



**HAL**  
open science

# A corpus-driven description of o in Naijá (Nigerian Pidgin)

Stefano Manfredi, Slavomír Čéplö

► **To cite this version:**

Stefano Manfredi, Slavomír Čéplö. A corpus-driven description of o in Naijá (Nigerian Pidgin). *Journal of Pidgin and Creole Languages*, 2023, 10.1075/jpcl.22016.man . hal-04109118

**HAL Id: hal-04109118**

**<https://hal.science/hal-04109118v1>**

Submitted on 1 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A corpus-driven description of *o* in Naijá (Nigerian Pidgin)

Stefano Manfredi (SeDyL, CNRS, INALCO, IRD)  
Slavomír Čéplö (CERES, Ruhr-Universität Bochum)

*If I dey lie o*  
*Make Osiris punish me o*  
*Make Ifa dey punish me o*  
*Make Edumare punish me o*  
*Make the land dey punish me o*  
*Make Edumare punish me o*  
*I read am for bookee o*  
*I see some my sefee o*  
*Wellee wellee o*

I.T.T. (International Thief Thief)  
Fela Anikulapo Kuti, 1980

## Abstract

Widely attested in both creole and non-creole languages of the Atlantic basin, the function word *o* has been traditionally described as a ‘sentence/phrase final particle’, owing to its typical syntactic behaviour, rather than to its multiple grammatical meanings. Based on the corpus-driven analysis of the NaijaSynCor, a ~400K words corpus of spoken Naijá (i.e., Nigerian Pidgin), this study suggests that sentence-final *o* can be better described as an ‘illocutionary force indicator’ whose main pragmatic function is to modify the illocutionary force associated with directive and assertive speech acts. The study also provides evidence for the emergence of new coordinating and subordinating functions of *o* in intra-sentential position that are semantically harmonic with its assertive (i.e. epistemic) meaning in sentence-final position. The corpus-driven analysis further shows that the higher occurrence of sentence-final *o* in (formal and informal) dialogic texts in comparison to monologic texts is a reflex of its basic illocutionary function.

## Keywords

*Naijá, corpus-driven analysis, illocution, modality, topicalization, coordination and subordination*

## 1. Introduction<sup>1</sup>

Owing to its large geographic and cross-linguistic distribution, the function word *o* (also found in the literature as *ó, ò, ô, oo*)<sup>2</sup> has been subject of considerable interest in the last decades. In

---

<sup>1</sup> A preliminary version of this paper has been presented at the Summer Meeting of the Society of Pidgin and Creole Linguistics (INALCO, Paris), 28-30 June 2021 with the title “*I see some myself o*” *A corpus-driven analysis of the pragmatic and modal functions of the marker o in Naija*. This study is part of the ANR project NaijaSynCor (*A Corpus-based Macro-Syntactic Study of Naija, aka Nigerian Pidgin*, ANR-16-CE27-0007, <http://naijasyncor.huma-num.fr/>). The authors would like to thank Kofi Yakpo, Maria Mazzoli, Bettina Migge and Bernard Caron for their constructive remarks that greatly contributed to improving the final version of this paper. Any mistakes, or mistaken conclusions, remain ours alone.

<sup>2</sup> These conflicting transcriptions are obviously a reflex of the different phonetic features of *o* across languages as well as of the diverse phonological approaches used by linguists for describing them. As far as Naijá is concerned, Faraclas (1986: 235, 266) claims that *ò* (following his transcription) is “commonly pronounced with a down-stepped high tone” and that there is no evidence for its cliticization. Though we align to this phonological analysis,

his ground-breaking paper *The Story of O*, Singler (1988) defines the sentence-final particle *o* as an “*areal feature par excellence*”, being attested on the two shores of the Atlantic Ocean, in both creole and non-creole languages. Singler convincingly claims that *o* stemmed from Kwa-Kru languages; the morpheme then went through a process of areal diffusion across other languages of the region and eventually entered English and French lexifier Atlantic creoles as a result of substrate interference. Singler further draws a functional parallel between *o* in Niger-Congo languages (ex. 1-2) and the Mandarin sentence-final particle *le* and he argues that both morphemes can be described as markers of a ‘currently relevant state’ since they have a purely pragmatic function.<sup>3</sup>

- 1) *mo rii o*  
I see-him *o*  
“I did see him.”  
(Yoruba [Kwa], Singler 1988: 124)
- 2) *e pu-é ô!*  
It shoot-PASS *ô*  
“It (the gun) was fired!”  
(Klao [Kru], Singler 1988: 123)

Interestingly enough, Singler also stresses that, in the case of Liberian English, *o* underwent a grammatical expansion and that it can signal a perfect aspect, as we can see in example (3).

- 3) *a dõ no wa ples de go o*  
I do not know what place they go *o*  
“I don't know where they have gone.”  
(Liberian English, Singler 1988: 137)

If we shift our attention to the Caribbean area, Migge (2020: 167) shows that in the Eastern Maroon Creoles of Suriname and French Guiana (i.e., Ndyuka, Nenge) *oo* has strong overtones of assertion and that it is largely used “*to draw attention to and alert someone to a problem*”. By adopting a discursive perspective of analysis, Migge demonstrates that *oo* creates a hierarchy whereby the speaker asserts greater knowledge or higher moral status in relation to the interlocutor. In this view, in example (4) *oo* points out that the interlocutor’s assumptions are wrong, whereas in (5) it enhances a vocative, while adding a complaint.

- 4) *a e si enke a kaba oo*  
“She thinks it is over mark my words!”  
(Eastern Maroon Creoles, Migge 2020: 167)
- 5) *M. oo, kon wasi den beenki*  
“M. for god’s sake! Come and do the dishes!”  
(Eastern Maroon Creoles, Migge 2020: 167)

Moving back to the African shores of the Atlantic Ocean, Yakpo (2019: 178, 199, 227-229) describes *ó* in Pichi (the English lexifier creole of Equatorial Guinea) as a ‘sentential modal

---

in this study we will adopt the transcription *o* in order to be consistent with the orthographic system of the NaijaSynCor corpus (§2.1).

<sup>3</sup> The glosses used for the examples of this introduction reproduce the glossing systems adopted by the cited authors.

element’ (glossed as SP, ‘sentential particle’) that marks assertion when signalling presentative/contrastive focus. In syntactic terms, Yakpo analyses *ó* as a sentence-final element, which has scope over all preceding material, being either a predicate-less sentence or a clause. Yakpo translates *ó* as “really, actually, even, at all” and he shows that it may occur in other contexts such as assertive and encouragement statements. Most notably, the addition of the particle *ó* gives directives an admonitive nuance, as we can see in example (6).

- 6) *mék yu mén=an ó!*  
 SBJV 2SG care.for=3SG.OBJ SP  
 “Make sure to take care of her!”  
 (Pichi, Yakpo 2019: 172)

In Pichi, *ó* can also be used vocatively when it occurs in combination with personal names in order to call people or get attention during conversation (7).

- 7) *Lindo ó, Charley ó, una de sí a bin tél dí gál sé*  
 Lindo SP Charley SP 2PL IPFV see 1SG.SBJ PST tell this girl QUOT  
  
*mék e nó hambóg mí ó, a go hát=an.*  
 SBJV 3SG.SBJ NEG bother 1SG.INDP SP 1SG.SBJ pot hurt=3sg.OBJ  
 “Lindo, Charley, you see, I told this girl not to bother me (lest) I might hurt her!”  
 (Pichi, Yakpo 2019: 460)

As far as Naijá is concerned, *ò* has been previously described by Faraclas (1996: 23, 114) as a ‘phrase-final particle’ (glossed as f, ‘final’) that increases the realis or the polite values of different kinds of sentences including perfect/completive VPs (8), imperatives (9), and optionally vocatives.

- 8) *im sel gari fòr maket ò*  
 3SG sell gari for market f  
 “S(he) sold *gari* even in the market.”  
 (Naijá, Faraclas 1996: 114)
- 9) *yu fo go fam ò*  
 2SG R go farm f  
 “You should go to the farm.”  
 (Naijá, Faraclas 1996: 23)

Faraclas (1996: 275) also stresses that, due to its realis connotation, *ò* is widely used in Naijá for signalling ‘emphasis’, but it never functions as question marker or as topicalizer (cf. §3.1, 3.2). In a similar vein, Unuabonah and Rotimi (2018) describe *o* as an ‘emphasis and/or mitigation’ marker in Nigerian English and they highlight its high degree of occurrence when compared to other pragmatic markers borrowed from Yoruba (e.g., *sha*, *abi*, cf. §3.3). Honkanen (forthc.) also confirms this tendency in her quantitative analysis of eighteen Naijá pragmatic morphemes in text-based online communication. On a syntactic ground, it is important to remind that Deuber (2005: 145) warns about a slight variation of *o* in Lagos. According to her, *o* typically occurs sentence-finally, “but in the news texts, it often appears after or even within a (non-final) noun phrase”, as shown in example (10). As we will see (cf. §4), Deuber attributes these intra-sentential occurrences of *o* to the overuse of a Naijá feature in the newscaster’s translations.

