

A computational grammar of Ga

Lars Hellan

Norwegian University of Science and Technology, Norway
lars.hellan@ntnu.no

Abstract

The paper describes aspects of an HPSG style computational grammar of the West African language Ga (a Kwa language spoken in the Accra area of Ghana). As a Volta Basin Kwa language, Ga features many types of multiverb expressions and other particular constructional patterns in the verbal and nominal domain. The paper highlights theoretical and formal features of the grammar motivated by these phenomena, some of them possibly innovative to the formal framework. As a so-called deep grammar of the language, it hosts a rich lexical structure, and we describe ways in which the grammar builds on previously available lexical resources. We outline an environment of current resources in which the grammar is part, and lines of research and development in which it and its environment can be used.

Keywords: Ga, Kwa, computational grammar, typed feature structures, construction level compositional analysis, nominal structures, multiverb constructions

1. Introduction¹

Ga is a Kwa language spoken in the Accra area of Ghana with about 745,000 speakers². Linguistic descriptions date back to early 1800 (cf. Rask (1828)), and although digital text resources are few, it is well studied linguistically, and has some advanced resources such as the Ga-English dictionary (Dakubu 2009). The present article describes digital resources which in important respects derive from this dictionary and its underlying Toolbox lexicon. We mainly focus on a computational grammar of Ga, whose development started in 2005, and also on a valence lexicon, whose development started in 2008. The developments were coordinated, although each at its own pace, and conducted jointly by Professor Mary Esther Kropp Dakubu and the author until Prof. Dakubu's death in 2016.

Computational grammars are programs which automatically assign various types of analysis to sentences of a language – they range from morphological parsers, which recognize words' part of speech (POS) and morphological build-up, via dependency parsers which recognize syntactic phrases and dependency relations between words internal to a phrase and between phrases, to so-called 'deep' parsers which also recognize lexical structures and semantic properties of words and their combinations. Deep parsers reflect frameworks of formal grammar such as Lexical Functional Grammar (abbreviated 'LFG', cf. Bresnan (2001)) and Head-Driven Phrase structure grammar (abbreviated 'HPSG', cf. Pollard and Sag (1994), Copestake (2002)); the grammar to be presented mainly follows HPSG but with some elements of LFG; it technically is developed at the LKB platform described in Copestake (2002).

Verb valence lexicons are lexicons giving concise enumerations of the valence frames of each verb, i.e., enumerations of the possible environments of a verb described in terms of the so-called *valence-bound* items in the environments (following the terminology of Tesnière (1959)). A principled meeting point between valence lexicons and deep grammars is that the verb lexicon of a

deep grammar will have explicit valence information. From either side one can thereby derive the other (and even in turn perform cyclical improvements, taking advantage of articulations on the derived side proving useful also on the other side, and vice versa). In the present case, once the valence lexicon was established, it was imported into the grammar.

Both types of resources have a solid foundation in Indo-European languages, and one can name various kinds of practical applications that they serve. However, equally interesting is what these resources for Ga can tell us regarding what are basic and necessary structures of grammar and valence. For instance, in an HPSG based grammar, the distinction between *argument* (i.e., valence-bound) and *adjunct* (i.e., not valence-bound) is basic like in linguistic traditions in Indo-European languages, and a question is whether it can be maintained in a grammar of a Kwa language. Likewise, the grammatical articulation of some semantic structures is quite different in Ga from what one expects in Indo-European languages.

To be more concrete, Kwa languages like Ga and Akan are known to make little use of prepositions and adjectives, so that constructions involving nouns and verbs may be seen as playing a larger role than, e.g., in Indo-European languages. Thus *multiverb expressions* are known to play a large role in the languages, subsuming *Serial Verb Constructions (SVCs)*, *Extended Verb Complexes (EVCs)* which are sequences of preverbs preceding a main verb, and *Verbid Constructions (ViD)*, where verb phrases play a role similar to what adverbials play in Indo-European languages (see Dakubu 2004a, 2008, Dakubu et al. 2007, Dakubu 2013 for analysis of many of the construction types). Such constructions raise the question whether there can be more than one verbal head per sentence; and if not, whether the argument-adjunct distinction is at all relevant to describing the relationships between the verbs. A further reflex of the lack of prepositions is that spatial specification often take the shape of transitive constructions. Moreover, pre-nominal specifiers manifest a complexity well beyond what one finds in Indo-European languages. The latter construction types will be exemplified and analyzed in section 2. In section 3 we exemplify and show the analysis of multiverb constructions. Section 4 recapitulates the development of the valence lexicon from the Toolbox

¹ I am grateful to the three reviewers for helpful comments.

² ISO-639-3 «gaa». Number of speakers in 2013.

<https://www.ethnologue.com/country/GH/languages>

source, and accompanying resources. Section 5 discusses possible further developments of the resources described. Examples throughout are from the works by Dakubu cited above, from the ‘Ga Appendix’ to Hellan and Dakubu 2010, and from Dakubu (Unpublished a). The latter is the presentation of the valence lexicon that we will be referring to, with about 2000 entries, where each entry represents *one* frame of a given verb. Thus, when a verb has *n* frames, it will be represented in *n* entries. To each entry is provided a short example, whereby this is also a corpus of short sentences.

