

The Value of Studying Grammaticalization in Semitic

Aaron D. Rubin - Pennsylvania State University
Dept. of Classics and Ancient Mediterranean Studies
108 Weaver Building, University Park, PA 16802 (U.S.A.)

[This paper will show the value of using cross-linguistic examples of grammaticalization in providing direction for an investigation, as well as in helping to make a seemingly unusual change perhaps seem more comprehensible. The major focus of this paper will be on the definite articles of Amharic and other South Ethiopian languages, though data from Mehri and numerous other Semitic languages will also be presented.]

Keywords: grammaticalization, Mehri, Amharic, reflexives, definite article.

One of the most enjoyable parts about studying grammaticalization is the opportunity for cross-linguistic comparison.¹ What I want to focus on in this paper is the value of using cross-linguistic evidence, not for proving a development in any one language, but for determining universal linguistic tendencies that can give direction to the investigation of a particular change. I want to emphasize that by using cross-linguistic comparison, a change that may seem strange or unfamiliar can suddenly become familiar.

As a simple and short example of this, we can think about the Hebrew reflexive pronouns *'ašmī* 'myself', *'ašmākā* 'yourself', and so on, which are all based on the word *'ešem* meaning 'bone'. In isolation, it might seem strange that the noun 'bone' should come to indicate reflexivity, but comparison with both Semitic and non-Semitic languages shows that words for parts of the body are a very common source for reflexives across the world's languages (Schladt 2000; Rubin 2005). Some examples from within Semitic are Ge'ez *rə's-* (< 'head'), Chaha *gäg-* (< 'body'), Old Assyrian *paḡr-* (< 'body'), and Classical Arabic *nafs-* (< 'soul') and *ayn-* (< 'eye'). Examples from non-Semitic languages are Yoruba *ara* (< 'body'), Hausa *kāi* (< 'head'), Tamazight *iḡef* (< 'head'), and Somali *laft-* (< 'bone'). This tendency

1. This article is a condensed version of the paper that was delivered in Turin on October 3, 2008. This is because some of the content appears in print elsewhere. It was a great pleasure and honor to be invited as one of the keynote speakers of this conference, and I would like to sincerely thank the organizers, Professor Fabrizio Pennacchietti and Professor Alessandro Mengozzi.

for the source of reflexives is, of course, well known, but it serves to show the usefulness of cross-linguistic comparison.

As another example, which may be less familiar, we can turn to Mehri. In Mehri the verb *hōm* has the basic meaning ‘to want’, as in the following sentences:

hōm moh ‘I want water’ (99:28)²
hōm əl-herəs ‘I want to get married’ (8:8)

This verb has become grammaticalized in a couple of different ways in Mehri.³ It can be used to indicate a proximative ‘to be about to do something’, as in:

bərhəm yəhəym yəšakfəm ‘they were about to go to sleep’ (75:17)
kəsk aḡayg də-bərəh yəhōm yəhwe ‘I found the man about to fall down’ (77:6)

It can also be used in the first persons (singular, dual, and plural) to indicate a mild form of obligation, corresponding to English ‘should’, as in:

nəhōm nəršān tāṭidayən ‘let’s tie each other up’ (24:25)
ḏōməh aḡəggēn wəkōna axayr mənay mət ’ākawr, wə-hōm l-əwtəḡəh ‘this boy will be better than me when he grows up, so I should kill him’ (76:12)

The first example of grammaticalization here, by which the verb ‘want’ comes to indicate a proximative ‘to be about to’ may strike one as unusual at first, but is actually very common cross-linguistically. A similar development can be found in Persian, Hungarian, Swahili, German, and Bislama (a Pacific creole). Following are some examples from German, where we see the verb *wollen* (+ *gerade*) indicating a proximative:

Wir wollten gerade gehen ‘we were just about to go’
Ich wollte gerade einkaufen gehen ‘I was about to go shopping’

The second type of grammaticalization, by which ‘want’ comes to indicate obligation may seem equally or more unusual, but this also has parallels around the world. Again, we find parallels in other languages, including Mandarin Chinese and, once again, German:

Wir wollen gehen! ‘we should go!’

So to reiterate, by studying these cross-linguistic tendencies, it allows us to recognize or understand a particular change that might in isolation seem very unexpected. Of course, simply recognizing a parallel development in another language does nothing to *explain* this development. But such a recognition can provide the starting point for an investigation.

2. All Mehri text numbers refer to Stroomer (1999). My own English translations often differ from those found in Stroomer’s edition. On Mehri in general, see Rubin (2010b).