- 10) *di Lagos State Government o say dem go work with Parents-Teacher Association*  
“The LSG said they would work together with Parents-Teacher Association.”  
(Naijá, Deuber 2005: 145)

In the light of the above-depicted multi-functionality of *o* in Naijá, the main aim of this paper is to provide a corpus-driven description of this function word in NaijaSynCor corpus, a 400K-word deeply annotated treebank. This means that the description of the syntactic and semantic properties of *o* in Naijá will aim to be comprehensive with respect to synchronic corpus evidence so that its linguistic category will be derived “*systematically from the recurrent patterns and the frequency distributions that emerge from language in context*” (Tognini-Bonelli 2001: 87). Our corpus-driven analysis will illustrate the occurrence of *o* in both sentence-final (finite and non-finite clauses) and intra-sentential positions, while exploring its different grammatical functions in these syntactic domains. As far as the most typical sentence-final *o* is concerned, we propose to synchronically describe its modal (i.e., speaker-oriented and epistemic) and evidential functions in the light of the broader category of ‘illocutionary force indicator’ (Searle 1969, 1975; Searle and Vanderveken 1985), that is a pragmatic device that can influence the expression of the intention to perform a certain illocutionary act (Searle 1969, 1975; Searle and Vanderveken 1985). As we will see, sentence-final *o* in Naijá can either add a polite value to directives or increase the illocutionary force of assertive speech acts by insisting on the speaker’s commitment to the truth of a proposition. This basic illocutionary function of *o* is further confirmed by an analysis of genre variation in the NaijaSynCor corpus that shows a higher occurrence of sentence-final *o* in (both formal and informal) dialogic texts in comparison to monologic texts. As a further matter, our synchronic analysis stresses that the occurrence of *o* in intra-sentential position is the result of an ongoing functional expansion reproducing cross-linguistically common semantic changes between modal and coordinating / subordinating functions. It is also worth noting that, when it appears in coordinated / subordinated clauses, inter-sentential *o* optionally cooccurs with conjunctions (e.g., *but sha (o)*, *if...(o)*, *wheda...(o)*) or complementizers (e.g. *(o) sey*, *(o) to*) that encode different coordinating/subordinating relations. In this regard, our analysis points out that, though synchronic syntactic variation can in long term produce the conventionalization of a morphosyntactic element (the element being a morpheme or a syntactic construction), intra-sentential *o* still represents a minor syntactic option in Naijá and it has not been integrated into narrower grammatical categories (e.g., conjunctions, subordinators or coordinated/subordinated constructions).

The paper is organized as follows. In section 2, we provide information about the NaijaSynCor corpus, the metadata and the methodology used for this study. This section also contains the first quantitative assessment of the occurrence of *o* within the corpus. Section 3 is devoted to a fine-graded description of the syntactic and semantic features of *o* within the NaijaSynCor corpus. We will first analyse the functions of *o* in both sentence-final (finite-clauses) and intra-sentential position, to then focus our attention to its occurrences in ‘Unipartite Clauses’ (i.e., ‘non-finite clauses’, Izre’el 2019). In section 4, we present the statistical analysis intended to assess the role of genre in determining the syntactic behaviour of *o*. In section 5, we finally sum up our findings in terms of synchronic multifunctionality of *o*, while providing an account of its ongoing functional expansion by means of a semantic map.

## 2. Corpus and methodology

### 2.1 The corpus

This study is based on NaijaSynCor, a ~400K words corpus of spoken Naijá. The corpus includes 321 texts representative of 320 speakers coming from over 300 geographical locations

all around Nigeria. That being so, NaijaSynCor reflects a higher degree of individual variation when compared to previous corpora of spoken Naijá (cf. Table 1).

|                        | <b>Faraclas 1996</b>      | <b>Deuber 2005</b> | <b>Mazzoli 2013</b>          | <b>NaijaSynCor 2020</b> |
|------------------------|---------------------------|--------------------|------------------------------|-------------------------|
| <b>Corpus size</b>     | ~400,000 words            | 80,000 words       | ~50,000                      | ~390,000 words          |
| <b>No. of speakers</b> | 60                        | 34                 | 23                           | 320                     |
| <b>Locations</b>       | 1<br><br>(Port Hartcourt) | 1<br><br>(Lagos)   | 2<br><br>(Ajengule, Ilupeju) | ~300                    |
| <b>Text count</b>      | 60                        | 40                 | ?                            | 321                     |

*Table 1: NaijaSynCor corpus and previous corpora of spoken Naijá*

The NaijaSynCor corpus consists of two subcorpora (see Table 2): one, the so-called gold corpus, has been manually annotated in accordance with the Surface-Syntactic Universal Dependencies (Gerdes et al. 2018) which is fully compliant with Universal Dependencies (version 2.10 as of this writing, Zeman et al. 2022). Each token is therefore assigned

- a lemma,
- a part of speech (POS) tag from the Universal POS tagset (Petrov et al. 2012),
- a morphological gloss,
- syntactic annotation (i.e., the head of the current token and the syntactic relationship of the current token to it),
- miscellaneous annotation.

Miscellaneous annotation includes a gloss (combining morphosyntactic and semantic information) and - in case of code-switching - a two-letter ISO 693-1 language code for non-Naijá tokens. Along with these levels of annotation, each sentence is provided with metadata which include sentence and speaker codes and a translation into English. The segmentation into maximal syntactic units has been developed in accordance with minimal morphosyntactic glossing (i.e., function words do not have glosses beyond their morphological features, Caron et al. 2019).<sup>3</sup> In addition to this annotation, the corpus contains a macrosyntactic annotation based on the system developed for the Rhapsodie corpus of French (Pietrandrea and Kahane 2019). This takes the form of special punctuation marking illocutionary units, hesitation, repair, breaks etc. For the purposes of corpus management and querying, this annotation is interpreted as punctuation tokens. The other subcorpus, i.e., the non-gold one, has been annotated in much the same manner as the gold one, but automatically and excludes morphological glosses. For the purpose of this paper, both corpora were combined into a single corpus and imported into an instance of NoSketchEngine (Rychlý 2007, Kilgariff et al. 2014). This involved direct conversion from the CONLL-U format to the vertical text format accepted by the NoSketchEngine. Table 2 below provides detailed information on the composition and annotation of both NaijaSynCor subcorpora.

|                                 | NaijaSynCor gold corpus   | NaijaSynCor non-gold corpus   |
|---------------------------------|---|---|
| Corpus size<br>(NoSketchEngine) | 140,729 tokens  | 250,083 tokens  |
| Text count                      | 88  | 233   |
| Speaker count                   | 87  | 268 (overlapping with the gold corpus)  |
| Annotation                      | macrosyntax, lemma, part of speech, morphology, syntax, other   | macrosyntax, lemma, part of speech, syntax, other   |
| Data access                     | <a href="https://arboratorgrew.elizia.net/#/projects/Naija_Gold_Reference">https://arboratorgrew.elizia.net/#/projects/Naija_Gold_Reference</a> | <a href="https://arboratorgrew.elizia.net/#/projects/Naija_Nongold">https://arboratorgrew.elizia.net/#/projects/Naija_Nongold</a> |
| User access                     | <a href="https://www.bulbul.sk/crystal/#dashboard?corpname=NaijaSynCor">https://www.bulbul.sk/crystal/#dashboard?corpname=NaijaSynCor</a>       |   |

Table 2: *NaijaSynCor* corpus and its division

As far as transcription is concerned, the corpus does not provide any phonetic nor phonological information about Naijá, as the corpus has been fully transcribed by Naijá speakers according to an informal orthography largely modelled after that of Standard English. In order to facilitate retrieving of allomorphic items, the corpus adopts different orthographic conventions for disambiguating lexical items (e.g., lexical verb *say* “to say”) from their grammatical counterparts (e.g., complementizer *sey* “that”). Despite this and the lemmatization in the gold corpus, some variation remains. The translation of all the sentences into English has equally been done by native speakers of Naijá, as has the marking of code-switching. For code-switching between Naijá and English which in many instances results in ambiguity, the native speaker’s intuition about the prosody of the utterance or passage in question was used as the main criterion to differentiate between the two languages.

## 2.2 *O* in the *NaijaSynCor* corpus

For the purposes of this paper, we examined and manually annotated the occurrences of the lemma *o* in the gold corpus only. A total of 685 relevant occurrences were identified which brings the relative frequency of the lemma *o* in the gold corpus to 4687 per million. The relative frequency of the lemma *o* across the entire *NaijaSynCor* corpus is 6026 per million, making it the 26th most frequent item, just behind the perfect marker *don* (23) and ahead of the conjunction *and* (27) and the verb *come* (28). The high frequency of *o* in the lexicon of Naijá is further evidenced by the fact that of the 87 speakers in the gold corpus, only 7 do not use it in their texts; this strongly suggests that there is no sociolinguistic variation in the use of *o*.

As we will see in the following sections, the syntactic annotation of the *NaijaSynCor* allows us to make a distinction between sentence-final and intra-sentential *o*, the former occurring before a major syntactic boundary (i.e., //, see Image 1a, §3.1), the latter before a minor syntactic boundary (i.e., <, see Image 1b, §3.2) or within the clause.

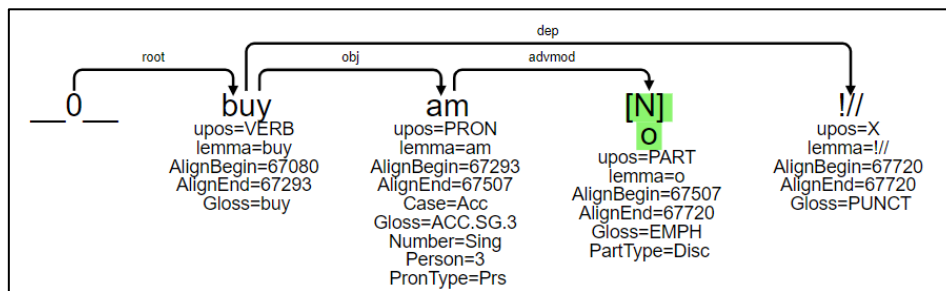


Image 1a. The syntactic annotation of sentence-final *o* in *NaijaSynCor*

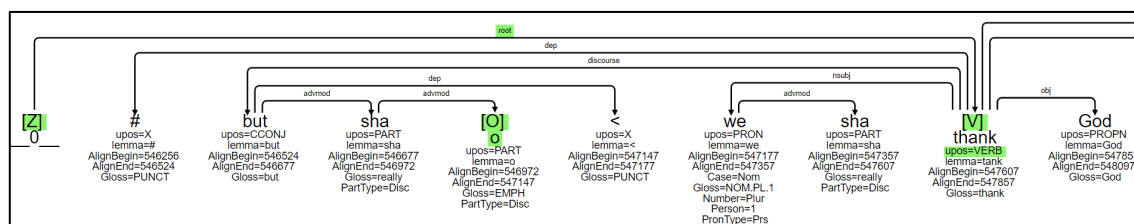


Image 1b. The syntactic annotation of intra-sentential *o* in NaijaSynCor

However, we will also show that *o* can also appear at final of Unipartite Clauses (§3.3), both within and at the end of major syntactic units.