2. Nominal complexes with relational nouns and possessive constructions

2.1 Examples

Nominal complexes with relational nouns and possessive constructions are exemplified in (1):

(1)
a.
v Ee-la e-daa-ŋ
3S.PROG-sing 3S.POSS-mouth-LOC"
V N
"He's murmuring incoherently to himself."
(literally: 'he is singing his mouth')

b.
E-ŋmra e-toi-ŋ
3S.AOR-scrape 3S.POSS-ear-LOC
V N
"She slapped him."
(literally: 'she scraped his ear')

c.
E-tsuinaa mii-funta lɛ
3S.POSS-desire PROG-nauseate 3S
N V PN
"She feels sick, nauseous."
(literally: 'her desire nauseates her')

d.
Mi-yitso mii-gba mi
1S.POSS-head PROG-split 1S
N V PN
"My head is aching."
(literally: 'my head splits me')

e.
O-he ɔ-ɔ bo
2S.POSS-self cool-HAB 2S
N V PN
"you are at ease."
(literally: 'your self cools you')

In each sentence, the full NP is headed by a relational noun which has a possessive specifier, and this specifier is coreferential with a pronoun (as prefix or freestanding). Of the 2000 sentences in the corpus mentioned, no less than 690 have an object headed by a relational noun, and 100 have a subject headed by a relational noun, often with

a bodypart or identity reading. This attests to the importance of analytically representing nominal complexes with relational nouns and possessive constructions.

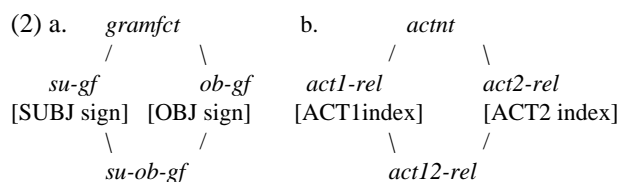
2.2 Analysis

A first installment of the grammar follows the HPSG Matrix (Bender et al. 2010), illustrated in Dakubu et al. 2007, while in a more recent version the grammar is designed according to the architecture outlined in Hellan 2019; both use the LKB platform, whose formalism is a Typed Feature Structure (TFS) system. Information in such a system is generally exposed through Attribute Value Matrices (AVMs), where each AVM belongs to a *type*, and attributes are introduced (declared) according to the following conventions:

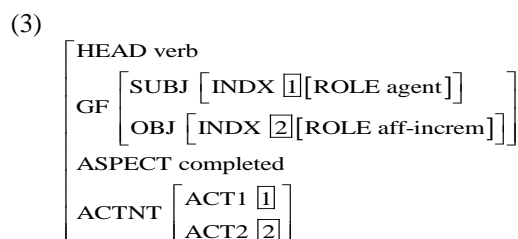
[A] A given type introduces the same attribute(s) no matter in which environment it is used.

[B] A given attribute is declared by one type only (but occurs with all of its subtypes).

In a TFS representing a grammar, there are many type hierarchies, representing POS, tenses, semantic roles, etc.; some of these hierarchies do without attributes, while the following ones do. Types for grammatical functions (values of the attribute ‘GF’) and actants (participants in semantic argument structure, represented as values of the attribute ‘ACTNT’) include those indicated below: the *gramfct* subtypes declare the GF attributes (‘SUBJ’ and ‘OBJ’) and the *actnt* subtypes declare the semantic argument structure attributes (‘ACT1’ and ‘ACT2’):



The way in which these attributes combine in an AVM of a transitive structure, as in a sentence like *John ate the cake*, is illustrated in (3); the co-numbering ‘1’ and ‘2’ indicate that the referential index of the subject is the ACT(ant)1 and the referential index of the object is the ACT(ant)2:



While a structure like (3) will reflect constructional features of a sentence like *John ate the cake*, a representation of what it *means* will also reflect the content of the various words. A strategy of ‘first stepping stone semantics’ is to simply put in a representation of the word itself in a slot designated for semantic argument structure, which for the sentence in question will mean

extending (3) as (4) (modulo definiteness marking of the object):

$$(4) \left[\begin{array}{l} \text{HEAD verb} \\ \text{GF} \left[\begin{array}{l} \text{SUBJ} \left[\begin{array}{l} \text{INDX } \boxed{1} [\text{ROLE agent}] \\ \text{ACTNT } [\text{PRED } \textit{John}] \end{array} \right] \\ \text{OBJ} \left[\begin{array}{l} \text{INDX } \boxed{2} [\text{ROLE aff-increm}] \\ \text{ACTNT } [\text{PRED } \textit{cake - rel}] \end{array} \right] \end{array} \right] \\ \text{ASPECT completed} \\ \text{ACTNT} \left[\begin{array}{l} \text{PRED } \textit{eat - rel} \\ \text{ACT1 } \boxed{1} \\ \text{ACT2 } \boxed{2} \end{array} \right] \end{array} \right]$$

To obtain this, each word must be lexically specified for its semantic contribution, along with a recipe for how it fits in relative to the overall structure (4). The use of *valence lists* in HPSG serves such a purpose; for the case in point, *eat* will thereby have as its lexical specification a structure like (5), where the valence list attributes *SPR* and *COMPS* enumerate the items with which the word has to combine (where failure for appropriate items to obtain in the word string to be analyzed means that this lexical structure is not appropriate for the analysis process):³

$$(5) \left[\begin{array}{l} \text{HEAD verb} \\ \text{ORTH "eat"} \\ \text{GF} \left[\begin{array}{l} \text{SUBJ } \boxed{1} [\text{INDX } \boxed{3} [\text{ROLE agent}]] \\ \text{OBJ } \boxed{2} [\text{INDX } \boxed{4} [\text{ROLE aff-increm}]] \end{array} \right] \\ \text{SPR } \langle \boxed{1} \rangle \\ \text{COMPS } \langle \boxed{2} \rangle \\ \text{ACTNT} \left[\begin{array}{l} \text{PRED } \textit{eat - rel} \\ \text{ACT1 } \boxed{3} \\ \text{ACT2 } \boxed{4} \end{array} \right] \end{array} \right]$$

A parsing algorithm is in principle necessary if one wants to construe a grammar as *compositional*, since compositionality resides in combinatorial relations between constituents, meaning that a grammar as a whole is compositional if all phenomena to be covered by it can be construed exclusively in terms of combinatorial operations involving all parts of the sentences analysed.

