

THE COMPLEXITY OF NOUN PHRASES IN WRITTEN VS. SPOKEN OSHIWAMBO

BACKGROUND:

Oshiwambo is a Bantu language of northern Namibia. It does not have a long history as a written language. The first written texts were developed by missionaries in the late 19th century and the early 20th century, and they were to a large extent translations of biblical texts. The fact that they were translations means that the languages they were translated from may have influenced the developing written Oshiwambo substantially, just like Latin had an influence on the written varieties of European languages that developed during the Middle Ages. Most of the written texts that were used in this study were newspaper articles which were not translated from another language, but were originally produced in Oshiwambo. Nevertheless, the journalists who wrote them are literate in at least one European language (English, and probably also Afrikaans), and their written Oshiwambo may be affected by that. Even if they were monolingual Oshiwambo speakers the fact that the written variety developed the way it did could lead to influences from written European languages in the texts. In other words, if we find that written Oshiwambo shows the same tendency to long and complex NP:s as European languages we cannot establish to what extent this is because of foreign influences, and - perhaps more importantly - we will not know whether this tendency would have developed if written Oshiwambo had been developed locally by indigenous people, without being based on translations from foreign languages.¹ Nevertheless, it might be of interest to look at the situation in a language which is structurally dissimilar to the Indo-European languages studied by Miller & Weinert.

THE DATA:

One half of the written material consists of twelve Oshindonga newspaper articles from the daily newspaper *The Namibian* (footnote: one section of one of the articles was ignored, since it was only a long list of names of all the people who had died in a minibus accident).

It has been noted that written English, Russian etc. are not homogeneous genres, but they consist of different genres such as academic monographs, legal documents, encyclopaedia articles, informal letters etc. The first three of the examples given in the previous sentence normally have more complex structures. In the case of Oshiwambo there are few academic monographs and legal documents, and probably no encyclopaedia articles, but within the present study I have also looked at something that can be described as an academic monograph, *Ongeleki methimbo lyuulunde wa putuka*, a book by Sebulon Ekanjo which discusses the impact of the Church on the society in Owamboland.

The spoken material consists of four narratives. They were chosen randomly from a corpus of spoken Ochikwambi (Ochikwambi is an Oshiwambo dialect) that I developed together with an assistant from 2001 to 2003 as part of a study where I compared Ochikwambi and Oshindonga. We cannot be sure about the extent to which these persons had talked about the same events on other occasions, but the narratives are not typical stories of the fairy-tale type that have been told in more or less the same form many times, but they must be considered to be examples of relatively spontaneous speech.

The total number of noun phrases in the written material is 1739, and in the spoken material there are 2176, which is comparable to the number of noun phrases in the corpora discussed by Miller & Weinert.

The written texts are in Oshindonga, whereas the spoken material is in Ochikwambi, but there is no reason to believe that differences found would be because of this dialect difference. Just as the difference between Miller & Weinert's Scottish spoken texts and written Standard English should not be thought of as indicating regional dialectal differences, the differences that we find in the Oshiwambo material must surely be because of the fact that one part of the material consists of written language and the other one of spoken

¹ There are reasons to believe that this tendency of having heavy NP:s in written language comes naturally because of the nature of written language (see for instance Miller & Weinert (1998) and Teleman (1983)).

language. The two dialects are closely related (more closely than for example Oshindonga and the other standardized Oshiwambo dialect, Oshikwanyama).

THE CODING:

In some cases we have two juxtaposed Oshiwambo clauses where the second clause corresponds to some smaller constituent in English, for example an adverb. For example, *oto popi to enderere* means 'you speak fast'. Such an Oshiwambo sentence would in the present study be considered to include two cases of 0-NP subject. On the other hand, in the case of habitual past sentences of the type *oya li haa lith'eengombe* and *okwa li ha ningi owaandaha* I have counted only one 0-NP subject (or no 0-NP subject if the construction is preceded by an explicit NP). In any case, these sentence types are not very common, which means that even if we were to analyze them in a different way the findings of the study would basically be the same. The reason for counting only one 0-NP in these constructions is that the the clause *oya li* (or *okwa li* etc.) is semantically rather empty in these cases, and it has a grammatical function rather than a semantic meaning. Some other similar structures have been treated the same way for the same reason, for example *mu kale mu n'ombiri* (=for you to have peace).

