

Using an online tool for the documentation of Èdó language

Ota Ogie

Norwegian University of Science and Technology (NTNU)

ota.ogie@hf.ntnu.no

Abstract¹

Language documentation is important as a tool for preservation of endangered languages and making data available to speakers and researchers of a language. A data base such as TypeCraft is important for typology studies both for well documented languages as well as little documented languages and is a valid tool for comparison of languages. This requires that linguistic elements must be coded in a manner that allows comparability across widely varying language data. In this paper, I discuss how I have used the coding system in TypeCraft for the documentation of data from Èdó language, a language belonging to the Edoid group of the Benue-Congo subfamily of the Volta-Congo language family and spoken in Mid-Western Nigeria, West Africa. The study shows how syntactic, semantic and morphological properties of multi-verb constructions in Èdó (Benue-Congo) can be represented in a relational database.

1. Introduction

In this paper², I show some ways in which I am using a shared methodology in my research on multi-verb constructions. My research is centered around the language Èdó, spoken in Mid-Western Nigeria, Ga and Akan (kwa), and the tool is the system TypeCraft, which has been developed in the ISK department, NTNU and first documented in Beermann and Prange (2006).

Èdó language belongs to the Niger-Congo, Atlantic-Congo, Volta-Congo, Benue-Congo-Edoid language family. The Ediod language family

consists of 33 languages and 19 of these languages have either very little documentation or no documentation available.

Multi-verb constructions are constructions in which the verbs in series must function as independent verbs in simple constructions, with at least one shared argument and no marking of syntactic dependency.

The paper shows how syntactic, semantic and morphological properties of multi-verb constructions in Èdó (Benue-Congo) can be represented in a relational database and the development of annotation standards that contribute to contrastive and typological research. The analysis is extended to multi-verb constructions in the following languages of the Niger-Congo: Ga and Akan (Kwa).

2. TypeCraft

TypeCraft is a tool for typological analysis that allows for annotation, classification and search of data along different morphological, syntactic, semantic and pragmatic criteria. In annotating it is important to have annotation schemes that allow for typological and contrastive studies.

In this paper I use an annotation scheme for verbal constructions currently being developed at NTNU and documented in Hellan and Dakubu (2009). Syntactic and semantic information about construction types are provided by templates composed by labels. The basic structural parts of a template are referred to as **slots** that are separated by hyphens. A template with a verbal head can consist of maximal 7 slots; (1) POS of the head, and diathesis information; (2) valence specification; (3) dependent specification; (4) participant roles; (5) aspect and aktionsart; (6) situation type; (7) provides a linking type between slot 6 situation type and the specifications in slots 2-4. Slots 1 and 2 are obligatorily filled, the others not. (cf. Hellan and Dakubu, 2009). At present annotation of the construction labels is manual and not incorporated into the TypeCraft. However TypeCraft provides a

¹ In *Proceedings of the Fourth Linguistic Annotation Workshop*. Held in conjunction with the 48th Annual Meeting of the ACL-2010. Sweden, July 11–16, 2010: 109–112 ©2010 Association for Computational Linguistics.
<http://www.aclweb.org/anthology/W10-1813>

² I thank Professor Lars Hellan, NTNU Norway for his comments on earlier versions of this paper.

construction tier where this information can be incorporated.

3. Sentence level and word level annotation

TypeCraft provides a set of glosses for syntactic and semantic coding and a set of parameters along which sentences may be classified that allow for standardized annotation and cross linguistic comparison as illustrated in figure1:

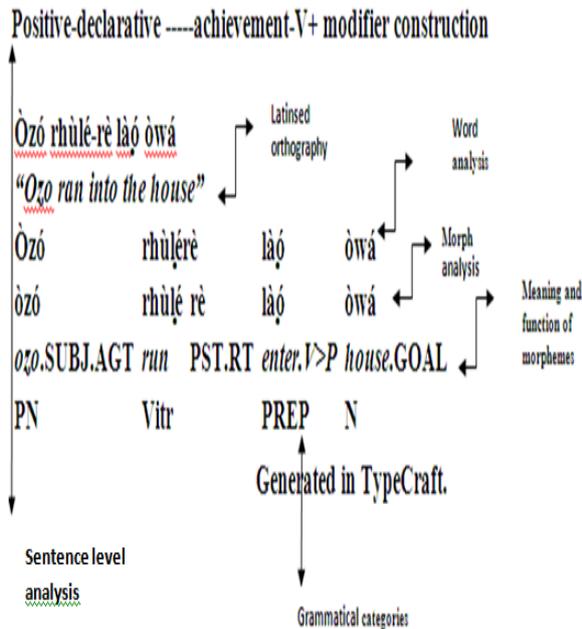


figure1: Word and sentence level annotation

3.1 Word level

Word level annotation allows for analysis of predicates in terms of syntactic and semantic properties including information about the subcategorization properties and argument structure of predicates.