3. For a more comprehensive study of the Mehri verb *hōm* and its grammaticalization, see Rubin (2009).

There are certain types of grammaticalization that are extremely common around the world, such as the development of future tenses based on verbs meaning ‘to want’ or ‘to go’; the development of indefinite articles from the numeral ‘one’; the development of prepositions from words for body parts like ‘head’, ‘foot’, and ‘back’; the development of quotative particles from the verb ‘say’; and the development of definite articles from demonstratives. To cite just a few examples for this last category, the development of definite articles from demonstratives, we can mention Italian *il* from Latin *ille*, Hungarian *az* (which is identical with the demonstrative), Coptic *p-*, and Vai *me*. We also find this development among several of the modern Semitic languages. For example, the Turoyo definite article *u-* (citing only the masculine singular form for simplicity) comes from a demonstrative, either from a form *haw* or *hu*. In modern Western Aramaic of Ma‘lula, the demonstrative *hanna* is often used as a simple definite article, and there is good evidence that this is a grammaticalization in progress. Likewise in Tigrinya, the demonstrative *'ətu* is becoming grammaticalized as a definite article (Rubin 2005).

Looking at the older Semitic languages, we find this development of the definite article in Central Semitic, where Hebrew *ha-*, Aramaic *-ā*, Arabic *al-*, and Sabaic *-(h)n* all almost certainly come from demonstratives. I discussed this issue at length elsewhere, where I began my arguments with the idea that while the strong universal tendency for articles to come from demonstratives does not prove anything with regard to any one individual language, it does provide a good starting point (Rubin 2005). And I concluded, as have others before me, that the articles of the Central Semitic languages also must derive from demonstratives. And I think most scholars today would agree that this is correct, though we may not agree precisely on how this happened, or precisely which demonstratives provided the source for the articles.

But what do we do when we find an example of a grammaticalization that seems to go against universal tendencies? As a test, we can look at cases in which we find definite articles that do not develop from a demonstrative. In some cases, definite articles are clearly borrowed. This is the case in several modern Aramaic dialects, which have borrowed a definite article from Kurdish, and it is likely the case in Mehri, which, as suggested by Sima (2002), has probably borrowed its article from neighboring Arabic dialects. But when a language has developed a definite article on its own, not through borrowing, and not from a demonstrative, I think that the practice of cross-linguistic comparison still proves very instructive, not as proof of anything, but as a source of ideas and as possible support for one’s argument. To illustrate this point, we can consider the case of the definite article in Amharic and several other South Ethiopian languages.⁴

The definite article in Amharic takes the form of a suffixed *-u* for masculine singular and the plural, while the feminine singular has the variant forms *-wa*, *-itu*, and *-it^wa*, which are used more or less interchangeably (Leslau 1995). The masculine singular and common plural article *-u* is identical in form to the third masculine singular possessive suffix, so, for example, Amharic *bet-u* can mean either ‘the house’ or ‘his house’, depending on the context, and a plural noun like *nəgusočč-u* can mean either ‘the kings’ or ‘his kings’. The feminine singular article *-wa* is identical to the third feminine singular possessive suffix, so a word like *lam-wa* can mean either ‘the cow’ or ‘her cow’. The variant form *-itu* is composed of a feminine suffix *-it* plus the third masculine singular possessive *-u*, and the variant *-it^wa* is a combination the feminine marker *-it* and the third feminine singular possessive suffix *-wa*.

There are parallels to the Amharic articles in several other South Ethiopian languages. So for example in Argobba, which is a close relative of Amharic, we find the definite articles *-u* and *-wa*, which are also

4. For a more comprehensive study of the Amharic and other South Ethiopian definite articles, see Rubin (2010a).

identical to the third person singular possessive suffixes. In Gafat, which is now extinct, the third masculine singular possessive suffix *-š* was used as a common definite article. In Chaha, one of the so-called Gurage languages, definiteness is not usually expressed, but the third person possessive suffixes *-(ä)ta*, *-(ä)çta*, *-(ä)^wxna*, and *-(ä)xnäma* can all function as definite articles when needed. In Harari, definiteness is also often left unexpressed, but can be made explicit with the addition of the third masculine singular possessive suffix *-zo*, as in *gār-zo* ‘the house’ or ‘his house’. Similar use of the third person possessive suffixes as articles are occasionally found in other South Ethiopian languages, e.g., Mäsqän, and also already in Ge‘ez.

The development of a third person possessive into a definite article is highly unusual, and there is no evidence in Amharic that allows us to see exactly how this development might have taken place. So in searching to explain this problem, we can turn to parallel developments in other languages for help. As I have discussed at length elsewhere (Rubin 2010a), Indonesian and several other Malayo-Polynesian languages also use definite articles that derive from possessives. In Standard Indonesian, however, the use of the possessive as a marker of definiteness is restricted syntactically to constructions in which the noun has not been previously mentioned. So, for example, we can contrast the following two Indonesian sentences:

Ibu sudah memasak nasi. Nasi itu di lemari.
 mother already cook rice rice that in pantry
 ‘Mother has cooked rice. The rice is in the pantry.’

Kalau mau makan, nasi-nya di lemari.
 if want eat rice-3POSS in pantry
 ‘If you want to eat, the rice is in the pantry.’