Nouns belonging to the locational Bantu noun classes 16, 17 and 18 have been considered to be single nouns even though the English translations have prepositional phrases. In order to be consistent one then has to analyze the demonstratives of these noun classes, like for instance *mpa(no)*, *mu(no)* etc., as one-word NP:s, even though they correspond in meaning to English adverbs like 'here' and 'there'. Sometimes one runs into the problem of how to analyze structures like *mUukwambi mu* (= 'here in Uukwambi' or 'in Uukwambi, here'). Should the demonstrative *mu* be considered to modify the class 18 noun *mUukwambi*, or is it an independent adverbial that is quickly added to the utterance so that we have two independent deictic markers in this case? In those cases one has to make an impressionistic judgement, based partly on intonation, partly on the context. Since words like *mUukwambi* have been considered to be nouns, the way one analyzes these structures makes a difference in the number of single noun NP:s vs. NP:s consisting of noun+demonstrative.

In cases where a description or title is added to a personal name a distinction has been made between those cases where there is a comma between the name and the description/title and those where there is no such comma, compare the difference between English 'Tony Blair, Prime Minister of the UK' and 'Prime Minister Tony Blair'. The first case would here be considered as two separate NP:s. In the second case we have a noun modified by another noun by simple juxtaposition, but it is not immediately obvious which noun is modifying which. Is 'Tony Blair' a modifier of 'Prime Minister', telling us which Prime Minister we are talking about, or is 'Prime Minister' a modifier of 'Tony Blair', attributing a quality to him? It makes sense to treat the name in a case like for instance *omukwaniilwa Iipumbu* (=King Iipumbu) as a postmodifier of *omukwaniilwa* (if nothing else at least because of the fact that almost all noun modifiers in Oshiwambo are postmodifiers). The title *tate* (corresponding to English 'Mr') was not considered to be a head noun though, but was just considered to be part of the name, so that for example "*tate Ndjendjela*" (=Mr Ndjendjela) was considered to be a single noun.

No difference has been made between cases similar to the structural ellipsis in English sentences like 'John bought a gun and killed his neighbour' and cases where an explicit subject is missing in non-coordinate structures (similar to for example Spanish 'Ya te he dicho que sabes poco de achaque de aventuras' = (I) have told you already that on the subject of adventures (you) know little).

Miller & Weinert's category 8 has been left empty in the chart. Even though there is a kind of compound nouns in Oshiwambo, like for instance *ondjalulamasiku* (=calendar, from *ondjalula* (=list, record) and *omasiku* (=days)), there seems to be no strong reason to treat these differently from other single nouns.

Even though verbal nouns of the type *okusa* (=to die) formally belong to noun class 15 they have not been treated as nouns in the present study, unless they occur with nominal modifiers like for instance *hu*, *kwawo*, etc. In most cases they rather occur with verbal complements such as objects and adverbs. Those few nouns from noun class 15 that denote physical objects, like for example *okuguru* (=wheel), have been treated as nouns at all times.

Miller & Weinert's category 5 is for those NP:s where a noun is modified by a prepositional phrase. In the case of Russian it also includes NP:s with nouns in the genitive case. In Oshiwambo there are no prepositional phrases of the Indo-European type, which means that most of the cases we find in category 5 in this study are NP:s with one or more genitive-marked nouns modifying a head noun. There are nouns

from the locational classes 16-18 that are used instead of English prepositions, like *pokati*, *konima*, *komesho* etc. These have been treated as nouns in spite of their functional similarity to the European prepositions. This means that a phrase such as *konima yegumbo* (=behind the house, but literally "at the backside of the house") is a NP belonging to type 5a.