It may be noted that standard HPSG in this respect is a parsing approach exclusively, in that it does not include GF notions, so that a parse representation will be without GF and with both valence lists empty, thus being less informative than (4). Adding such notions to the parsing formalism strengthens the formalism, which might be unwanted on other grounds, yielding a situation where one chooses between formats on other grounds than plainly empirical. However, as we turn to the selected areas of Ga grammar to be considered, we will see that even from parsing perspectives, there may be reasons to use GF in the formalism. In (1d), repeated:

(1d)
 Mi-yitso mii-gba mi
 1S.POSS-head PROG-split 1S
 N V PN
 "My head is aching."
 (literally: my head splits me')

we want to represent the subject as a possessive phrase, where the referent of the whole phrase is a (body)part of the specifier 'mi', and this specifier is also identical to the object; in terms of semantics. The first of these constellations we may represent as in (6), labeled as 'subject is a BodyPart of subject's specifier' (of course speaking of their referents), and the second as (7), in a similar vein labeled as 'subject's specifier is Identical to object':

$$(6) \left[\begin{array}{l} \text{suBPsuSpec} \left[\begin{array}{l} \text{GF} \left[\begin{array}{l} \text{SUBJ} \left[\begin{array}{l} \text{INDX } \boxed{1} \\ \text{GF} [\text{SPEC} [\text{INDX } \boxed{2}]] \end{array} \right] \\ \text{ACTNT} \left[\begin{array}{l} \text{PRED 'is-bodypart-of-rel'} \\ \text{ACT1 } \boxed{1} \\ \text{ACT2 } \boxed{2} \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right]$$

$$(7) \left[\begin{array}{l} \text{suSpecIDob} \left[\begin{array}{l} \text{GF} \left[\begin{array}{l} \text{SUBJ} \left[\text{GF} [\text{SPEC} [\text{INDX } \boxed{2}]] \right] \right] \\ \text{OBJ} [\text{INDX } \boxed{2}] \end{array} \right] \end{array} \right]$$

In a full representation of the sentence, (6) and (7) should unify with the verb representation as (8):

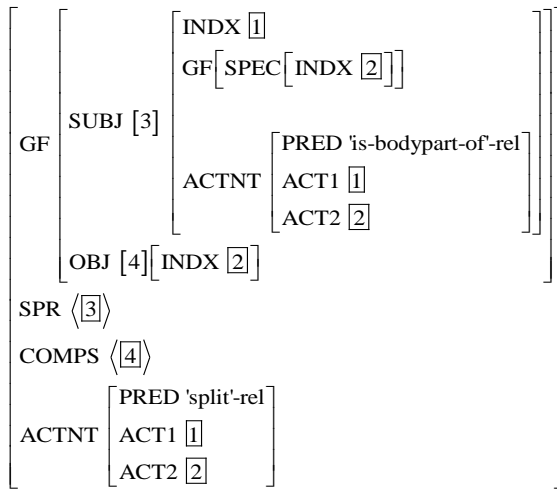
$$(8) \left[\begin{array}{l} \text{GF} \left[\begin{array}{l} \text{SUBJ} \left[\begin{array}{l} \text{INDX } \boxed{1} \\ \text{GF} [\text{SPEC} [\text{INDX } \boxed{2}]] \end{array} \right] \\ \text{ACTNT} \left[\begin{array}{l} \text{PRED 'is-bodypart-of-rel'} \\ \text{ACT1 } \boxed{1} \\ \text{ACT2 } \boxed{2} \end{array} \right] \end{array} \right] \\ \text{OBJ} [\text{INDX } \boxed{2}] \\ \text{ACTNT} \left[\begin{array}{l} \text{PRED 'gba'-rel'} \\ \text{ACT1 } \boxed{1} \\ \text{ACT2 } \boxed{2} \end{array} \right] \end{array} \right]$$

However, although *gba* 'split' is a transitive verb, defining it simply on the model of (5) will fail to induce the relations between the specifier of the subject and the subject and object. What is needed is a lexical representation able to 'look down' into the subject, thus 'seeing' the item that relative to a valence list of the head noun would be representable as the list '<[SPR]>'. This represents a pattern of 'non-locality' for which the valence list notation is not defined (i.e., meaning specifying a list inside of an item inside the 'SPR' list). A way in which we avoid violating this restriction is by, instead of an extra embedded list, using GF attributes reflecting the way they are used in (8), so that the lexical

³ See Hellan (2019a) for details on introduction of lexical types.

specification for *gba* relative to the kind of frame in question is (9):

(9) Lexical entry of *gba* in (1d):



With similar reasoning for the other cases in (1) and related constructions, this demonstrates the use of including GFs as a construct also in the parsing algorithm. Similar cases have not been prominent in the discussion of the design of standard version of HPSG, and so the phenomenon of Nominal complexes with relational nouns and possessive constructions may represent motivation for this item of modification of the general formal design, and thus a motivation coming from Ga. It has been adopted in the current grammar.

3. Multiverb expressions

Multiverb expressions types are well exemplified in the literature,⁴ so here we just point to two types and discuss some aspects of their analysis.