### 3. The syntax and semantics of *o* in the NaijaSynCor corpus

This section provides a description of the syntactic and semantic features of *o* in Naijá. The different grammatical meanings expressed by *o* will be analyzed according to their distribution across three syntactic domains: in sentence-final position (§3.1), in clause-internal position (§3.2), and in Unipartite Clauses (§3.3).

#### 3.1 Sentence-final *o*

In the NaijaSynCor corpus, *o* typically occurs in sentence-final position (577 occurrences, 84,2% of the total occurrences), immediately before a major syntactic boundary (i.e., //; cf. 2.2). When appearing sentence-finally, *o* has scope over all preceding material and it can be associated with both modal and evidential meanings. **First** of all, *o* is often (72 occurrences) related to a speaker-oriented modality when it occurs in directive utterances. In this case, *o* adds a polite value to the expression of conditions on the agent with respect to the completion of the predicate situation (Bybee et al. 1994: 177). The politeness-marking function of *o* in such directive contexts could be explained by the fact that ‘flat directives’ are often unsuitable in conversation as speakers tend to find indirect means to reach their illocutionary goals in accordance with requirements of politeness (Mauri and Sansò 2011). In fact, *o* can be well be considered as an integral part of Naijá ‘directive strategies’ (i.e., strategies through which the speaker orders someone to do something), as it can occur in combination with imperative, prohibitive, hortative, and jussive moods. In the case of imperatives (ex. 11) and prohibitives (ex. 12), *o* occurs at the end of a sentence in which the addressee corresponds to a 2<sup>nd</sup> person, the verb being unmarked.

- 11) # *buy am o !!!*  
“Buy it!”  
(ENU\_17\_Buying-Grocery\_DG\_\_47)
- 12) < # *no enter second service o !!!*  
“Don’t go for the second service”  
(WAZP\_04\_Ponzi-Scheme\_MG\_\_102)

Hortatives (ex. 13) and jussives (ex. 14-15) featuring a sentence-final *o* are instead introduced by the auxiliary verb *make* and their agent corresponds to either a 2<sup>nd</sup> person or to a 1<sup>st</sup>/3<sup>rd</sup> person subject, respectively.

- 13) < *make you use your brain o !!!*  
“Apply wisdom!”  
(IBA\_33\_News-Comments\_MG\_\_3)



- 14) # *make she kukuma just arrange hersef o* //  
 “Let her just adjust herself.”  
 (ENU\_13\_School-Life\_DG\_\_94)
- 15) # *hey make I buy am for my children o !* //  
 “Hey, let me buy it for my children!”  
 (ENU\_17\_Buying-Grocery\_DG\_\_46)

Despite the above-described propensity to express a speaker-oriented modality in combination with directives, assertive utterances are by far the most common context of occurrences of sentence-final *o* (362 occurrences, i.e., 63% of the sentence-final occurrences and 53% of the total occurrences). When it is used assertively, *o* expresses an epistemic modality and it functions as clausal-scope indicator of the speaker’s commitment to the truth of a proposition. In this context, *o* typically follows pragmatically unmarked utterances, whether positive (ex. 16-17) or negative (ex. 18-19) NPs/VPs, without any specialization with regard to aspect (i.e., perfective/imperfective) or mood (i.e., realis/irrealis).

- 16) # *e cost o !* //  
 “That’s expensive!”  
 (ENU\_17\_Buying-Grocery\_DG\_\_9)
- 17) # *you go use cube o* //  
*You will use seasoning cubes.*  
 (PRT\_01\_Banga-Soup\_MG\_\_36)
- 18) # *I no fit go like dat o* //  
 “I can’t go like that.”  
 (ABJ\_GWA\_08\_David-Lifestory\_MG\_\_180)
- 19) # *no be small o* //  
 “It’s not a small thing.”  
 (IBA\_02\_Igwe-Festival\_MG\_\_12)

The epistemic meaning of sentence-final *o* can also increase the assertive value of pragmatically marked utterances. This is the case of relative clefts in which the left-dislocated referent is introduced by the focus particle *na* (see Caron 2019 for a detailed description), with *o* at the end of the following relative clause (82 occurrences). Here, *o* has scope over the cleft construction and does not mark a contrastive focus. Rather, it underscores the assertive value of the utterance, as shown by the following examples.

- 20) # *dis one < na im >+ dey come o* //  
 “This one, this is the one who is still on the way.”  
 (IBA\_32\_Tori-By-Samuel\_MG\_\_62)
- 21) # *I want tell di person sey [ hope sey [ you no mind // = na Pidgin >+ we dey speak o !/ ] ]* //  
 “I want to tell the person: “Hope you don’t mind, it’s Pidgin what we speak.””  
 (WAZL\_08\_Edewor-Lifestory\_MG\_\_58)

- 22) # *na only fifty naira* >+ *can come out o* //  
 “It’s only fifty naira that I can take off.”  
 (ENU\_17\_Buying-Grocery\_DG\_\_24)
- 23) *na im* [*naim*] *God change my life o* //§  
 “And that is how God who changed my life.”  
 (WAZL\_15\_MC-Abi\_MG\_\_48)

The same applies to focused NPs that feature the focus marker *na*, but without any left-dislocation (47 occurrences). In this case too, *o* is relevant only for the interpretation of the speaker’s commitment to the truth of the proposition.

- 24) < *na hot slap o* // §  
 “That’s a big slap.”  
 (PRT\_05\_Ghetto-Life\_MG\_\_15)
- 25) # *na my son o* // §  
 “He’s my son.”  
 (PRT\_07\_Drummer\_MG\_\_32)
- 26) # *Hey dis one na wahala o* // §  
 “Hey this is a problem.”  
 (ABJ\_GWA\_02\_Market-Food-Church\_DG\_\_102)

As is widely known, there is a cross-linguistic overlap between the expression of epistemic modality (i.e., assertion) and that of evidentiality, intended as the grammatical encoding of the information source (Aikhenvald 2004). In this regard, Bybee et al. (1994: 323) point out that *evidentials* indicate the source of the information on which an *assertion* about a situation is based. De Haan (1999) instead argues that evidentiality and epistemic modality differ in that the former *assert* the nature of the evidence for the information in the sentence, while the latter *evaluate* the speaker’s commitment to the statement. Furthermore, De Haan claims that evidential morphemes and epistemic modals often present different etymologies. In a similar vein, Cornillie (2009) explains that, even if a source of information can be attributed different degrees of reliability, this does not automatically correspond to the epistemic commitment of the speaker. On the other hand, Aikhenvald openly mentions the possibility for epistemic modal markers to be used as ‘evidentiality strategies’ (i.e., evidential extensions of non-evidential categories, 2004: 276–279, 284–285). This actually seems to be the case of *o* in Najjá as it can insist on the speaker’s commitment to the utterance as direct information source, with reference to both a recent visual (27) and non-visual (28) past.

- 27) # *na so* >+ *we see am o* //  
 “That’s how we see it”  
 (KAD\_15\_Money-Wahala\_MG\_\_39)
- 28) # *my sister* < *na so* >+ *I hear am o* //  
 “My sister, it’s like that, I heard it”  
 (D\_ABJ\_GWA\_02\_Market-Food-Church\_TRANS, tx@SP2)

More generally, our data confirm Faraclas’s claim (cf. §1) that *o* cannot occur at the end of interrogative sentences. In our corpus, there are only 2 ambiguous occurrences of sentence-final *o* in Yes/No questions in which it has neither modal nor evidential meaning. There are, however,

7 occurrences of a sentence-final *o* in wh- questions (ex. 29). In this case, *o* rather has an affective meaning linked to the expression of the speaker’s astonishment or surprise, as shown in the following example.