A spoken sentence like *Kandi chi ng'oodh'iire po nгаа eenkutuwa...nomapushu* (= I don't know whether the 'eenkutuwa' and the 'omapushu' came to the missionaries) has been interpreted as having two NP subjects. As far as I know, postverbal subject NP:s would not occur in this way in written Oshiwambo, which would make it tempting to treat the sentence as having a 0 NP with the subjects quickly added as an afterthought, but in doing that we might be forcing the rules of the written language on the spoken language. There is no intonation break between the verb and the first following NP. In any case, this sentence type is not frequent even in spoken language, so the question of what is the correct way of analyzing it does not affect the final result of the study substantially.

Additional modifiers that seem to be added parenthetically have been ignored (footnote: for example 'gwomimvo 38' in *Cornelia Uxas, gwomimvo 38* (=Cornelia Uxas, 38 years old), or, from the spoken corpus: *mombara hashekuru haNegumbo lyaKandenge* (=in the palace of his uncle, of Negumbo lyaKandenge), even though there is no clear intonation break in the spoken example given here.

For the same reason that Miller & Weinert chose not to consider the structure 'which of the books' as having two parts which themselves are counted as smaller NP:s I prefer not to count any smaller NP:s which form part of a larger one in the chart. In a case such as *omulongi gweehapu dhaKarunga* (=a teacher of the words of God) I have not considered *eehapu dhaKarunga* as a separate NP or *Karunga* as a single noun NP, because the speaker is not uttering a two-word or a one-word NP but a three-word sequence. However, a second set of figures is given in brackets to show what the results would be if we use Miller & Weinert's way of counting. When using their method NP:s inside relative clauses are counted as well as NP:s modifying nouns in NP:s belonging to type 5, but NP:s inside of complex clauses are still ignored.

Miller & Weinert use the term 'determiner'. This is not a term that I have seen being used in the description of Bantu languages, but in an attempt to make the findings of this study somewhat comparable to those of Miller & Weinert I decided to use this term to refer to a group of noun modifiers with some similarities to those labelled 'determiner' by Miller & Weinert. In my study the term then refers to numerals, demonstratives, possessive pronouns and a few less frequently used modifiers. However, genitive-marked nouns are not included, and neither are adjectives.

This leaves us with the following nine categories (with further subcategories):

- 1) NP:s lacking a head noun and having a pronoun or a noun modifier as the sole constituent
- 2) NP:s consisting of only a noun
- 3) NP:s consisting of noun+determiner (or, in a few cases, determiner+determiner (*kehe gumwe* = each one))
- 4) NP:s with an adjective
- 5) NP:s where a noun is modified by another noun or nouns. The modifier is then typically a genitive-marked noun, but there are also cases with simple juxtaposition. In the written material the simple juxtaposition of a locational noun (from classes 16-18) is not uncommon, but in the spoken material it is very rare. One cannot help but to suspect that this phenomenon in written Oshiwambo is because of influence from European languages, where prepositional phrases are common as noun modifiers ('the man on the roof', 'a village in Norway', etc.). In spoken Oshiwambo the locational noun would normally be genitive-marked, like other nouns functioning as noun modifiers (e.g. *aahona yomoVenduka* = the masters in Windhoek, lit. "The masters of in Windhoek").
- 6-7) NP:s with a relative clause. At an early stage in the study clauses starting with a demonstrative (e.g. *iipango mbyoka nde yi mu pe nena ndjika*) were treated as relative clauses and those lacking one (e.g. *epulo tali tsakanekwa ngashingeyi komilungu dhoyendji*) were considered as corresponding to the term "participial clause" of Miller & Weinert's study, because there was a tendency (although weak) to translate the latter type with English participial clauses in the Ochikwambi research project mentioned in the section about the data. However, the similarities between the latter type and English participial clauses were then not regarded as strong enough, and both types of noun-modifying clauses have now been grouped together, but separated as two subcategories a) (having an initial demonstrative) and b) (lacking one). (footnote: In the end it also turned out that the NP:s with "participial clauses", i.e. 6-7 b), are more common than NP:s belonging to type 6-7 a) in the spoken material, whereas the opposite is true in the written material. This of

course means that it would not have made sense to have the NP:s with "participial clauses" in a higher category like in Miller & Weinert's study, since the idea is to have the structures more typical of written language in higher categories)

9) Complex NP:s. The definition given in Miller & Weinert is rather impressionistic, and I have no clear definition in my study either, but a NP with for example a determiner or an adjective in addition to three relative clauses would count as a complex NP, as would one with three relative clauses including relatively complex NP:s within them, or NP:s with few or no relative clauses but with very complicated genitive modifiers or many relatively complicated ones (like for instance *etidho lyaaniilonga ye li 79 mokampani yokulanditha iiparte yoohauto yaCymot*).

