HaG — A Computational Grammar of Hausa

Berthold Crysmann
CNRS, Laboratoire de linguistique formelle, Paris–Diderot

1. Introduction

In this paper, I shall give an overview of HaG (=**Ha**usa **G**rammar), an emerging computational grammar of Hausa[1], developed within the framework of Head-driven Phrase Structure (Pollard and Sag, 1987, 1994; Sag, 1997). Since HPSG is an integrated theory of syntax and semantics, meaning representations are built up in tandem with syntactic analysis. Semantics in HaG are represented using Minimal Recursion Semantics (MRS; Copestake, Flickinger, Pollard and Sag, 2005), essentially providing predicate-argument structures with an unspecified representation of (quantifier) scope.

The grammar described in this paper runs on top of the Lingo *Linguistic Knowledge Builder* (LKB; Copestake, 2002), a platform for typed feature structure grammars originally developed at CSLI, Stanford. The LKB system not only provides a bottom-up chart parser (Oepen and Carroll, 2000), but also a chart generator (Carroll, Copestake, Flickinger and Poznanski, 1999; Carroll and Oepen, 2000). Consequently, HaG was designed from the start as a reversible grammar, i.e. a grammar that is suitable for both analysis and synthesis. In addition to the development platform LKB, HaG can also be run using the efficient C++ parser *Pet* (Callmeier, 2000), and since autumn 2011, on the reversible *ace* parser/generator developed in C by Woodley Packard (`http://sweaglesw.org/linguistics/ace/`). The grammar, as well as all the development and run-time systems, are available under free and open source licenses. As an alternative to a full-fledged install of the grammar and development systems, we provide a web demo (`http://hag.delph-in.net`) which provides a concise interface to the grammar, displaying full semantic representations, as well as the constituent structure.

Development of the grammar started in 2009, based on the LinGO grammar matrix (Bender, Flickinger and Oepen, 2002), a core system of basic types extracted from the English Resource Grammar (ERG; Copestake and Flickinger, 2000) which ensures basic compatibility of semantic representations between LKB grammars using MRS.

Implementation of a formal grammar of Hausa is motivated by two major goals: first, the availability of an implemented grammar will contribute a reusable linguistic resource for a computationally underresourced language. Second, the implementation of a competence grammar based on a linguistically motivated formalism such as HPSG will provide testable models of linguistic theories. Since HaG is the first implemented grammar of a tone language that systematically integrates suprasegmental phonology, we hope to also further our understanding of the computational treatment of (African) tone languages.

Current development of HaG focuses on the implementation of syntactic constructions and the system of morpho-syntactic rules. This decision is deliberate, since we plan to expand the lexicon using grammar-based machine learning techniques (Zhang and Kordoni, 2006). As a result, the grammar already covers a substantial part of Hausa core grammar, despite the comparatively small lexicon. In this paper, I shall provide an overview of the main constructions of the language as implemented in HaG. Following a detailed overview of the central issues concerning the treatment of tone and length (section 1), I shall briefly discuss the implementation of inflectional morphology (section 2). Section 3 will be devoted to morpho-syntax, including direct object marking, mixed categories, and pronominal

[1] Throughout this paper, I use the following conventions: long vowels are marked by a macron, whereas short vowels are left unmarked. A grave accent signals a low (L) tone and circumflex a falling (HL) tone, while high tones (H) are left unmarked. In addition to conventions of the Leipzig Glossing Rules `http://www.eva.mpg.de/lingua/resources/glossing-rules.php`, I use the following glosses: CONTINUATIVE for continuative aspect, LINKER for the genitive linker *-n/-r*, and A, B, and C to identify the A-, B-, and C-forms of verbs, gerunds, nouns, adjectives, and genitive prepositions.

 Selected Proceedings of the 42nd Annual Conference on African Linguistics, ed. Michael R. Marlo et al., 321-337. Somerville, MA: Cascadilla Proceedings Project.

affixation. The paper will conclude with an outline of the current treatment of non-local dependencies. Besides serving as a documentation of the grammar, this paper also serves the purpose of connecting the concrete implementation to the underlying theoretical analysis.

2. Autosegmental representations

It is a well-known fact about Hausa that both tone and vowel length are distinctive suprasegmental properties: at the surface level, Hausa distinguishes two level tones (H and L), as well as a falling contour tone (HL), typically analysed in terms of two level tones associated with a single tone-bearing unit. As is common for African tone languages, suprasegmental information such as tone and vowel length not only serve to distinguish lexical meaning, but also grammatical notions. One of the major challenges for a computational treatment of Hausa is contributed by the fact that different writing systems reflect suprasegmental information to differing degrees: while standard orthography (*boko*) does not provide any suprasegmental marking, neither tone, nor vowel length, the traditional Arabic script (*ajami*) does mark vowel length, but not tone. In scientific and foreign language learning contexts, i.e. scholarly work, textbooks, and dictionaries, both tone and length tend to be marked, albeit by different marking regimes: while most grammars mark long vowels (either diacritically by means of a macron or else by vowel gemination), Newman and Ma's *Sabon ƙamus na Hausa zuwa Turanci* (Newman and Ma Newman, 1977) marks short vowels (by means of a cedilla diacritic). Similarly, for tone, there are high-marking regimes (Jungraithmayr, Möhlig and Storch, 2004) alongside the more common strategy of marking low tone. The diversity of suprasegmental marking must of course have direct implications for grammar design: on the one hand, processing should be robust towards the absence of (crucial) suprasegmental information, whereas, on the other, any amount of suprasegmental annotation present should be taken into consideration, no matter the detail or the marking convention.

