanati & Neftali Villanueva (eds.), *Context-dependence, perspective and relativity*, 93–126. De Gruyter Mouton Berlin.

Etxeberria, Urtzi & Anastasia Giannakidou. 2019. Definiteness, partitivity, and domain restriction: A fresh look at definite reduplication. *Studies in Diversity Linguistics* (25).

Frege, Gottlob. 1892. On sense and reference. In P. Geach & M. Black (eds.), *Translations from the philosophical writings of Gottlob Frege*, 56–78.

Grierson, George & A.F. Hoernle. 1885. *A comparative dictionary of the Bihari language*. Calcutta, Bengal Secretariat Press.

Hawkins, John. 1978. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Croon Helm.

Heim, Irene Roswitha. 1982. *The semantics of definite and indefinite noun phrases*: University of Massachusetts Amherst dissertation.

Jenks, Peter. 2015a. Determiners, bare nouns, and donkey sentences in numeral classifier languages. Presentation at the 89th Annual Meeting of the LSA.

Jenks, Peter. 2015b. Two kinds of definites in numeral classifier languages. In *Semantics and Linguistic Theory*, 103–124.

Jenks, Peter. 2018. Articulated definiteness without articles. *Linguistic Inquiry* 49(3). 501–536.

Kumar, Chandan. 2020. *Typology of nominal modifiers in major languages of Bihar: Magahi, Bhojpuri, Maithili, Angika & Bajjika*: Jawaharlal Nehru University dissertation.

Kumar, Chandan. 2022. Multiple determiners in Magahi: A case beyond agreement. *Indian Journal of Language and Linguistics* 3(1). 32–45.

Lahiri, Bornini. 2021. *The case system of Eastern Indo-Aryan languages: A typological overview*. Routledge India.

Löbner, Sebastian. 1985. Definites. *Journal of Semantics* 4(4). 279–326.

Longobardi, Giuseppe. 1994. Reference and proper names: A theory of N-movement in syntax and logical form. *Linguistic Inquiry* 609–665.

Owusu, Augustina P. 2022. *Cross-categorial definiteness/familiarity*: Rutgers The State University of New Jersey dissertation.

Potts, Christopher. 2007. The expressive dimension. *Theoretical linguistics* 33(2). 165–198.

Prasad, Saryoo. 2008. *Magahī phonology: a descriptive study*. Concept Publishing Company.

Roberts, Craige. 2003. Uniqueness in definite noun phrases. *Linguistics and Philosophy* 26. 287–350.

Russell, Bertrand. 1905. On denoting. *Mind* 14(56). 479–493.

Saah, Kofi Korankye. 1994. *Studies in Akan syntax, acquisition, and sentence processing*: University of Ottawa dissertation.

Saul, Janice E. & Nancy F. Wilson. 1980. *Nung grammar*. Summer Institute of Linguistics, University of Texas at Austin.

Schwarz, Florian. 2009. *Two types of definites in natural language*: University of Massachusetts Amherst dissertation.

Schwarz, Florian. 2013. Two kinds of definites cross-linguistically. *Language and Linguistics Compass* 7(10). 534–559.

Schwarz, Florian. 2019. Weak vs. strong definite articles: Meaning and form across languages. *Studies in Diversity Linguistics* (25).

Simpson, Andrew. 2005. Classifiers and DP structure in Southeast Asia. In Guglielmo Cinque & Richard S. Kayne (eds.), *The Oxford Handbook of Comparative Syntax*, 806–838. New York: Oxford University Press.

Simpson, Andrew & Priyanka Biswas. 2016. Bare nominals, classifiers, and the representation of definiteness in Bangla. *Linguistic Analysis* 40(3-4). 167–198.

Sinha, A.C. 1966. *Phonology and morphology of a Magahi dialect: A thesis submitted to the University of Poona*. Poona (Deccan College Postgradute and Research Inst.).

Syed, Saurov. 2015. Focus-movement within the Bangla DP. In *Proceedings of WCCFL 32*, 332–341.

Syed, Saurov. 2016. Decomposing definiteness: Arguments for a split D-domain in Bangla. In *Proceedings of WCCFL 33*, 390–397.

Syed, Saurov. 2017. *The structure of noun phrases in Bengali: What it tells us about phases and the universality of DPs*: University of Southern California dissertation.

Syed, Saurov & Andrew Simpson. 2017. On the DP/NP status of nominal projections in Bangla: Consequences for the theory of phases. *Glossa: a journal of general linguistics* 2(1).

Verma, Sheela. 2003. Magahi. In Danesh Jain & George Cardona (eds.), *The Indo-Aryan Languages*, chap. 13. Routledge.

Wang, Ruoan. 2023. Honorifics without [HON]. *Natural Language & Linguistic Theory* 41(3). 1287–1347.

Yip, Ka-Fai, Ushasi Banerjee & Margaret Chui Yi Lee. 2023. Are there "weak" definites in bare classifier languages? In *Semantics and Linguistic Theory*, 253–275.