10) 0 NP:s, i.e. cases where there is no explicit subject. As in Spanish and many other familiar languages, subject NP:s are often dropped in Oshiwambo, since the identity of the implicit subject can easily be inferred from the subject concord. These cases have been classified as 0 NP:s.

11) Objective concords. Object concords, like in for example *aantu taye yi pumbwa* (=People who need them), are written disjunctively in the standard orthography, and have been treated as similar to English object pronouns. The decision to treat these as noun phrases may be controversial, but they do resemble unstressed English object pronouns. If one would rather prefer to treat these as verb affixes we are instead dealing with another type of 0 NP:s here, 0 NP objects.

The subclassifications are based on which other modifiers are in the NP:s. 3a) are simple cases where one determiner modifies a noun (or, in a few cases, another determiner), 3b) are those where the head is modified by more than one determiner. 5a) are simple cases where one noun modifies another noun, 5b) are slightly more complex in that there is more than one modifier of the head, for instance a determiner as well as a noun modifying a noun, and 5c) are complex in that a noun modifying the head itself has a noun modifier, for example *ondjokonona yonkalamwenyo yaantu* (=the history of the existence of people). 5d) are a combination of 5b) and 5c), i.e. there is more than one modifier of the head and at least one of them is a noun which itself has a noun modifier. One example of 5d) would be *ekende enene lyonamunate yOsprite* (=a big bottle of the soft drink Sprite). The difference between 6-7a) and 6-7b) has already been explained above.

In cases where more than one type of modifier occurs within a NP the NP is regarded as belonging to the highest category, i.e. if there is both an adjective and another noun it belongs to 5), and if there is both another noun and a relative clause it belongs to 6-7). This is in line with the method used by Miller & Weinert. The categories introduced by me in the Oshiwambo study fall outside of this ranking, since they have just been added as categories 10) and 11).

Some example NP:s are given below:

1) Spoken: ngaye, shiya, oko, se, ngwiya, koole, nayo, oyo, yo - Written: opo, mpoka, shaShaanda, shimwe, yo

2) Spoken: tate, Iipumbu, aantu, naChongora, mUukwambi - Written: koRundu, aanona, omidhimba, konima, mokambesa

3) Spoken: omakopi gangapi, nomeengereka moka, ethimbo alishe, kombara ngaa ndyo, ehauto ndi lyetu - Written: kaakwanezimo yawo, aawiliki yalwe, mombila yimwe, aaniilonga mbaka, kiituthi hoka

4) Spoken: omusaane omukwawo, okakadhona lela okanene, okaskora okachuuchuuchuu - Written: oshiwike oshiwana, aaposi yakwawo, oshinima oshaanawa, lwopotundi onti13h00

5) Spoken: komband'eevi, mombara hashekuru, iinima yaatiligane, iipiriri yetu yomashini - Written: oshituthi shomagongo, omukalimo gwomOluno mOndangwa Silas Nangombe, ekende enene lyonamunate yOsprite, koshipangelo shaShakati, etando lyawo lyomeho, omundohotola Ndume, konima yomimvo odhindji

6-7) Spoken: ehokololo ndi lya za ko, iikaro mbi taya ara yi mu tule koshi, oskora ndjiya hi li hwiya, uumati ya hongwa uupokoro - Written: shoka ta lombwelwa kaakomeho ye, aaniilonga mboka ya tidhwa, mehuliloshiwike lya za ko, oondjambi dha nanga noonkondo, pethimbo oshinyolwa shika tashi yi komanyanyangidho