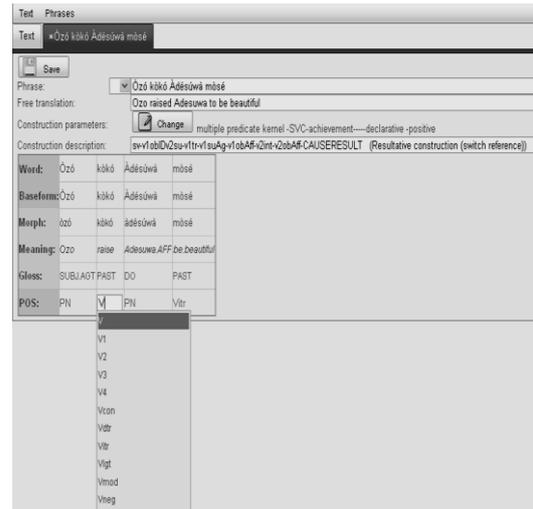


Figure2: Text editor in TypeCraft showing word level annotation

Type craft features 7 tiers that provide information at the word level as shown in the Èdó example below.

(1). Èdó

Construction parameters: TransitiveVerb-
accomplishment-----declarative -positive
Construction labels: v-tr-suAg_obThincrem-
COMPLETED_MONODEVMT

“He/she wrote books”

èbé

èbé

3SG.SUBJ.NOM.AG write PL.PST. book.DO.T
T H H
PRON V CN

Generated in TypeCraft

The construction labels are explained as follows: v in Slot1 in the example above states that the head of the construction is a verb. tr in Slot2 states that the verb is transitive, suAg_obThincrem in Slot 4 states that the NP that is the subject of the construction bears an agent theta role and the object an incremental theme theta role. Lastly slot 5 gives the information that the aktionsart of the construction is completed monodevelopment.

3.2 Sentence level

TypeCraft provides a set of global tags at the sentence level that allows for classification in terms of syntactic and semantic automatically generated construction parameters such as constituent type, core constituent vs adjunct, transitivity, thematic roles, situation and aspect types, propositional types and polarity. Polarity is based on the assumption that States Of Affairs (SOA) comes in pairs: positive and negative. Figure 3 is used as illustration:

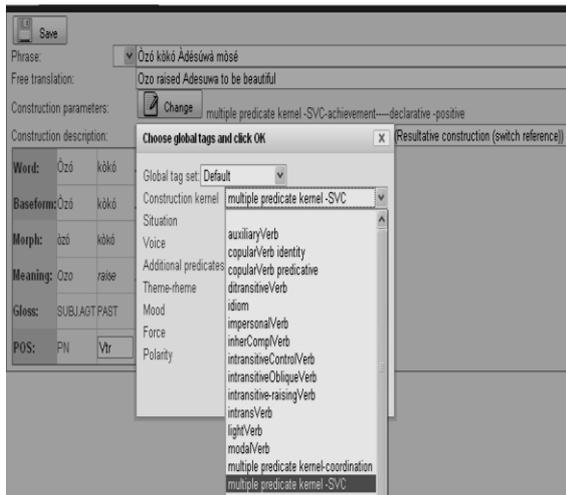


Figure3: Text editor in TypeCraft showing sentence level annotation

(2). Èdó

Construction parameters: multiple predicate kernel -SVC-achievement-----declarative -positive
 Construction labels: svSuObIDALLsuAgobAff-v1tr-v2tr-EVENTSEQ

“Ozo cooked rice and sold”

Òzó	lé		
òzó	lé		
Ozo.SUBJ.AG	cook.PST.	rice.AFF.D	sell.PST.
T	H	O	H
PN	V	N	V

Generated in TypeCraft.

The construction parameter is explained as follows: the global tags *multiple predicate kernel -SVC-* provides information about constituent type, *achievement* provides information about situation and aspect types, *declarative* provides information about propositional types and *positive* about polarity.

The construction labels have the following structure: Area1 (in italics for ease of exposition) gives the global labels, the number of verbs in series (*ie sv, sv3, sv4*) as well as argument sharing information (coded by the label *IDALL*) and information about thematic relations holding across the verb in series. Area 2 gives the valence information as well as information about grammatical function and thematic roles (underlined for ease of exposition). Information about the situation type of the construction is provided by Area 3 and is written in capital letters.

Information about tense, aspect, mood and negation is also provided by area 1 in the construction labels. Sharing of these features across verbs in series is represented as with sharing of arguments as in example (3) from Akan below.

(3). Akan

Construction parameters: multiple predicate kernel -SVC-achievement-----declarative -positive
 Construction labels: svsuAspIDALLsuAgaspCompl-v1tr-v1obAff-v2intr- CAUSERESULT

Ama twe-e Kofi hwe-e fam

“Ama pulled Kofi and fell (Ama fell) (covert reference subject sharing)”

Ama twee Kofi hwee fam

ama tw e kofi hwe e fam

Ama.SUBJ. pul COM kofi.AFF. fall.COM und
 AGT l PL DO PL er
 PN Vtr PN Vitr

Generated in TypeCraft.