As the first of these two sentences illustrates, when the noun has been explicitly mentioned before, it is the demonstrative *itu* that can be used as the definite article. In this sentence, *nasi* meaning ‘rice’ is mentioned once, and upon its repetition the anaphoric demonstrative *itu* refers back to this initial mention: ‘Mother has cooked rice’ and ‘that rice which I just mentioned is in the pantry’. However, when the reference is not explicit, but is only understood, as in the second sentence, then the possessive suffix *-nya* is used as a marker of definiteness. In the first part of this sentence, rice is not explicitly mentioned, but there is an implication of food in the offer made to the addressee. Therefore, the suffix *-nya* is used in this context as the definite article.

Upon further investigation, I found that this same set of syntactic restrictions can also be found in an Ethiopian language, in Chaha. Compare the following two sentences:

muk^yər bā-x^wet enät yəšäği. at-äta yä-däng^ya muk^yər, at-äta yä-gämya muk^yər
 bonfire in-two kinds one.divides one-its of-children bonfire one-its of-adults bonfire
 ‘The bonfire is divided into two kinds. The one is the children’s bonfire; the other is the adults’ bonfire.’ (Leslau 1950, 63)

yančän bāxärä yä-šäg^wäre yarlä. šäg^wärä xuta yəbər...
 it does not leave CONDIT. to-sorcerer one.goes.for.him sorcerer he/that says

‘if it (the sickness) does not leave him, they go to the sorcerer for him. The sorcerer says...’
(Leslau, 1950, 58)

In first of these sentences, it is the word ‘one’ that is being made definite with the possessive suffixes. The noun ‘one’ has not been explicitly mentioned before, and so the reference is implicit. The possessive clearly refers back to the word ‘bonfire’. I translated ‘the one is the children’s’ and ‘the one is the adults’, but if I had translated ‘its one is the children’s’, the meaning would still be clear. In the second sentence, we find *šäg^wärä xuta* ‘the sorcerer’ where *xuta* is the third person pronoun being used as a demonstrative (like Hebrew or Aramaic *hū*), because the word for ‘sorcerer’ has already been mentioned.

So, the expression of definiteness in Chaha, whether with the possessive suffixes or with the independent pronoun/demonstrative, shows similar syntactic restrictions to those of Indonesian. And, in fact, there are other languages in which we find this same distribution of definite marking, including Yucatec Mayan and some Turkic languages.⁵ This type of definite marking must underlie the definite articles of Amharic and other South Ethiopian languages, in which the use of the possessives as a definite article has become much more widespread and unrestricted in usage. We find a similar trend in Colloquial Jakartan Indonesian (Sneddon 2006).

So to summarize, by looking at evidence from Indonesian, we learn how a possessive can originate as a marker of definiteness in certain contexts, and we find an exact parallel in Chaha. By looking at examples of grammaticalization from outside the Ethiopian Semitic sub-family, we found a starting point from which we could uncover the history of the Amharic definite article. The situation that may have seemed strange, namely, the grammaticalization of a third person possessive suffix into a definite article, certainly seems less so, in light of comparative evidence from Indonesian (and elsewhere). As all of the examples of grammaticalization in this paper have shown, cross-linguistic evidence such as this is not proof of anything, but it can provide an excellent starting point for an investigation, and can make a change that might seem strange seem a little bit more comprehensible.

5. I am grateful to Prof. Christian Lehmann for providing me with the information on Yucatec Mayan, and to Prof. Andrzej Zaborski for alerting me to the Turkic data.

REFERENCES

- Leslau 1950 W. Leslau, *Ethiopic Documents: Gurage*, New York 1950.
- Leslau 1995 W. Leslau, *Reference Grammar of Amharic*, Wiesbaden 1995.
- Rubin 2005 A. Rubin, *Studies in Semitic Grammaticalization*, Winona Lake, IN 2005.
- Rubin 2009 A. Rubin, “*Hōm* Sweet *Hōm*: The Unusual Mehri Verb ‘Want’”, *Zeitschrift für Arabische Linguistik* 51 (2009), 76-84.
- Rubin 2010a A. Rubin, “The Development of the Amharic Definite Article and an Indonesian Parallel”, *Journal of Semitic Studies* 55/1 (2010), 103-114.
- Rubin 2010b A. Rubin, *The Mehri Language of Oman*, Leiden 2010.
- Schladt 2000 M. Schladt, “The Typology and Grammaticalization of Reflexives”, in Z. Frajzyngier and T.S. Curl (Eds.), *Reflexives: Forms and Functions*, Amsterdam-Philadelphia 2000, 103-124.
- Sima 2002 A. Sima, “Der bestimmte Artikel im Mehri”, in W. Arnold and H. Bobzin (Eds.), “*Sprich doch mit deinen Knechten aramäisch, wir verstehen es!*”. 60 Beiträge zur Semitistik: Festschrift für Otto Jastrow zum 60. Geburtstag, Wiesbaden 2002, 647-668.
- Sneddon 1996 J.N. Sneddon, *Indonesian: A Comprehensive Grammar*, London-New York
- Sneddon 2006 J.N. Sneddon, *Colloquial Jakartan Indonesian*, Canberra 2006. of T.M. Johnstone, Wiesbaden 1999.
- Stroomer 1999 H. Stroomer, *Mehri Texts from Oman. Based on the Field Materials* 1996.