9) Spoken: nomukulukadhi gumwe a za mombala mono i iyaka e ya okoolama mpano, nochirumbu hachi zi kOutjo haku ti Omboroma shi na ofurasha heeshako tahe eta pwEelim - Written: oongala ndhoka dha kala komoni dhawo dha kuthwa ko koondohotola yomeho mewiliko lyomundohotola a tseyika nawa Helen Ndume, omutaambi gwayehe mboka taa galukile moshilongo melongelokumwe nOmahangano ge lile po uthemba nonkalonawa yaantu muuyuni,

11) Spoken: otandi ku tumu, a li ku tega, otagu ya li, otaa ya nyenge - Written: ngele otamu shi ningi, okwe shi longitha, oye ya kanithitha, otatu gu nyanyangidha ngeyi

FINDINGS:

The Oshiwambo data are presented below, followed by the data for the Indo-European languages studied by Miller & Weinert. For German only the figures for the face-to-face conversation has been included in the comparison below, since the other types of spoken material for German in Miller & Weinert's study consisted of telephone conversation and what Miller & Weinert call "the Map Task", and these two types of spoken language must be considered to be quite different from the spoken language used for Oshiwambo.

Oshiwambo

	Spoken (Narratives)			Written (newspaper articles)			Written (ac. monograph)		
	%			%			%		
1)	266	12.2	(291/11.4%)	49	5.2	(71/5.2%)	79	9.8	(88/8.2%)
2)	613	28.2	(905/35.3%)	240	25.7	(494/36.1%)	215	26.7	(365/34.1%)
3)	185	8.5	(197/7.7%)	120	12.8	(164/8.8%)	94	11.7	(122/11.4%)
	a)	178	8.2	a)	118	12.6	a)	88	10.9
	b)	7	0.3	b)	2	0.2	b)	6	0.7
4)	19	0.9	(20/0.8%)	8	0.9	(10/0.7%)	11	1.4	(14/1.3%)
5)	126	5.8	(142/5.5%)	178	19.0	(264/19.3%)	114	14.2	(145/13.6%)
	a)	85	3.9	a)	83	8.9	63	7.8	
	b)	28	1.3	b)	30	3.2	17	2.1	
	c)	10	0.5	c)	38	4.1	33	4.1	
	d)	3	0.1	d)	27	2.9	1	0.1	
6-7)	110	5.1	(112/4.4%)	107	11.4	(113/8.2%)	94	11.7	(101/9.4%)
	a)	42	1.9	54	5.8		59	7.3	
	b)	68	3.1	53	5.7		35	4.4	
9)	3	0.1	(3/0.1%)	11	1.2	(11/0.8%)	6	0.7	(6/0.6%)
10)	765	35.2	(793/31.0%)	187	20.0	(207/15.1%)	166	20.6	(186/17.4%)
11)	89	4.1	(99/3.9%)	35	3.7	(36/2.6%)	25	3.1	(43/4.0%)
Total)	2176			935			804		

Miller & Weinert English

	Spoken (Narrative)	Spoken (Conversation)	Readers' letters to newspaper
1)	311 (44.9%)	1308 (48.9%)	63 (14.1%)
2)	49 (7%)	127 (4.8%)	68 (15.2%)
3)	51 (7%)	449 (16.9%)	40 (8.9%)
4)	41 (5.6%)	159 (6.0%)	88 (19.7%)
5)	44 (6%)	135 (5.6%)	84 (18.8%)
6)	23 (3.2%)	100 (3.8%)	14 (3.0%)
7)	4 (0.5%)	2 (0.07%)	5 (1.1%)
9)	0	0	14 (3.0%)
Total)	730	2657	447

The big difference between spoken Oshiwambo and written Oshiwambo is the small number of NP:s of type 5 in the spoken material. Especially the number of NP:s belonging to the more complex types 5b, 5c and 5d are quite rare in the spoken material. In this respect it is especially the language of the newspaper articles that is strikingly different from the spoken language.