3.1 Serial Verb Construction (SVC)

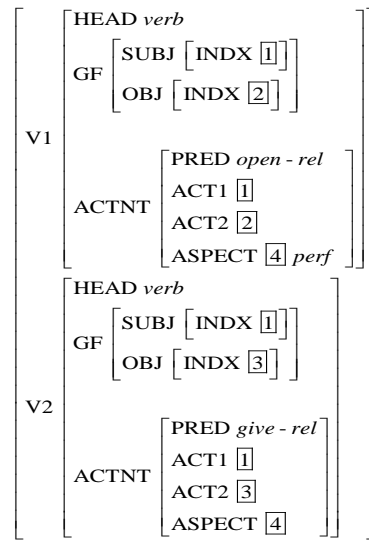
An SVC appears as a sequencing of any number of VPs, with pervasive uniformity between the verbs, both in their morphology and regarding their arguments. Interpretations range from temporal sequences of events reflecting the sequencing of VPs to pairwise more special combinations. (10) is an example of the latter (from Dakubu (unpublished a)):

| | | | | |
|------|-------------------------------------|------|------------|------|
| (10) | Á-gbele | gbe | á-ha | bo |
| | 3.PRF-open | road | 3.PRF-give | 2S |
| | V | N | V | Pron |
| | 'You have been granted permission.' | | | |

This SVC has two verbs, and as is often the case in Ga SVCs, both with the subject expressed by a clitic; the subjects are identical, and likewise the aspects of the verbs.

In AVM form, this can be provisionally exposed as follows, where the notions 'V1', 'V2' are standard labels for the VPs in an SVC sequence:

(11)



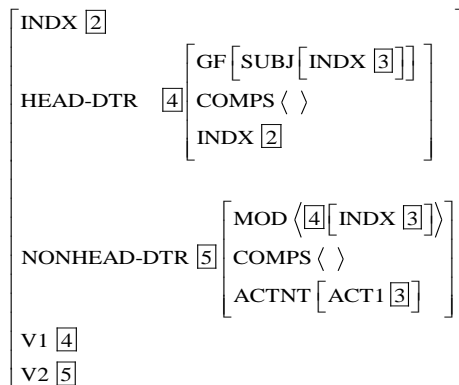
We now face the question whether the SVC should be counted as multi-headed, or whether there are linguistic reasons to count one VP as head and the other as something else. VPs in SVCs are generally too tightly integrated to count as coordination structures – cf. Hellan et al. (2003) for a discussion relative to temporally sequenced VPs in SVCs in Akan (commonly referred to as 'chaining SVCs), an argumentation which may well also hold for Ga, and especially for a case like this where the interpretation is not one of temporal sequencing. If not a coordination structure, another possibility may be a structure of complementation: although *gbele* ('open') does not have a meaning which would motivate counting the subsequent VP as a complement, one might perhaps count the whole construction as a phraseological unit and technically count the first verb as binding the second VP to it as a fixed part. The third option is an analysis of the sequencing of VPs is as *adjunction* between the VPs; this is in general plausible for cases of temporally sequenced VPs where any number of VPs can freely occur; structures like (10) could then be treated as a limiting case of such structures.

Given this as most plausible for the 'free' VP sequences, how can this be formally implemented, and how would the attributes V1 and V2 in case be introduced? Adjunction to, or modification of, VPs is commonly construed as the adjunct being a predicate of the event expressed by the head VP. This will be false for the VP sequencing, since the adjoined VP is predicated of the same entity as the head VP is predicated of. To express this, the grammar must contain, in addition to the 'event modification' rule, a modification rule imposing coreference between the subjects. (12) is such a rule, equating the ACT1 of the adjunct ('NON-HEAD-DTR') not to the event index of the head, but to the index of its

⁴ For a recent overview concerning Akan and Ga, see Beermann and Hellan (2018).

subject; we here also indicate the introduction of ‘V1’ and ‘V2’⁵.

(12) *Head-Modifier rule II* (partial formulation)



This rule schema will apply recursively when there are more than two VPs.

The motivated status of such a rule of modification is again a respect in which Ga and similar languages may be seen as adding a formal possibility to the formal inventories sustained so far.⁶

The by far most common pattern of ‘argument sharing’ in SVCs is one of identical subjects. In the literature also identity between objects has been recognized – as ‘object sharing’ – and even identity between the object of one VP and the subject of the following VP, called ‘switch sharing’. These are rare in Ga, but exemplified in Akan by examples like (13):⁷

(13) ‘switch sharing’ between object of one verb and subject of the subsequent verb (Akan)
 Kofi to-o ne nan wo-o Kwame
 Kofi throw-PRF 3Poss leg pierce-PRF Kwame
 N V Pron N V N
 ‘Kofi kicked Kwame’

Unlike the case above, where both COMPS lists are empty at the point where two VPs combine, one here needs to be able to state that the object of the first VP is identical to the ACT1 of the second VP. Such an identity is hard to state if all one can refer to in the first VP is an empty COMPS list. Once the object is also represented in the GF specification of the first VP, however, one has a

⁵ By convention, in the constellation (i), the path ‘V2.V2’ can be read as ‘V3’, and similarly for higher numbers of VPs.

(i)
$$\left[\begin{array}{l} \text{V1} \\ \text{V2} \left[\begin{array}{l} \text{V1} \\ \text{V2} \end{array} \right] \end{array} \right]$$

⁶ The schema will also be relevant in Indic and other languages with so-called co(n)-verb constructions, where the syntax by itself is marked as modification – and thus unlike the Volta Basin SVC pattern – whereas the argument identities which can obtain match the present case.

⁷ Presented by Clement Appah at a seminar in Legon, Ghana.

reference point.⁸ An example with object sharing in Ga is given in (16) below. Although the status of both of these patterns involving objects is a bit controversial, both construction types are accommodated in the grammar, and may potentially provide further motivation for including GF notions in the grammar formalism.

3.2 Extended Verb Complex (EVC) and more

For descriptions of this construction in Ga see Dakubu (2008), Dakubu et al. (2007), Dakubu (2004a), and in Dangme, see Dakubu (1987). The construction type holds a particular interest in that EVCs act as single verbs relative to the environment, but are dividable into word-like units, namely a limited number of *preverbs* (up to three in a row) together with the *main verb*. A simple example is the Ga sentence (14), where *kε* is a preverb:

(14) Tete kε-ba-biε
 Tettey take-come-here
 N V
 “Tettey brought it here.”