- 29) # *I tell am sey [ which kind work be dat o ?//] //*  
 I asked him: what kind of job is that?  
 (ENU\_22\_Barman-Interview\_MG\_\_5)

Lastly, sentence-final *o* can also limitedly occur in combination with the Yoruba-sourced pragmatic marker *sha* (3 occurrences, cf. §3.2). Table 3 summarizes the distribution of sentence-final *o* in the gold portion of the NaijaSynCor corpus.

| Type                                  | Absolute count | Relative percentage |
|---------------------------------------|----------------|---------------------|
| directive                             | 72             | 12.48%              |
| assertive (unmarked)                  | 362            | 62.74%              |
| relative cleft with <i>na</i>         | 82             | 14.21%              |
| focused NP ( <i>na</i> without cleft) | 47             | 8.15%               |
| evidential                            | 2              | 0.35%               |
| yes/no questions                      | 2              | 0.35%               |
| wh- questions                         | 7              | 1.21%               |
| <i>sha o</i>                          | 3              | 0.49%               |
| <b>TOT</b>                            | <b>577</b>     | <b>100%</b>         |

Table 3: the functional distribution of sentence-final *o*

### 3.2 Intra-sentential *o*

Intra-sentential *o* can occur either immediately before a minor syntactic boundary (i.e., <) or within a sentence ending with a major syntactic boundary (i.e. //). With 42 occurrences (6.1% of the total occurrences), intra-sentential *o* undoubtedly represents a minority syntactic option in Naijá. However, it is also important to stress that, when it occurs clause-internally, *o* constitutes an integral part of syntactically complex constructions. First, the most common context of occurrence for intra-sentential *o* in the NaijaSynCor corpus is represented by frame setting topics (14 occurrences) in which it optionally marks the boundary of the left dislocated frame. Following Chafe (1976: 50), we here define ‘frame setting topics’ as adverbial expressions that set “*the spatial, temporal or individual framework [...] which limits the applicability of the main predication to a certain restricted domain*”. The pragmatic interpretation of frame setting topics involves no anaphoric processing, because there is no referent that gets activated in the left-dislocated frame. In Naijá, *o* most typically marks temporal frame setting topics (ex. 30-31).

- 30) # *because ehn I remember < dat Jonathan time o < everybody just dey flex |*  
 “Because I remember that, in Jonathan’s era, everyone was just showing off.”  
 (LAG\_07\_Johns-Biography\_MG\_\_4)
- 31) # *around ten p.m. on January twenty one o < # na im di guy go di restaurant |*  
*go threaten im oga o //*  
 “Around ten pm on January 21, the guy went to threaten his Mistress in the restaurant.”  
 (IBA\_33\_News-Comments\_MG\_\_65)

*O* also rarely occurs at the boundary of left-dislocated argument topics in which a nominal referent is activated (5 occurrences). This is shown in example (ex. 32) in which the referent *gote* in the *o*-marked topic is co-referential with the pronoun *it* in the following comment.

- 32) # *and dis gote o* < # *we like it very much because it is very nutritional* //  
 “And this *gote* (soup), we like it very much because it is very nutritional”  
 (OS\_12\_How-To-Prepare-Gote-Soup\_MG\_\_41)

Different from what we have seen in the previous examples of frame-setting and arguments topics, intra-sentential *o* can optionally occur in combination with conjunctions and complementizers that encode different kinds of coordinating/subordinating relations. In fact, conditional clauses represent the second most common context of occurrence for intra-sentential *o* (12 occurrences). This is actually not surprising, since it is often assumed that conditional clauses and topics share a number of syntactic, semantic and prosodic properties in many languages (Haiman 1978, cf. §5). *O*-marked conditional clauses in *Naijá* include both truth and concessive conditionals. In the case of truth conditionals, *o* follows a protasis introduced by *if*, while insisting on the condition under which the following subordinated clause holds true, as shown by examples (33-34).

- 33) < # *if you must drive o* < *if e be sey you must drive motor* < #  
*make you no take alcohol* > #  
 “If you must drive, if you have to drive a car, don’t drink alcohol.”  
 (LAG\_31\_Road-Safety\_MG\_\_69)

- 34) < # *if na fish you want to use o*  
 < # *no problem* > #  
 “If you want to use a fish, there’s no problem”  
 (PRT\_01\_Banga-Soup\_MG\_\_12)

In the case of alternative concessive conditionals, *o* occurs at the end of a protasis introduced by *wheda* ‘whether’. Here, *o* rather stresses the factual nature of the conditional clauses, since the main function of concessives is to say that, in spite of some negative implication, the main clause assertion holds true (Bybee et al. 1994: 175, 225).

- 35) # *as you see* < # *wheda { Christian |c Muslim } o* //  
 < *we dey live togeder* //  
 “As you see, whether you are a Christian or a Muslim, we all live together.”  
 (JOS\_20\_Beauty-Of-Jos\_MG\_\_36)

- 36) # *weda {you be my daughter |c or not } o* //  
*I go call you* //  
 “Whether you are my daughter or nor, I will call you.”  
 JOS\_10\_Mothers-Against-Mini-Skirts\_DG\_\_51

The third context of occurrence for intra-sentential *o* are contrastive adversative coordinated clauses (7 occurrences). In *Naijá*, a contrastive adversative coordination is expressed at the beginning of the clause by means of the complex conjunction *but sha (o)* “but well, but in fairness” which features both the English-derived conjunction *but* and the Yoruba-sourced

pragmatic marker *sha* (cf. Unuabonah and Rotimi 2018), in optional combination with *o*. In Naijá, *but sha (o)* is always prosodically isolated and it can express the denial of both the speaker's (37) and the interlocutor's (38) expectations.

37) # *but sha o < di Permanent Secretary <*  
*dem say [ she no gree go for di Committee //] //*  
 “But in fairness, the Permanent Secretary, they said she refused to meet the Committee.”  
 (BA\_32\_Tori-By-Samuel\_MG\_\_64)

38) # *but sha o < before you do ofofo to collect money <*  
*reason am { well |r well } o //*  
 “But well, before you engage in gossip in order to collect money, think about it deeply.”  
 (IBA\_33\_News-Comments\_MG\_\_5)

Lastly, the NaijaSynCor corpus also gives evidence of the occurrence of intra-sentential *o* in complement clauses. On the one hand, *o* can occur in inter-sentential position within perception complement clauses that refer to the way an experiencer perceives the occurrence of a given state of affairs (2 occurrences marked as “source of information” in Table 4; cf. Cristofaro 2003: 105). In this case (ex. 39), *o* appears between the perception predicate (e.g., “hear”) and the complementizer *sey* (Eng. “that”) introducing the following subordinate clause. This subordinating function of *o* in perception clauses is harmonic with its evidential meaning in sentence-final position (cf. ex. 27-28, §2).

39) # *we hear o sey [ Otuyi < na im >+ get di restaurant ] //*  
 “We heard that Otuyi is the one who owns the restaurant”  
 BA\_33\_News-Comments\_MG\_\_47

On the other hand, intra-sentential *o* can also appear in modal complement clauses featuring implicative verbs such as “try” or “manage to”. In this case (ex. 40), *o* appears between the main verb and the following *to*-infinitive subordinate clause. This subordinating function (4 occurrences total) is instead related to the epistemic (i.e., assertive) meaning of *o* in sentence-final position (cf. ex. 16-19, §2), given that implicative verbs indicate that the main clause commits the subject to an implied proposition, which consists of the complement clause (Karttunen 1971).

40) # *e dey make attempt o to withdraw some money from one ATM ni #*  
 “He tried to withdraw some money from an ATM”  
 IBA\_33\_News-Comments\_MG\_\_103

In the light of the above, we can argue that, if the function of intra-sentential *o* in topic/comment structures is similar to that of its sentence-final counterpart, in that it signals assertion with reference to the comment, the use of *o* in adversative coordinated and subordinated (conditional and complement) clauses rather points to a gradual functional expansion from a marker of epistemic (i.e., assertive) modality to a coordination/subordination device. It should, however, be stressed that *o* is only optionally used in combination with conjunctions and complementizers such as *but sha (o)*, *if...(o)*, *wheda...(o)*, *(o) sey*, and *(o) to*, and that these phrasal compounds are not conventionalized yet. The fact remains that the multifunctionality of *o* in intra-sentential position reflects cross-linguistically common semantic connections

between modal and post-modal meanings (cf. §5). The following table summarizes the functional distribution of intra-sentential *o*.

| Type                     | Absolute count | Relative percentage |
|--------------------------|----------------|---------------------|
| frame-setting topic      | 14             | 31.82%              |
| conditional              | 12             | 27.27%              |
| adversative coordination | 7              | 15.91%              |
| source of information    | 2              | 4.55%               |
| subordinator             | 4              | 9.09%               |
| argument topic           | 5              | 11.36%              |
| <b>TOT</b>               | <b>44</b>      | <b>100.00%</b>      |

Table 4: The functional distribution of intra-sentential *o* (percentages are rounded)

### 3.3 *O* in Unipartite Clauses

The above analysis of the semantic features of *o* has been primarily driven by the syntactic annotation of the NaijaSynCor corpus that allows for a distinction between minor (i.e., <) and major (i.e., //) syntactic boundaries (Courtin et al. 2018, Caron et al. 2019, cf. §2). However, in almost 10% of its total occurrences (63 occurrences), *o* appears in the final position of Unipartite Clauses that can occur both within and at the end of major syntactic units.<sup>4</sup> Following Izre’el (2018), we define a Unipartite Clause as a minimal syntactic unit whose only necessary and sufficient component is the predicate (with no subject required). Here, the predicate is viewed as the element (not necessarily a verb or a noun) that carries the informational load of the clause, while expressing a broad modal sense (ranging from assertive through interrogative and vocative to imperative, Izre’el 2018: 242). In this perspective, Unipartite Clauses must be seen as syntactically coherent utterances that fall outside the scopes of clausal finiteness (i.e., ‘non-finite clauses’). Vocatives, used for getting attention from the interlocutor, are a common example of Unipartite Clauses. As in other Atlantic Creoles (cf. §1), Naijá vocatives can be optionally marked by *o* (7 occurrences, cf. 1). Though always prosodically independent,<sup>5</sup> *o*-marked vocatives can occur before minor (ex. 41) and major (ex. 42) prosodic boundaries within our corpus.

41) *ah my sister o < mtschew toh I eat eba o !//*  
 “Ah my sister, well, I ate eba!”  
 (ABJ\_GWA\_02\_Market-Food-Church\_DG\_\_79)

42) *thanks > ma o //*  
 “Thank you, ma.”  
 (ENU\_17\_Buying-Grocery\_DG\_\_159)

The same is true for the courtesy interjection *abeg* “please” that is often associated (6 occurrences) with both sentence-final (ex. 43) and intra-sentential (ex. 44) *o* for adding a polite value to the expression of a co-hortative.