The number of relative clauses is also smaller in the spoken material, especially those belonging to type 6-7 a). However, we should note that the Oshiwambo relative clauses, especially those belonging to type 6-7 b), are quite loosely added to the head noun (to the extent that it often becomes difficult to tell whether we are really dealing with a relative clause, as mentioned earlier in the discussion of the complex NP:s), and if such a relative clause in itself has a simple structure it does not necessarily make the sentence complex or difficult to produce or to "decode" in a spoken language situation. The relatively high number of NP:s with relative clauses in Oshiwambo when we compare it with the Indo-European languages is partly due to the fact that numerals are often put in a small relative clause (of type b) instead of being simply juxtaposed to the noun, e.g. *aantu ye li 26* (literally 'people which are 26') rather than *aantu 26*, and also partly due to the fact that the equivalents of expressions like 'last week', 'last month', 'next week', 'next month' etc. have relative clauses of this type in them. (footnote: We should also not forget that the percentages of the Oshiwambo data add up to 100 %, whereas those of the European languages do not (see Miller & Weinert 1998:144))

The number of 0 NP:s is higher in the spoken corpus, which should come as no surprise, since many of them are referring to first or second person subjects, i.e. people whose identity can be recovered from the context, with no explicit NP needed.

The very high number of NP:s from types 5b, 5c and 5d in the written material as opposed to the spoken material is in itself a good reason to say that NP:s in written Oshiwambo are more complex. Additional evidence that the written NP:s are more complex is the fact that NP:s of type 1, where nothing modifies the head of the NP, are more common in the spoken material (although the difference is not as striking as in the languages studied by Miller & Weinert, the main reason probably being that in those European languages a pronoun would be used in many cases where Oshiwambo would have a zero subject NP). Furthermore, type 3) is significantly more common in the written material. Even though this type does not include very heavy NP:s the NP:s in this group at least have something modifying the head, whereas those of type 1), which is more common in the spoken material, do not.

CONCLUSIONS AND IMPLICATIONS:

Even a relatively small sample of Oshiwambo texts gives a clear indication that written Oshiwambo has much more complex NP:s than spoken Oshiwambo. Complex NP:s (footnote: Here the expression 'complex NP:s' is used not in the sense 'NP:s belonging to category 9 in the chart above', but in the sense 'NP:s which have a relatively complex structure') is not the only phenomenon that makes written language different from spoken language, but we have at least found one way in which Oshiwambo-speaking children learning to read and write their first language have to go through the same difficult process that English-speaking children go through when they cross the gap from spoken to written language. It is not just a matter of learning the alphabet.² On the more positive side we can say that the similarity in the complexity of NP:s in Oshiwambo and English should mean that a child who is familiar with the complex structures of written

² Interestingly, one of the differences between spoken and written English is not present in Oshiwambo. What Miller & Weinert call "the NP-Clause construction" (like in for example 'The driver he's really friendly') - otherwise often referred to as 'left-dislocation' - can be found in written as well as spoken Oshiwambo. Some examples from my written material: *Uuyuuki nelandulathano oyo iinima mbyoka tayi yuulukiwa nondjuulukwe onene kumboka ya putukile methimbo uuna mbika ya li tayi longo* (literally "Righteousness and order they are the things that...") and *Opolisi yomOshakati nayo oya lopotelwa oshinima shika* (literally "The police of Oshakati they too have been informed about this"). It is also interesting to find in the written material the existence of a relative clause construction like *ya ka tale omulumentu edhina lye Simon Petrus* (=to go and find a man his name is Simon Petrus). This is similar to some relative clause constructions that Miller & Weinert found in spoken language. More research into the differences (and similarities) between spoken and written Oshiwambo is needed.

Oshiwambo can benefit from this when making the transition to (written) English, the only official language of the Republic of Namibia.³

REFERENCES:

- Ekandjo, S. (2002): *Ongeleki methimbo lyuulunde wa putuka*. Oniipa: Eloc Printing Press.
Miller, J. & Weinert, R. (1998): *Spontaneous Spoken Language*. Oxford: Oxford University Press.
Teleman, U. (1983): *Har tal- och skriftspråk olika grammatiker?* Nordlund. 3. Skrifter från institutionen för nordiska språk i Lund.

³ In reality, my experience tells me that most young people of Owamboland today have only a very basic level of literacy in their mother tongue.