The valence of the main verb determines the valence of the whole relative to the containing clause, its subject is necessarily the subject of all the preverbs with the same role, and its Aspect, Modality and Polarity marking is wholly determined from left to right. Most preverbs are intransitive, only *kε* is transitive.

EVCs, which are perhaps themselves frozen SVCs, can serve in the role of verbs also in SVCs, which gives a complex structure of verb embeddings. Unlike SVCs, in EVCs a complementation analysis is reasonable, since a preverb needs to be followed by a verb – another preverb, or the main verb. These analyses are laid out in Dakubu et al. (2007) and Hellan and Dakubu (2010), and included in the grammar formalism.

Also to be mentioned is the construction in (15):

(15) E-ba tsu mli
 3S-AOR-come room inside
 “he entered the room .”

The phrase *tsu mli* in our resources is counted as a noun phrase, and acting as object of *ba* (‘come’, which in addition has intransitive and other uses).⁹

4. Developing the resources

The backbone of a deep computational grammar is its lexicon. The starting point for this lexicon in the Ga grammar is a Toolbox project holding data of the general-purpose published dictionary (Dakubu 2009). The lexicon file in this project consists of 80,000 lines of code, with 7080 entries, of which 5014 for nouns, and 935 for verbs, of which 722 were annotated for valence. From this

⁸ See Hellan 2019 for details.

⁹ Cf. Beermann and Hellan (to appear). In the lexical resource addressed in section 4, this use of *ba* has the code ‘ba_3 : v-tr-qbPostp-suAg_obLoc-MOTIONDIRECTED’.

Toolbox repository a valence lexicon was created. As a first step the Toolbox lexicon was augmented by valence information such that each entry reflects a unique valence frame or multiverb environment. For instance, for the verb *su* as used in the sentence (16),

(16) E-su lɛ e-gbe lɛ
 3S.AOR-bewitch 3S 3S.AOR-kill 3S
 ‘she killed him by magic’

the design of a lexical entry in the amended Toolbox version is as shown in Figure 1; the valence codes are written into the lexical entry following the general ‘field’ style of Toolbox, where the fields marked `\pdl-\pdv` represent inflectional information of the lexeme, and the fields `\xe`, `\xg`, `\xv` together constitute a standard linguistic glossing with `\xv` as a word-and-morph break-up, `\xg` as morphological and English gloss, and `\xe` as a free English translation; the valence (or, as here, SVC environment) is encoded as the fields starting with `\sl...`:¹⁰

```
\lx su
\hm 3
\ph su_`
\ps verb annotated
\sn 1
\de poison
\sn 2
\ge bewitch
\de bewitch, practice black magic, kill by
magic
\sl1 svSuAspIDALL_suAg-
\sl2 v1tr-
\sl3 v1obIDv2ob-
\sl4 obTrgt-
\sl2 v2tr-
\sl4 v2obTh-
\sl5 CAUSATIONwithCAUSINGEVENT-
\sl6 CHANGEofSTATE
\xv E-su lɛ e-gbe lɛ
\xg 3S.AOR-bewitch 3S 3S.AOR-kill 3S
\xe she killed him by magic.
\pdl v. iter
\pdv susui
\pdl n. ag
\pdv sulo
\pdl n. ger
\pdv suu
\dt 15/Jan/2010
```

Figure 1 Example of Ga Toolbox entry enriched with CL valence/construction annotation

A verb with more than one valence frame having one entry specified per frame, the verb *ba*, for instance, is represented by 15 different entries in this edition of the Toolbox file. 547 verb lexemes here received altogether 2006 entries annotated in this fashion. In Figure 1, the

¹⁰ With such IGTs illustrating verbs and smaller phrases illustrating nouns and other POS, these specifications in the Toolbox file constitute a large corpus which however yet remains to be implemented on a standard corpus format; this situation may apply to Toolbox files for other languages as well.

specification ‘`\hm 3`’ indicates that this is the third entry with the form *su*.

The code specifications in the `\sl`-fields are pulled together in a single string as in (17) (omitting `\sl15` and `\sl16`)¹¹, read as ‘a SVC where subject and aspect are identical in all VPs, the role of subject is agent, the head of the second VP is transitive, the first VP’s object is identical to the second VPs object, and the object in the second VP is theme’.¹²

(17) svSuAspIDALL_suAg-v2tr-v1obIDv2ob-v2obTh

Labels in this style were independently developed as the system *Construction Labeling* formalism (CL) (cf. Hellan and Dakubu 2010, Dakubu and Hellan 2016), and one of the languages to which it was applied was Ga, in a construction type inventory given in Hellan and Dakubu 2010.

The verb part of the lexical resource was turned into a lexical data structure of the type used in HPSG grammars, consisting of 1980 sequentially numbered entries, with the CL specification indicating the *lexical/construction type* to which the entry belongs.¹³ Figure 2 is the direct counterpart to the Toolbox entry in Figure 1, with *su_1448* as the entry identifier, and the formula part

meaning ‘belongs to the construction type ‘...’; this information is stated relative to the first verb, which is thus, formally, counted as a head:

```
su_1448 := svSuAspIDALL_suAg-v2tr-v1obIDv2ob-v2obTh &
[STEM <"su">,
PHON <"su">,
ENGL-GLOSS <"bewitch">,
EXAMPLE "E-su lɛ e-gbe lɛ",
GLOSS "3S.AOR-bewitch 3S 3S.AOR-kill 3S",
FREE-TRANSL "she killed him by magic."].
```

Figure 2 Grammar style counterpart to entry in Figure 1

While this is a constructional representation, a valence representation can be derived in a similar manner from a Toolbox entry, thus, the entry in Figure 3 will directly go into the grammar lexicon as a valence entry of type ‘transitive with agentive subject and theme object’:

¹¹ This is a *Situation Type* label, an aspect of analysis not so far fully integrated in the grammar; as an annotation resource, cf. Hellan (2020). It’s a common observation that many SVCs express a ‘unique situation’, thus that the verb meanings do not constitute separate events but are merged into a single event. Hellan (2019b) is an attempt to give formal expression to this notion, in terms of a layer of semantic representation called *Situation Structure*, whose interaction with *Semantic Argument Structure* is outlined in Hellan (2019a), and whose encoding in *Situation Type labels* is outlined Hellan (2020).