<sup>4</sup> The notion of Unipartite Clause is a part of the framework underlying the syntactic annotation of the NaijaSynCor corpus. That being so, Unipartite Clauses are not associated with syntactic boundaries within the corpus. In this paper, we chose to adopt the notion of Unipartite Clause in order to better describe the grammatical roles played by *o* in the NaijaSynCor corpus, as well as to diachronically reconstruct its semantic expansion across different syntactic domains (cf. §5).

<sup>5</sup> The independent prosodic status of vocative is most commonly evidenced by a pause with an optional pitch reset.

43) *abeg o !//*  
 “Please!”  
 (ABJ\_GWA\_02\_Market-Food-Church\_DG\_\_99)

44) *abeg o < make we moders dey talk to our daughters now //*  
 “Please, let us mothers give guidance to our daughters.”  
 (JOS\_10\_Mothers-Against-Mini-Skirts\_DG\_\_33)

Other self-standing courtesy and addressing interjections that can occur with *o* include *na wa o*, *na wa for (...)* *o* “wow, it is a shame, congratulations to...” (9 occurrences), *hey o* “hey” (4 occurrences) and *oh o* (2 occurrences). This broad category also includes greetings and religious invocations, e.g. *thank God o* (single occurrence) and *amen o* (2 occurrences). Furthermore, *o* can also express an assertive meaning when occurring in Unipartite Clauses. This epistemic function is typical of answers to Yes/No questions (ex. 45, 24 occurrences) or absolute negative/positive assertions (ex. 46, 4 occurrences).

45) *< yes o ! //*  
 “Yes!”  
 (IBA\_03\_Womanisers\_MG\_\_13)

46) *< na di youth dem ] < no o ! //*  
 “This is the youth. No!”  
 (WAZA\_05\_Big-Mo\_MG\_\_33)

*O*-marked answers to yes/no questions also come with other interjections, including *okay o* (8 occurrences), *alright o* (3 occurrences) and even the Yoruba-sourced *aye o* (1 instance). In summary, the functions of attention calling/addressing and yes/no answering expressed by *o* in Unipartite Clauses are evidently harmonic with the directive and assertive meanings expressed by sentence-final *o*, and also give evidence of common semantic connections (cf. §3.1, 5). Table 5 summarizes the distribution of *o* in Unipartite Clauses.

| Type                        | Absolute count | Relative percentage |
|-----------------------------|----------------|---------------------|
| vocative                    | 7              | 11%                 |
| courtesy                    | 28             | 44%                 |
| yes/no                      | 24             | 38%                 |
| positive/negative assertion | 4              | 7%                  |
| <b>TOT</b>                  | <b>63</b>      | <b>100%</b>         |

Table 5: The distribution of *o* at final of Unipartite Clauses

#### 4. *O* and genre variation

In her variationist analysis of Naijá in Lagos, Deuber argues that “[...] one would expect *o* to be most frequent in the dialogic or persuasive text categories – conversations, drama and advice. However, it is in the news texts that *o* is used most often.” (2005: 145). In this section, we address the question of whether, why and how the occurrence of *o* in Naijá is affected by genre (i.e., text type or text category) variation. More particularly, we aim at refuting Deuber’s claim that intra-sentential *o* is a product of its “overuse” in newscaster’s translations (cf. §1). In this context, it should be recalled that the NaijaSynCor corpus reflects a large range of text genres, including news translations in radio broadcasting (see table 6 below for a detailed breakdown); as such, it represents an optimal test for genre variation. As a further matter, it has

already been argued that text genre predominates over social factors such as age, sex, and degree/kind of formal education in ruling morphosyntactic variation within the NaijaSynCor corpus (Manfredi et al. 2019).

| Genre                      | Text count (gold corpus) | Wordcount     | Wordcount percentage |
|----------------------------|--------------------------|---------------|----------------------|
| drama                      | 2                        | 1317          | 0.94%                |
| formal conversation        | 0                        | -             | -                    |
| informal conversation      | 9                        | 18423         | 13.08%               |
| opinion                    | 5                        | 6845          | 4.86%                |
| persuasion and instruction | 14                       | 21880         | 15.54%               |
| radio - conversation       | 0                        | -             | -                    |
| radio - monologues         | 7                        | 12490         | 8.87%                |
| radio - reading news       | 7                        | 7314          | 5.19%                |
| religion                   | 3                        | 5396          | 3.83%                |
| storytelling               | 41                       | 67172         | 47.69%               |
| <b>TOT</b>                 | <b>88</b>                | <b>140837</b> | <b>100.00%</b>       |

Table 6: Genres in NaijaSynCor (gold corpus)

With regard to Deuber’s analysis in general, there is a large caveat. The NaijaSynCor classification of genres is somewhat different from Deuber’s: where NaijaSynCor has three different text categories for recordings of two or more people speaking differentiated by social context, Deuber classifies such dialogic texts into three different groups, “Interviews”, ‘Discussions’ and ‘Conversations’ without making a clear distinction between these text types (see Table 5.10, Deuber 2005: 145, reproduced as Figure 1 below).

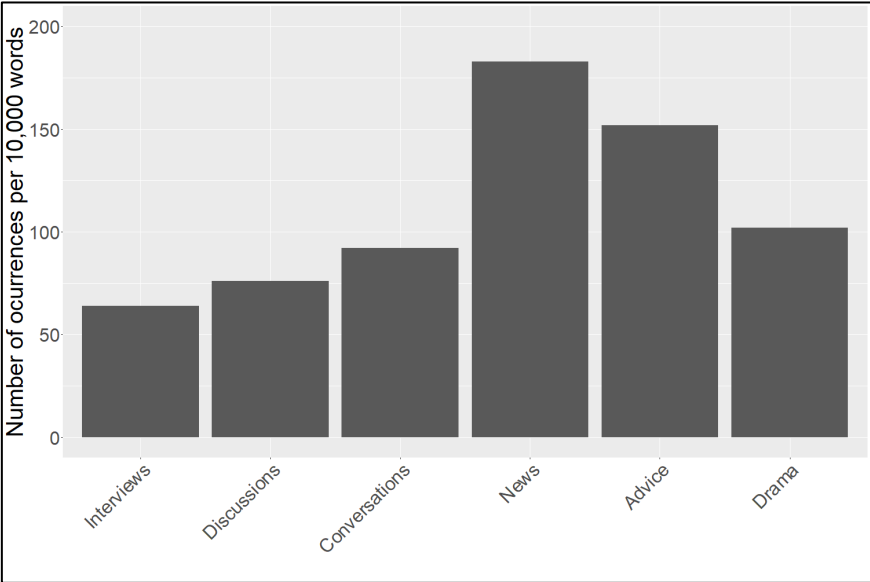


Figure 1: Deuber’s analysis of the distribution of o across different genres (2005: 145)

This has, of course, ramifications on Deuber’s analysis: if one were to group all three dialogic categories together, there would be little justification for her conclusion that “it is in the news texts that o is used most often” cited above. Compare the statistical analysis of o in the gold part of NaijaSynCor in Figure 2 below which clearly shows that in NaijaSynCor, radio news texts rank far below two other genres (informal conversations and radio monologues) when it comes to the occurrence of o.



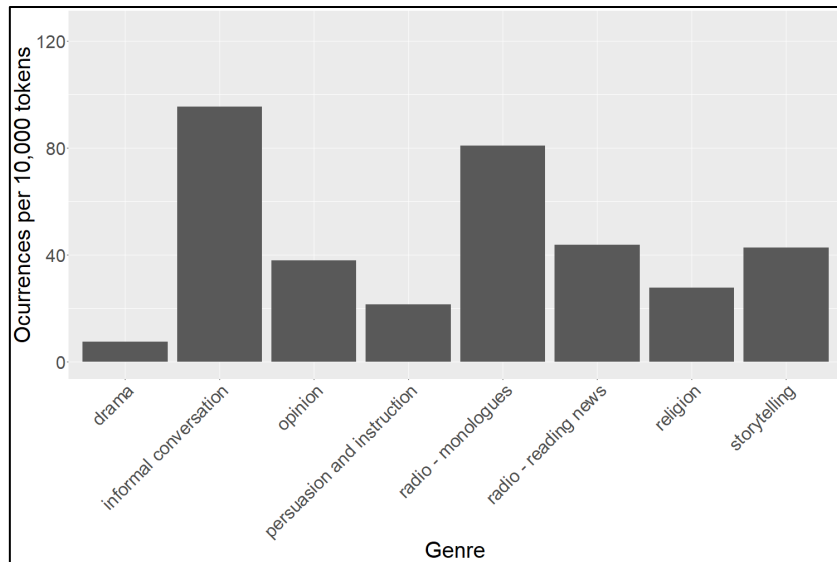


Figure 2: Statistical distribution of *o* across genres in the NaijaSynCor gold corpus

The difference between Deuber’s and NaijaSynCor data is even more evident when grouping the NaijaSynCor fine-grained genres into supergenres based on their monologic and dialogic nature. According to Deuber, one would expect that dialogic (and persuasive) texts would contain more occurrences of *o* and based on her data she concludes that this is not the case. With NaijaSynCor data, however, this expectation holds: when grouping the genres into dialogic (‘informal conversation’, ‘formal conversation’, ‘radio - conversation, drama’) and monologic (‘storytelling’, ‘persuasion and instruction’, ‘opinion’, ‘religion’, ‘radio - reading news’, ‘radio – monologues’), a clear picture emerges where *o* is twice as likely to appear in dialogic texts than in monologic ones (see Figure 3 below).