With respect to the global labels in area 1, Hellan and Dakubu (2009) uses also the global label *ev* to represent Extended Verb Complexes and the label *pv* for preverbs in EVCs. In addition, to the labeling conventions used by Hellan and Dakubu (2009) for SVCs (*sv*) and EVCs (*ev, pv*), the following global labels are introduced to account for the range of multi-verb constructions in my data.

- cc – covert co-ordination
- mvc- multi-verb construction
- mc- modifier construction

4. Text , phrasal and construction search

TypeCraft allows for search using different word level and sentence level parameters. This facilitates comparative analysis in multi-verb constructions. For example, argument sharing is a property that

identifies types of multi-verb constructions. A search using the construction label *svsuObIDALL* is used as illustration. The result gives an output of serial verb constructions in Èdó and Ga consisting of two verbs in series with the subject and object arguments of the verbs in series sharing reference:

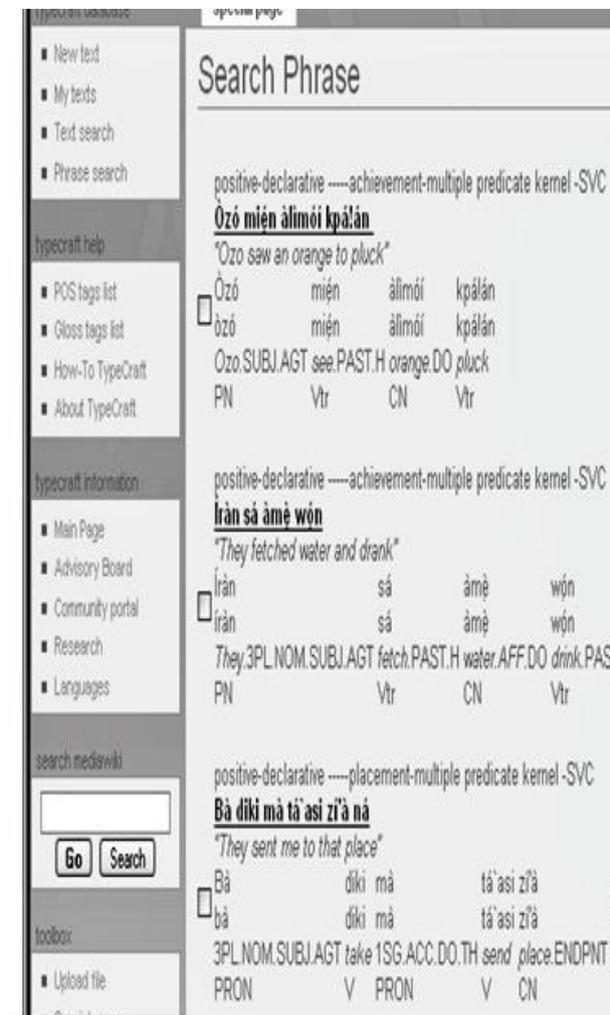


Figure 4: Search for phrase using global tag *svsuObIDALL*

The standardized annotation, search parameters and online nature of TypeCraft makes it advantageous compared to toolbox, a linguistic data management file based system used by many linguists in the documentation of African languages.

5. Conclusion

Standardized annotations and online databases such as TypeCraft aid linguists and speakers of a language in research, preservation of languages and in producing literacy materials that aid education and literacy. My research on multi-verb constructions in Èdó is the first in-depth annotation for Èdó and will be easily available for language researchers/teachers/students all over the world.

6. References

- Beermann, Dorothee and Atle Prange. 2006. "Annotating and archiving Natural Language Paradigms online", presentation at the *Texas Linguistic Society Annual Conference*. Austin, Texas. 2006.
- Hellan, Lars and Dakubu, Mary Esther Kropp. 2009. Identifying verb constructions cross-linguistically, *Studies in the Languages of the Volta Basin 6. Part 3*. Accra Ghana University Press 2009.
- Ogie, Ota 2009. Multi-verb constructions in Edo. PhD dissertation. Norwegian University of Science and Technology. Trondheim. Norway. Also to appear in VDM publishers. Germany.
- Ogie, Ota . 2009. In-depth annotation of multi-verb constructions in Èdó. <http://www.typecraft.org.NTNU>.
- Ogie, Ota . 2009. Multi-verb constructions in Edo. <http://www.typecraft.org.NTNU>.
- Ogie, Ota 2009. An annotation tool for less-resourced languages. Presented at The Nordic African Days 2009. Africa- in search for alternatives. Trondheim 1-3 October 2009. The Nordic African Institute.
- Ogie, Ota & Dorothee Beermann 2009. Typecraft Software for the documentation of minority languages. Presented at ISFITovation February 26th 2009. Slides 29- 44. <http://itovation.wordpress.com/2009/02/26/watch-isfitovation-live/>. NTNU, Trondheim, Norway