¹² The identity of the objects is here not included in the ‘IDALL’ part, leaving open if it is a matter of argument sharing rather than pronominal coreference. Since Ga often realizes subject sharing through pronominal pro-clitics, a counterpart to this strategy for objects could be conceived.

¹³ Conducted by Tore Bruland; also cf. Hirzel 2006.

```

fee_244 := v-tr-suAg_obTh &
[STEM <"fee">,
PHON <"fee">,
ENGL-GLOSS <"make">,
EXAMPLE "E-fee flɔɔ, samala",
GLOSS "3S.AOR-make stew, soap",
FREE-TRANSL "she made stew, soap."].

```

Figure 3 Ga Grammar style valence entry

The sentence in (16) also parses by the grammar (formally treating the second VP as an adjunct, as outlined above), but only relative to a transitive frame for *su* corresponding to the meaning ‘bewitch’.

For a language with many multiverb construction types, it might be tried to feed total expressions like (5) into the verb frame of the first verb, and construe also the formal adjunct as ‘foreseen’ by the head verb, and thus separating the technical combination frame from what is intuitively a valence frame. This, however, is not a possibility that the present grammar ventures into.

To have an impression of how frequently a multiverb expression may be associated with a given verb, the following table indicates how often the types SVC (as ‘sv’), EVC (as ‘ev’) and Verbid construction (as ‘trVid’ or ‘intrVid’) are among the environments in which a verb can occur (for instance, 14 verb lexemes can occur in both intransitive, transitive and SVC environment), according to the resource built on Dakubu (unpublished a):

Table 1 Distribution of verbs over valence frames and construction types in Ga

| | |
|----------------------------|-----|
| {tr} | 144 |
| {intr} | 51 |
| {intr,tr} | 44 |
| {tr,ev} | 23 |
| {tr,sv} | 15 |
| {ev} | 15 |
| {intr,tr,sv} | 14 |
| {tr,ditr} | 9 |
| {intr,tr,ev} | 6 |
| {intr,tr,ditr} | 6 |
| {tr,ditr,ev} | 6 |
| {intr,intrVid,tr} | 6 |
| {tr,ditr,sv} | 5 |
| {intr,tr,ev,sv} | 4 |
| {intr,tr,trVid,ditr,ev,sv} | 4 |
| {intr,tr,ditr,ev,sv} | 4 |
| {intrComp,tr} | 3 |
| {tr,ev,sv} | 3 |
| {intr,tr,ditr,sv} | 3 |
| {tr,trVid} | 3 |
| {ditr,ev} | 3 |
| {intr,tr,ditr,ev} | 2 |
| {intrVid,tr,sv} | 2 |
| {intrVid,tr} | 2 |
| {intrVid,tr,trVid} | 2 |
| {tr,ditr,ev,sv} | 2 |
| {intr,tr,trComp} | 2 |

Calibrating these kinds of multiverb environments into a lexicon or grammar resource otherwise will be among the interesting next steps in dealing with digital resources for Volta Basin Kwa languages.

5. The grammar and its environment

As said above, a first version of the Ga grammar follows the formal architecture of Pollard and sag (1994), as used in many grammars adopting the Grammar matrix as a common feature structure repertory (Bender et a. (2010)). The later version, as illustrated here, has a simpler feature structure, which aims at more directly accommodating variation among languages. A guide to its feature structure is given in the pdf ‘Building Global Grammar’, which takes as point of departure a simple introductory grammar for English used as illustration in Copestake (2002), and stepwise builds up what is there called a ‘global’ feature structure, both with implementations and description. The description can be accessed at <https://typecraft.org/tc2wiki/TypeGram>, with a link from the Introduction, while the implementation is linked from <http://regdili.hf.ntnu.no:8081/typegramusers/menu>, with instructions at the TypeGram site. The specifications concerning Ga sit in a common repository of features covering also Germanic (Norwegian¹⁴), Bantu (Luganda) and Ethio-semitic (Kistaninya), as illustrated in Figure 5, here with a highlight on the Ga grammar (‘GaGram’):

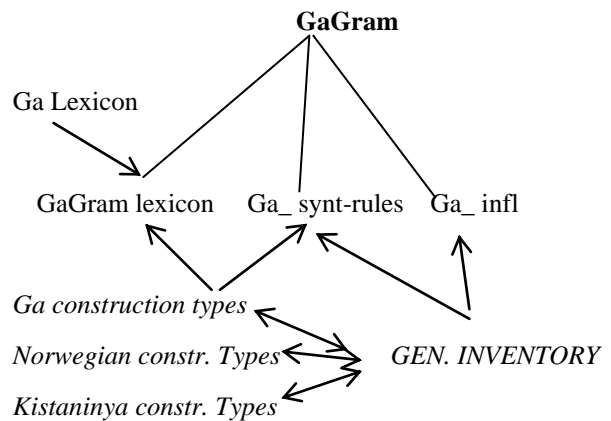
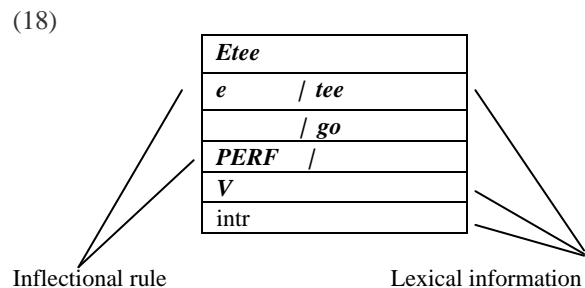