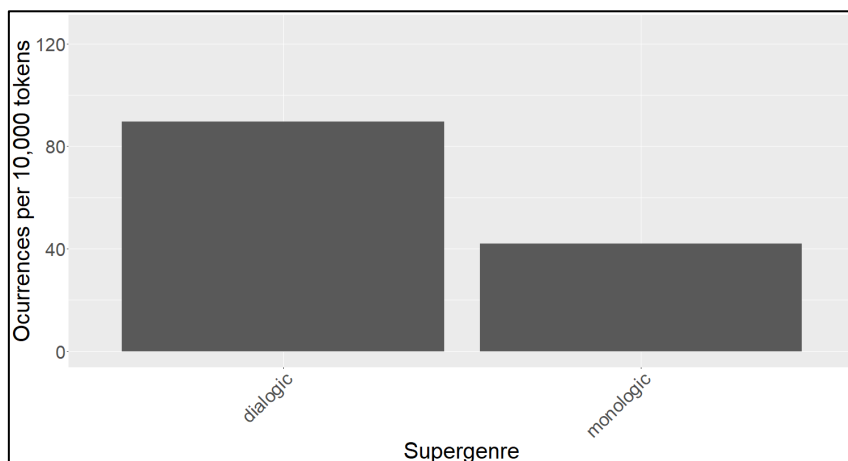


Figure 3: Statistical distribution of *o* across monologic/dialogic texts in the NaijaSynCor gold corpus

Even with other definitions of supergenres, as with a three-way split where the NaijaSynCor genres ‘persuasion and instruction’ and ‘religion’ (the latter of which consists mostly of sermons) are classified as persuasive texts (Figure 4), the prevalence of *o* in dialogic texts holds.

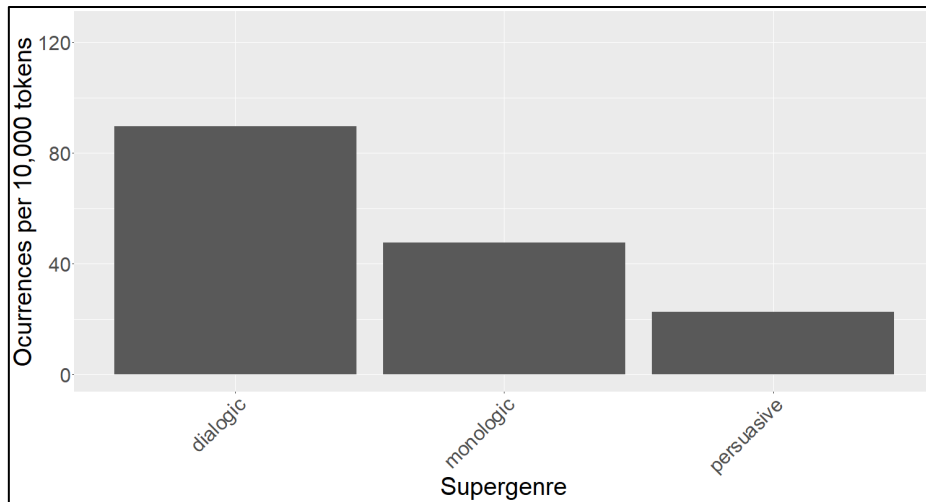


Figure 4: Statistical distribution of *o* across monologic/dialogic/persuasive texts in the NaijaSynCor gold corpus

In short, the NaijaSynCor data contradicts Deuber’s analysis, but confirm her initial intuition that based on the functions of *o*, it is more likely to appear in dialogic contexts. Concerning the grammatical/pragmatic functions of intra-sentential *o*, it should be remarked they are evidently the product of a semantic expansion from its basic assertive meaning in sentence-final position (cf. §5). Such semantic developments can hardly be explained (exclusively) by the “overuse” of Naijá features in news translations, as they rather reflect a series of diachronically coherent changes across different grammatical domains (i.e., modality, information structure, coordination, and subordination). As a further matter, recent variationist studies have shown that the use of epistemic and evidential markers such as *o* (cf. §3.1) is more typical of informal registers than the formal ones (Gonzales et al. 2017). The statistical analysis of the occurrence of intra-sentential *o* in NaijaSynCor across genres supports this view, as it shows no significant increase in the use of intra-sentential *o* in radio newscasts as opposed to other genres (Figure 5).

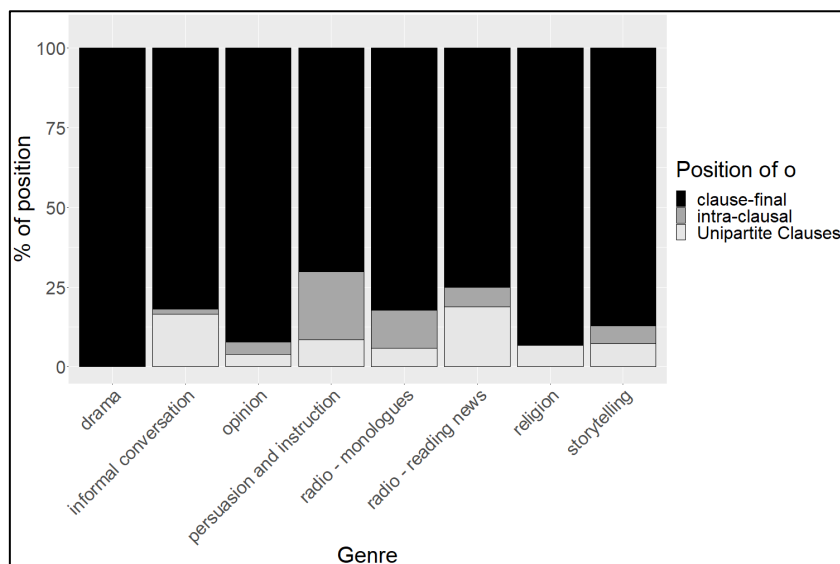


Figure 5: Statistical distribution of *o* (by position) across genres in the NaijaSynCor gold

In fact, the ratio of intra-sentential *o* in radio newscasts (2.73 per 10,000 tokens) is roughly the same as that of storytelling texts (2.38/10,000). This and the relatively high occurrence of intra-sentential *o* in radio monologues (9.6/10,000) and persuasive/instructional texts (4.57/10,000)

compared to the relatively low rate for conversational texts (none in drama, 1.63/10,000 in informal conversations) would suggest a differentiation in the functional load of *o* in the different positions. However, the low absolute numbers at our disposal make any definite conclusion premature.

## 5. A corpus-driven description of *o* in Naijá

The previous corpus-driven analysis of *o* in Naijá widens our grammatical understanding of this function word in several aspects. On syntactic grounds, the synchronic corpus clearly shows that *o* cannot simply be described as a ‘sentence-final particle’. In actuality, despite the fact that *o* most commonly occurs sentence-finally (i.e., at the end of both finite and non-finite clauses), it can also appear in intra-sentential position (i.e., before a minor syntactic boundary or within a sentence). On semantic grounds, we have shown that sentence-final *o* is related either to a directive or an assertive function, the latter providing the context for the ongoing emergence of new coordinating and subordinating functions in intra-sentential position (see below). This brings us to the first question we want to address in this section, namely how can we synchronically describe the multiple functions of sentence-final *o* in Naijá?

First, it should be remarked that, contrary to what has been written by Faraclas (1996), sentence-final *o* is related neither to the expression of a perfective aspect nor of a realis mood in Naijá, as it can occur with irrealis-marked verbs (i.e., future, cf. ex. 17-18). We have instead seen that, when occurring sentence-finally, *o* typically expresses a subject-oriented modality in directive utterances (ex. 11-15) or an epistemic modality in pragmatically unmarked assertive utterances (ex. 16-19). The epistemic value of sentence-final *o* also applies to pragmatically marked utterances introduced by the focus marker *na*, whether relative clefts (ex. 20-22) or focused NPs (ex. 23-25). We have also pointed out that, when sentence-final *o* co-occurs with *na*, it is not directly associated with the expression of a contrastive focus as it rather functions as clausal-scope indicator of a speaker’s commitment to the truth of a focused proposition. Furthermore, sentence-final *o* in Naijá cannot be described as an ‘emphasis’ marker (Faraclas 1996; Unuabonah and Rotimi 2018), since it does not have nuclear scope over a single element of the clause, but it rather modifies the clause as a whole.

In view of the above, we suggest to synchronically describe sentence-final *o* in Naijá as an ‘illocutionary force indicator’ (Searle 1969, 1975; Searle and Vanderveken 1985) whose role is to modify the illocutionary force associated with different ‘illocutionary (speech) acts’. Austin (1962: 116, 121, 139) first introduced the notion of ‘illocutionary (speech) act’, which is intended not simply as the act of *uttering* something, but also as the act of *doing* something (i.e., asserting, ordering, warning, asking, promising, etc.). Searle (1969) then pointed out that every illocutionary speech act has the structure F (P), where F represents the illocutionary Force and P is the Propositional content. F and P are represented in the syntactic structure of a sentence by different linguistic elements, which are called ‘illocutionary force indicators’ and ‘propositional content indicators’, respectively. According to this model, illocutionary force indicators can influence the expression of the intention to perform a certain illocutionary act. Van Valin and La Polla (1997: 51) also point out that, on a syntactic level, illocutionary force indicators are utterance modifiers, just like e.g. evidentials. Additionally, they stress the importance of distinguishing speech act types from sentence types, as it is possible for a language to have a single basic sentence type and for speech act distinctions to be signalled either by prosody alone or by the addition of sentence-final particles (Van Valin and La Polla 1997: 42). If we accept the idea that the illocutionary force is a part of the grammatical meaning, then the intended force of an utterance can be contextually strengthened or decreased by the (optional) use of illocutionary force indicators. Accordingly, sentence-final *o* in finite-clauses can either reduce the illocutionary force of directive speech acts by adding a polite value to imperatives / prohibitives / hortatives / jussives or increase the illocutionary force of assertive

speech acts by insisting on the speaker's commitment to the truth of the proposition (cf. §3.1). The same analysis fits with *o* in final position of non-finite (i.e., Unipartite) clauses in which it can add nuance to the illocutionary force of vocatives and other addressing/courtesy interjections, while intensifying the assertive value of answers to Yes/No questions (cf. §3.2). All things considered, the main function of sentence-final *o* in Naijá is to show how a given propositional content is to be taken, that is, what illocutionary force the speaker applies to directive and assertive speech acts. This pragmatic interpretation is also supported by our statistical analysis of the occurrence of *o* within the NaijaSynCor that shows that sentence-final *o* is typical of dialogic contexts such as (formal and informal) conversations and drama texts as opposed to monologic texts.

As a matter of fact, the illocutionary function of *o* is not an innovative feature of Naijá insofar as many Niger-Congo use sentence-final *o* for “*emphasizing or, alternately, attenuating an assertion*” (Singler 1988: 130). Indeed, despite the fact that Singler categorizes *o* as a pragmatic marker of ‘currently relevant state’ (§1) based on the fact that it can either “correct a wrong assumption” (cf. also Migge 2021) or “determines what happens next”, many of his examples (Singler 1988: 124-125) clearly show its original illocutionary function related to speaker-oriented and epistemic modalities (cf. also ex. 1-2).

- “In Kroumen (Kru) *o* marks 'une sorte d'atténuation (que l'on trouve aussi après un ordre, un conseil, un enseignement)', ex. *nà wió o!* *your greeting* ‘Hello!’ (Thalmann m.s.: 54)”
- “In Nkonya (Kwa) *o* 'follows an expression containing a salutation or follows an exclamation expressed to a person'. *Ne mô to bo okulu a, santyi ba o!* ‘If it is deep inside, come back!’ (Reineke 1972: 126, 100)”
- “In Wobe (Kru) '*o* est un élément atténuateur. '*mɔ-je aun o. moi qui-voir te*(FEM) ‘Bonjour!’ [lit. ‘C'est moi qui t'ai vu.’] (Egner 1983:232)”
- “In Twi (Kwa) /oo/ is an emphasis marker.' *Kasa berɛɛ-oo.* talk softly ‘Talk slowly.’ (Redden et al. 1963:34)”
- “In Mende (Mande) '. . . in commands, calls, *o* "softens" the command or call.' *Wa ó!* ‘Come!’ (Innes 1969:121)”
- Kpelle (Mande) *o* is a 'particule exclamative'. *N'ga li tay o.* ‘Je pars au village.’ (Casthelain 1952:245,101)

This brings us to the second question to be addressed here, namely how can we analyse the ongoing emergence of new grammatical functions of *o* in both sentence-final and intra-sentential positions in the light of its basic directive (i.e., speaker-oriented) and assertive (i.e., epistemic) meanings? In this regard, it is interesting to note that in his grammar of Pichi, Yakpo (2020: 460) claims that “*it is this function of alerting (i.e., vocative) which lies at the heart of the other uses (of ó)*” (cf. §1). This view presumes a diachronic change from pragmatic-interactional (i.e., conative or phatic) to speaker-oriented / epistemic meanings. Our analysis rather shows that in Naijá, being synchronically dominant, it is the assertive meaning of *o* which provides the basis for its ongoing grammatical expansion. On the one hand, due to its large use as a marker of epistemic modality, sentence-final *o* can also be considered as an evidentiality strategy when it specifies the source of information of recent visual and non-visual events (ex. 27-28). This is because the evidential extension of the epistemic categories actually reflects a common cross-linguistic semantic change (cf. Aikhenvald 2004, §3.1). On the other hand, the assertive use of *o* is directly linked to the emergence of new coordinating and subordinating functions of *o* in intra-sentential position (§3.2). Concerning the optional use of *o* in topic structures, we have seen that intra-sentential *o* predominately marks frame-setting topics (ex. 30-31), while being rarer in argument topics (ex. 32, see also below). If we consider that in

frame-setting structures, the assertion bears on the predication inside the comment, the left dislocated being a merely stated adverbial frame, then intra-sentential *o* does not function as a true topicalizer (i.e., it has no scope on the left-dislocated topic), as it rather insists on the assertive nature of the following comment.

At this point, it should also be remarked that, if we exclude topic-comment structures, intra-sentential *o* always combines with other function words encoding either coordinating or subordinating syntactic relations. We have seen that intra-sentential *o* occurs either with *if* in truth conditionals (ex. 33-34) or with *wheda* in concessive conditionals (ex. 34-35). As already noted (§3.2), Haiman (1978: 587) argues that conditionals and topics have the same distribution at the initial of sentence because they both provide the information framework for the following discourse. In contrast to this hypothesis, Caron (2006) stresses that adverbial adjuncts (i.e., frame setting topics) and conditional clauses have to be distinguished from argument topics (i.e., topics proper). More specifically, Caron claims that, even if it is true that both frames and argument topics function as ‘preamble’ for the following predication, topic is a referential construct that will become an argument of the comment, whereas frame is the area in which the following comment holds true. In this perspective, adverbial adjuncts and conditionals are different types of frames (Caron 2006: 75) and this is most plausibly the reason why inter-clausal *o* in Najjá only rarely marks argument topics, while being more frequent in frame-setting and conditional clauses.

The use of intra-sentential *o* in concessive conditionals also explains its occurrence in contrastive adversative coordinated clauses introduced by *but sha (o)* (ex. 37-38). In fact, according to Malchukov (2004: 179), there is a common semantic overlap between concessive conditionals and contrastive adversative coordinated clauses, because they both share the function of denial of an expectation (Malchukov 2004: 179). Malchukov (2004: 181) further argues that “*an adversative connective is like a concessive in that it characterizes two situations as incompatible, but it is less committal than the concessive with regard to the nature of this incompatibility*”. More generally, the adversative meaning supposedly precedes the concessive meaning. However, due to the paucity of occurrences of intra-sentential *o* in these constructions (12 concessive occurrences vs 7 adversative occurrences), it is difficult to confirm the directionality of this semantic change based on synchronic corpus evidence. The point remains that both concessive and the adversative constructions are related to assertion as far as they both contains factual statements (cf. 3.2).

Lastly, concerning the use of intra-sentential *o* in subordinate complement clauses, it should be recalled that modal meanings and the dependent status of certain types of clauses are related phenomena, since they are both related to (non-)assertiveness (Cristofaro 2005; Nordström 2010). In this context, Frajzyngier (2005) observes that general subordinators are lexically separated modality markers that happen to occur in a specific syntactic position. As such, “*general subordinators are just one of the means by which languages encode modality*” (Frajzyngier 2005: 475). In view of that, it is not difficult to link the epistemic (i.e., assertive) meaning of *o* to its use in perception subordinate clause in which it co-occurs with the subordinator *sey* “that” (ex. 39). This subordinating function of intra-sentential *o* with perception verbs (e.g., “I hear that”) might equally been fostered by its emerging epistemic value (i.e., marker of the source of information) in sentence-final position (see above). In the case of modal subordinate clauses (e.g., “I’m able to”), the intra-sentential *o* instead precedes the English-derived subordinator *to*. Here, the subordinate clause is asserted relative to the epistemic modality introduced by *o* (ex. 21). In this case too, the subordinating function of intra-sentential *o* is congruent with its assertive value in sentence-final position. All things considered, the combination of inter-sentential *o* in phrasal compounds such as *but sha (o)*, *if... (o)*, *wheda... (o)*, *(o) sey*, and *(o) to* gives evidence of the ongoing semantic expansion from its core modal (i.e., assertive) to new coordinating/subordinating meanings. However, it is also

important to remark that the aforementioned phrasal compounds are not yet conventionalized, insofar as inter-sentential *o* still represents a minor syntactic option. Still, it is worth noting that *o* in Naijá presents a remarkably higher degree of multi-functionality when compared with its cognates in other Atlantic creoles (e.g., Pichi and Nengee; Kofi Yakpo and Bettina Migge, p.c.).