Figure 5 Architecture of resources

Among applications surrounding the grammar system is a procedure for grammar induction from Interlinear Glossed text (IGT), described in Hellan and Beermann (2011, 2014) and Bruland 2011, schematically indicated with an IGT snippet in (18):



¹⁴ The Norwegian branch of this system is distinct from the large scale Matrix-based grammar Norsource, (cf. Hellan and Bruland 2015), which has been maintained since 2001. Code files are downloadable from GitHub: <https://github.com/Regdili-NTNU/NorSource/tree/master>, and it has a web demo at <http://regdili.hf.ntnu.no:8081/linguisticAce/parse>. Also this grammar uses GF features, and its lexical types belong to the same system as what is presently described.

The inflectional rule here induced is

```
verb-Perf_irule := %prefix (* e) word & [ TAM.T perf,
DTR < v-lxm > ],
```

and the lexical information induced is

```
tee-v := v-intr_lx & [ ORTH <"tee">, ACTNT.PRED
tee_rel ].
```

In addition, one can also explore the GLOSS line specifications to obtain ‘meta-string’ versions of sentences of the language, in the case in point with

```
verb-Perf_irule := %prefix (* PERF) word & [ TAM.T
perf, Stem < v-lxm > ].
```

as inflectional rule and

```
go-v := v-intr_lx & [ORTH <"go">, ACTNT.PRED go_rel ].
```

as lexical information induced. Both courses are described at <https://typecraft.org/tc2wiki/TypeGram>. The latter course instantiates a procedure where the modules of syntactic and semantic parsing can be conducted separate from morphology.¹⁵ Thus, relative to a sentence like (19),

(19)

| | | | | |
|------------|---------|-----|--------|------|
| Ame-wo | tsone | le | mli | yele |
| 3P.AOR-put | vehicle | DEF | inside | yam |
| V | N | Art | N | N |

‘They put vehicle’s inside yam’ = ‘They put yams in the lorry.’

such a ‘meta’ approach will address the construction in the shape (20a) rather than (20b):

- (20) a. 3PputAor *vehicle* DEF *inside* *yam*
b. *Ame-wo tsone le mli yele*

This method sustains a use of 145 sentences on the format of (20a) serving as an intermediate test suite for the full set of valence and construction types described in Hellan and Dakubu (2010).

6. Ga valence and verb construction dictionary

The Ga valence dictionary resources are represented as Dakubu (unpublished a) as a conversion from the enriched Toolbox version described in section 4, and in Dakubu (unpublished b) as a larger monograph. The material in Dakubu (unpublished a) is also online accessible as part of MultiVal, a comparative valence resource based on lexicons from LKB grammars for four languages.¹⁶

The point where this resource fits into the view of Figure 5 is, through its use of the CL labeling, marking its place within a compact cross-linguistic comparison of *language valence profiles*, which are enumerations of the valence and construction frames realized in a language. Preliminary comparisons of valence profiles for Ga and English suggest that they have less than 20% of their valency frames in common (see, e.g, Dakubu and Hellan (2016, 2017)).

¹⁵ See Dakubu (2002) on the tone system, whose lexical and syntactic impact is not yet reflected in the grammar.

¹⁶ Cf. Hellan et al (2014). Online site: http://regdili.hf.ntnu.no:8081/multilanguage_valence_demo/multivalence

An investigation of valence types in Ga can be related to the research into valence classes started with Levin 1993, followed up, i.a., in VerbNet and in the Leipzig Valency Classes (LVC) Project,¹⁷ being attempts to associate commonalities in morpho-syntactic patterns with semantic factors, both language internally (like Levin and VerbNet) and cross-linguistically (LVC). Establishing valency classes for Ga has a tie to VerbNet in aiming at a fairly large coverage of the language’s verbs,¹⁸ and to LVC in establishing one more coordinate point in the attempt to attain a typologically broad basis for generalizations within this domain.

Given the large discrepancies in valence frames between Ga and English, a good strategy may be to first explore commonalities between Ga and other West African languages¹⁹. In the present setting, a natural step will for instance be to build a mapping between Ga and Akan lexical information, assuming that the valence labels used for Ga are adequate also for Akan.²⁰

7. Conclusion

The view taken on the creation of a ‘deep’ computational grammar is that it allows one to

(i) through execution, create a formally tractable representation of structures of the language, where the execution binds one to consistency:

(ii) reflect on what are the essential structures of the grammar studied, and their relation to structures of other languages for which similar formally consistent investigations have been made;

(iii) effectively port one’s findings to, or into the creation of, other resources.

Not least in the setting of African languages, an additional concern is to identify efficient ways of utilizing existing linguistic resources for the language in question, combined with a formal framework allowing for proper representation of the facts.

For languages with few previous digital facilities, a goal is to be able to develop a number of resources and applications in interaction but at a speed which allows one to digest and actively explore given and new connections. This is the goal of the resources here described, and we have highlighted the role of Prof. Dakubu’s lexical resources, the way a grammar’s organization of lexical information can lead to further resources, and we have discussed bearings that linguistic structures of Ga, and presumably Volta Basin Kwa in general, have on the formal structures of a grammar framework.

¹⁷ Cf. for LVC, Malchukov and Comrie (eds) 2015 and <http://valpal.info/>; for VerbNet <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>.