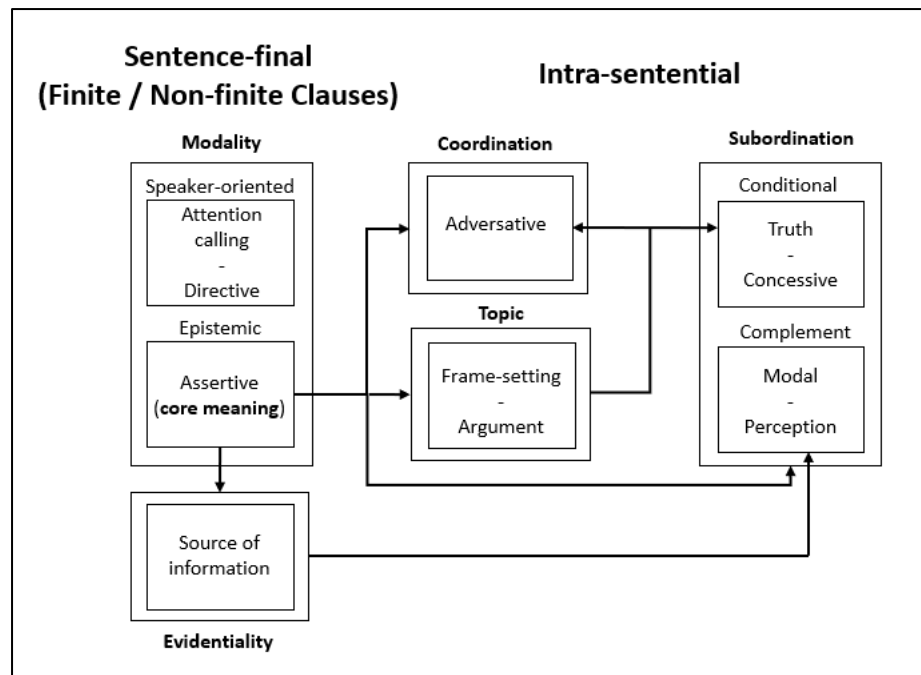


Figure 6. A corpus-driven semantic map of *O* in Naijá

In lieu of a conclusion, we present here a corpus-driven semantic map in Figure 6 that summarizes the grammatical meanings expressed by *o* in Naijá. The map represents cross-linguistically relevant synchronic connections (Haspelmath 2003) between the modal and the topicalizing and coordinating/subordinating functions *o*. The map organizes the different grammatical meanings according to their syntactic distribution (i.e., sentence-final vs intra-sentential). Intertwined together, the functions expressed by *o* in sentence-final (in both finite and non-finite clauses) and intra-sentential position eventually compose a harmonic semantic network explaining its ongoing grammatical expansion. The semantic map does not only allow us to draw a dynamic representation of the grammatical meanings of *o* in Naijá based on synchronic corpus evidence, but it also establishes a basis for its semantic comparison with similar polyfunctional networks in other Atlantic Creoles. In conclusion, the adoption of a corpus-driven methodology allowed us to synchronically describe sentence-final *o* as an ‘illocutionary force indicator’. This basic illocutionary function is confirmed by the fact, regardless of the formal/informal genre variable, sentence-final *o* most commonly occurs in dialogic texts. As a further matter, the occurrence of intra-sentential *o* in Naijá reveals an ongoing process of functional expansion that reflects a number of cross-linguistically common semantic changes from modal to subordinating meanings. Overall, the analysis of the recurrent patterns and the frequency distributions of *o* in the NaijaSynCor turns out to be useful not only for deriving the grammatical value of this multifunctional word in Naijá but also for identifying and comparing ongoing cross-linguistically relevant semantic changes.

## Bibliography

Aikhenvald, Alexandra, 2004. *Evidentiality*. Oxford: Oxford University Press.

- Austin, John Langshaw, 1962. *How to Do Things with Words*. Second edition (2005). Cambridge: Harvard University Press
- Bybee, Joan L., Revere Perkins, and William Pagliuca, 1994. *The Evolution of Grammar Tense, Aspect, and Modality in the Languages of the World*. Chicago/London: Chicago University Press.
- Bybee, Joan L. and Suzanne Fleischman, 1995. *Modality in Discourse*. Amsterdam: John Benjamins.
- Caron, Bernard, 2006. Condition, topic and focus in African languages: why conditionals are not topics. *ZAS Papers in Linguistics* (46): 69-82.
- Caron, Bernard, 2019. Clefts in Naijá, a Nigerian pidgincreole. *Linguistic Discovery*, 17 (1): 149-174.
- Caron, Bernard, Marine Courtin, Kim Gerdes and Sylvain Kahane, 2019. A Surface-Syntactic UD Treebank for Naijá. TLT 2019, Treebanks and Linguistic Theories, Syntax fest, Aug 2019, Paris, France. Accessed on <https://hal.archives-ouvertes.fr/hal-02270530>.
- Chafe, Wallace, 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Charles Li and Sandra Thompson (eds.), *Subject and Topic*. London – New York: Academic Press, 25-56.
- Cornillie, Bert, 2009. Evidentiality and epistemic modality: on the close relationship between two different categories. *Functions of Language* 16 (1): 44-62.
- Courtin, Marine, Bernard Caron, Kim Gerdes and Sylvain Kahane, 2018. Establishing a language by Annotating a Corpus: The Case of Naijá. In Sandra Kübler and Heike Zinsmeister (eds.), *Proceedings of annDH 2018, Annotation in Digital Humanities*. Sofia: 7-11.
- Cristofaro, Sonia, 2003. *Subordination*. Oxford: Oxford University Press.
- De Haan, Ferdinand. 1999. Evidentiality and Epistemic Modality: Setting Boundaries. *Southwest Journal of Linguistics* 18: 83-101.
- Deuber, Dagmar, 2005. *Nigerian Pidgin in Lagos: Language Contact, Variation and Change in an African Urban Setting*. London: Battlebridge.
- Faraclas, Nicholas G., 1996. *Nigerian Pidgin*. London: Routledge.
- Frajzyngier, Zygmunt 1995. A functional theory of complementizers. In J. Bybee & S. Fleischman (eds.), *Modality in Grammar and Discourse*. Amsterdam: John Benjamins. 473-502.
- Gerdes, Kim, Bruno Guillaume, Sylvain Kahane and Guy Perrier. 2018. SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD. *Universal Dependencies Workshop 2018*, Nov 2018, Brussels, Belgium.
- Gonzales, Montserrat et al. 2017. Epistemic and evidential marking in discourse: effects of register and detability. *Lingua*, 186-187: 68-87.
- Haiman, John, 1978. Conditionals are topics. *Language* 54 (3): 564-589.

- Haspelmath, Martin, 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In M. Tomasello (ed.), *The New Psychology of Language, Vol. 2*. Lawrence Erlbaum, 211-242.
- Honkanen, Mirka, forthcoming. “This word no get concrete meaning oo”: Pragmatic markers in Nigerian online communication. In In S. Mohr, J. A. Anderson, & K. Schneider (Eds.), *Communicative action and interaction in Africa*. Pragmatics & Beyond New Series. Amsterdam: John Benjamins.
- Izre’el, Shlomo, 2018. Unipartite clauses. A view from spoken Israeli Hebrew. In M. Tosco (ed.), *Afroasiatic: Data and Perspective*. Amsterdam: John Benjamins, 235-260.
- Karttunen, Lauri, 1971. Implicative verbs. *Language* 47 (2) : 340-358.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý and Vít Suchomel. 2014. The Sketch Engine: ten years on. *Lexicography* 1(1): 7–36.
- Malchukov, Andrej L. 2004. Towards a semantic typology of adversative and contrast marking. *Journal of Semantics* 21: 177-198.
- Manfredi, Stefano, Slavomír Čéplö, Francis Egbokhare, and Christine Ofulue, 2019. A first variationist approach to the NaijaSynCor corpus. Paper presented at the *International Naijá Symposium*. University of Ibadan, Nigeria, 27-29 June 2019.
- Mauri Caterina and Andrea Sansò, 2011. How directive constructions emerge: Grammaticalization, constructionalization, cooptation. *Journal of Pragmatics* 43 (14): 3489-3521
- Mazzoli, Maria, 2013. Copulas in Nigerian Pidgin. Ph. D. Dissertation, University of Padova.
- Migge, Bettina, 2020. Broadening Creole Studies. From grammar to discourse. *Journal of Pidgin and Creole Studies* 35 (1): 160-177.
- Nordström, Jackie, 2010. Modality and Subordinators. Amsterdam: John Benjamins.
- Pietrandrea, Paola and Sylvain Kahane. 2019. Macrosyntactic annotation. In Lacheret-Dujour, A., Kahane, S. and Pietrandrea, P. (ed.), *Rhapsodie: A prosodic and syntactic treebank for spoken French*. Amsterdam: John Benjamins. 97-125
- Petrov, Slav, Dipanjan Das and Ryan McDonald. 2012. A Universal Part-of-Speech Tagset. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12), 2089–2096. Istanbul, Turkey: European Language Resources Association (ELRA).
- Rychlý, Pavel. 2007. Manatee/Bonito-A Modular Corpus Manager. In RASLAN, 65–70.
- Searle, John R. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge, Cambridge University Press.
- Searle, John R. 1975. A Taxonomy of Illocutionary Acts. In Gunderson, K. (ed.), *Language, Mind, and Knowledge*, Minneapolis, vol. 7.
- Searle, John R. and Daniel Vanderveken, 1985. Speech Acts and Illocutionary Logic. In Daniel Vanderveken (eds) *Logic, Thought and Action*. Logic. Dordrecht: Springer. 109-134.
- Singler, Victor, 1988. The story of *o*. *Studies in Languages* 12: 123-144.
- Tognini-Bonelli, Elena, 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.



Unuabonah, Foluke and Oladipupo, Rotimi, 2018. “You’re not staying in Island *sha o*”: *O, sha* and *abi* as pragmatic markers in Nigerian English. *Journal of Pragmatics* 135, 8-23.

Van Valin, Robert D. Jr. and Randy J. LaPolla, 1997. *Syntax. Structure, Meaning and Function*. Cambridge: Cambridge University Press.

Yakpo, Kofi, 2019. *A Grammar of Pichi*. Berlin: Language Science Press.

Zeman, Daniel, et al., 2022, *Universal Dependencies 2.10*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11234/1-4758>.

### **Author 1**

Stefano Manfredi

SeDyL (CNRS UMR8202, INALCO, IRD UR135)

Campus CNRS de Villejuif, Bâtiment D

7, rue Guy-Môquet, 94801 Villejuif Cedex (France)

[stefano.manfredi@cnrs.fr](mailto:stefano.manfredi@cnrs.fr)

### **Author 2**

Slavomír Čéplö

CERES, Ruhr-Universität Bochum

Universitätsstr. 90a

44789 Bochum, Germany

[bulbul@bulbul.sk](mailto:bulbul@bulbul.sk)