¹⁸ It is worth noting that while the VerbNet resource is essentially just a knowledge base, the present system is also used as an integral part of a syntactic parser.

¹⁹ Perspectives are offered in Atoyebi (2015), Schaefer and Egbokhare. (2015), Creissels (2015), in the frame of LVC.

²⁰ Cf. Beermann and Hellan (2018) and (to appear).

8. Bibliographical References

- Atoyebi, Joseph Dele Valency classes in Yorùbá. In: Malchukov, A., and B. Comrie (eds), pages 299-326.
- Beermann, Dorothee, and Lars Hellan. 2018. West African Serial verb constructions: the case of Akan and Ga. In: Agwuele, Augustine, and Adams Bodomo (eds) *The Routledge Handbook of African Linguistics*. London and New York: Routledge. Pg. 207-221.
- Beermann, D. and L. Hellan. To appear. Enhancing grammar and valence resources for Akan and Ga
- Bender, Emily.M., Drellishak, S., Fokkens, A., Poulson, L. and Saleem, S. 2010. Grammar Customization. In *Research on Language & Computation*, Volume 8, Number 1, 23-72.
- Bruland, T. (2011). Creating TypeGram data from TypeCraft. Presentation at *India 2011*, NTNU.
- Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. CSLI Publications.
- Creissels, Denis. 2015. Valency properties of Mandinka verbs. In: Malchukov, A., and B. Comrie (eds) Pages 221-260
- Dakubu, M.E. Kropp, 2002. *Ga Phonology*. Institute of African Studies, Legon, Ghana.
- Dakubu, M.E. Kropp, 2004a. The Ga preverb *kɛ* revisited. In Dakubu and Osam, eds., *Studies in the Languages of the Volta Basin 2*: 113-134. Legon: Linguistics Dept.
- Dakubu, M.E. Kropp, 2004b. Ga clauses without syntactic subjects. *Journal of African Languages and Linguistics* 25.1: 1-40.
- Dakubu, M.E. Kropp, 2008 Ga verb features. In Ameka and Dakubu eds., *Aspect and Modality in Kwa Languages*. Amsterdam & Philadelphia: John Benjamins Publishing Co. p. 91-134.
- Dakubu, M. E. Kropp, 2009. *Ga-English Dictionary with English-Ga Index*. Accra: Black Mask Publishers
- Dakubu, Mary Esther Kropp. Unpublished a. 'Ga_verb_dictionary_for_digital_processing'. Accessed for download at: https://typecraft.org/tc2wiki/Ga_Valence_Profile
- Dakubu, Mary Esther Kropp. Unpublished b. Ga Verbs and their constructions. Monograph ms, Univ. of Ghana.
- Dakubu, M.E.K., L. Hellan and D. Beermann. 2007. Verb Sequencing Constraints in Ga: Serial Verb Constructions and the Extended Verb Complex. In St. Müller (ed) *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar*. Stanford: CSLI Publications. (<http://csli-publications.stanford.edu/>)
- Dakubu, M.E. Kropp and Lars Hellan. 2016. Verb Classes and Valency classes in Ga. Paper read at Symposium on West African Languages (SyWAL) II, Vienna.
- Dakubu, M.E. Kropp and Lars Hellan. 2017. A labeling system for valency: linguistic coverage and applications. In Hellan, L., Malchukov, A., and Cennamo, M (eds) *Contrastive studies in Valency*. Amsterdam & Philadelphia: John Benjamins Publ. Co.
- Hellan, Lars. 2019a. Construction-Based Compositional Grammar. March 2019. *Journal of Logic Language and Information*. DOI: 10.1007/s10849-019-09284-5
- Hellan, Lars and M.E. Kropp Dakubu. 2010. *Identifying verb constructions cross-linguistically*. In *Studies in the Languages of the Volta Basin* 6.3. Legon: Linguistics Department, University of Ghana.
- Hellan, Lars, and Dorothee Beermann. 2014. Inducing grammars from IGT. In Z. Vetulani and J. Mariani (eds.) *Human Language Technologies as a Challenge for Computer Science and Linguistics*. Springer.
- Hellan, L., D. Beermann, T. Bruland, M.E.K. Dakubu, and M. Marimon. 2014. *MultiVal*: Towards a multilingual valence lexicon. In Calzolari, Nicoletta et al. (eds.) *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, pages 2785–2792, Reykjavík, Iceland. ELRA.
- Hellan, Lars and Tore Bruland. 2015. A cluster of applications around a Deep Grammar. In: Vetulani et al. (eds) *Proceedings from The Language & Technology Conference (LTC) 2015*, Poznan.
- Hirzel, Hannes. 2006. "Porting lexicon files from Toolbox into LKB-grammars: A case study for a grammar of Ga." https://typecraft.org/tc2wiki/File:Toolbox-LKB-Link-slides_-_version_4.pdf.
- Levin, B. 1993. *English Verb Classes and Alternations*. Chicago IL: University of Chicago Press.
- Malchukov, Andrej L. & Comrie, Bernard (eds.) (2015). *Valency classes in the world's languages*. Berlin: De Gruyter Mouton. 2015.
- Pollard, Carl and Ivan Sag (1994) *Head-driven Phrase Structure Grammar*. Chicago Univ. Press.
- Rask, R. 1828. *Vejledning til Akra-Sproget på Kysten Ginea* ('Introduction to the Accra language on the Guinea Coast').
- Schaefer, R.B, and Francis O. Egbokhare. 2015. Email valency classes and their alternations. In Malchukov, A. and B. Comrie (eds) 2015. Pp. 261-298.

9. Language Resource References

<https://typecraft.org/tc2wiki/TypeGram>

https://typecraft.org/tc2wiki/Ga_Valence_Profile