Apart from input conventions, the amount of suprasegmental information required may also vary according to processing direction (parsing/generation) or application scenario: in text-to-speech applications, e.g., suprasegmental information will typically be absent in the textual input, but needs to be recovered prior to speech generation. Similarly, in the context of CALL (=Computer-assisted Language Learning) suprasegmentally unmarked input text may get annotated with tone and length information.

While orthographic and processing considerations already militate in favour of autosegmental representations in the grammar, the conclusive argument for a separation of segmental and suprasegmental information is contributed by the language's morphology: alongside essentially local processes, i.e. affixation of toned material, possibly triggering local sandhi, Hausa exhibits a number of morphological processes featuring suppletive assignment of entire tonal melodies, often unrelated to those of the base. This is most evident with Hausa plural formation, which is characterised by a high number of "tone-integrating" suffixes (Newman, 2000). Segmental material, however, as well as vowel length, is largely preserved, modulo highly local morpho-phonological adjustment. It follows that a satisfactory treatment of morphological tone in Hausa necessitates the clean separation of tone, length, and segmental information along the lines of Autosegmental Phonology (Goldsmith, 1976; Leben, 1973).

To summarise, the kind of separation of levels embodied by autosegmental representations not only provides the key to a flexible and robust treatment of suprasegmental information in the input, but also constitutes a necessary prerequisite for a generalised and linguistically adequate treatment of Hausa morphology.

Within HaG, suprasegmental processing happens at essentially two levels: first, tone and length information are systematically integrated into the grammar, both at the lexical level, as well as at the level of morphological (and phrasal) rules. Second, in order to accommodate different degrees and conventions for suprasegmental marking of the input, the grammar is equipped with a set of pre-processing rules that serve to build autosegmental representations from diacritically marked input representations.

Grammar-internally, tone and length information of lexical items is represented as lists, specifying sequences of *high*, *low*, *fall* for tone, or *long* and *short* for length. Owing to the fact that Hausa morphology is predominantly suffixal and that the underlying formalism (LKB) facilitates operations at beginning of the list, but not the end, tone and length are encoded in reverse order. Furthermore, the reverse encoding lends itself pretty well to a treatment of tonal spreading (see below). A sample lexical representation of the word *àlmùbazzàrī* 'spendthrift' is given in Figure 1 below.

$$\begin{bmatrix} \text{ORTH} & \text{“almubazzari”} \\ \text{SUPRA} & \begin{bmatrix} \text{TONES} & \langle \textit{high, low, high, low, low} \rangle \\ \text{LEN} & \langle \textit{long, short, short, short, short} \rangle \end{bmatrix} \end{bmatrix}$$

Figure 1: Lexical representation of suprasegmentals: *àlmùbazzàrī*

While grammar-internal representations are fully specified for all aspects of phonology, i.e. segmental and suprasegmental, textual input tends to be underspecified with respect to suprasegmental information.

Tone and length information, if specified in the input, are typically encoded by means of diacritics, i.e. as a property of (vocalic) segments. In order to separate levels of information, we employ an input chart rewriting mechanism (Adolphs, Oepen, Callmeier, Crysmann, Flickinger and Kiefer, 2008) to transform the (diacritically marked) input string into a feature structure representation (Figure 2).[2]

As part of the conversion, the preprocessor also normalises across different marking conventions: at present, three different ways of length marking are recognised, namely double vowels, colon (`:`), and macron. Furthermore, as to the representation of diacritics, the preprocessor accepts precombined accented Unicode characters, Unicode combining diacritics, as well as combinations of the two, e.g., precombined ē with a combining grave accent (ȅ).

$$\begin{bmatrix} \text{ORTH} & \text{“almubazzari”} \\ \text{SUPRA} & \begin{bmatrix} \text{TONES} & \langle \textit{utone, low, utone, low, low} \rangle \\ \text{LEN} & \langle \textit{long, ulen, ulen, ulen, ulen} \rangle \end{bmatrix} \end{bmatrix}$$

Figure 2: Feature structure representation of low tone marked input *àlmùbazzàrī*

$$\begin{bmatrix} \text{ORTH} & \text{“almubazzari”} \\ \text{SUPRA} & \begin{bmatrix} \text{TONES} & \langle \textit{high, utone, high, utone, utone} \rangle \\ \text{LEN} & \langle \textit{long, ulen, ulen, ulen, ulen} \rangle \end{bmatrix} \end{bmatrix}$$

Figure 3: Feature structure representation of high tone marked input *almubázzaríí*

Minimally, token rewrite rules convert diacritically-marked input segments into unmarked segments and introduce a corresponding tone and length type into the autosegmental representation. Syllable nuclei unmarked for tone or length give rise to an underspecified suprasegmental specification (*utone* or *ulen*). Before parsing, input tokens are instantiated with lexical entries, unifying segmental and suprasegmental descriptions. Standard orthography input then just constitutes a special sub-case where suprasegmental constraints come exclusively from the grammar. Likewise, input in ajami will only contain length distinctions, with tonal information being crucially underspecified.

In addition to extracting tone and length, the preprocessor also registers the marking regime used. The parser can be configured at run-time whether it should assume a consistent marking regime, where segments unmarked for tone/length are interpreted bearing the complementary tone of the ones overtly marked, or rather a sporadic marking regime, where no inferences are drawn regarding unmarked syllables.

Under the assumption of a consistent marking regime, presence of, e.g., a low-marked segment will lead to an interpretation of unmarked segments as high. Conversely, if a single high-marked segment

[2] The switch to input chart mapping marks the main difference between the current approach to tone/length diacritics to the earlier one discussed in Crysmann (2009).

is detected in the input, all unmarked segments will be interpreted as low. Assumption of a consistent marking regime is most useful for parsing edited scientific text or the output of a speech recogniser, where is is safe to assume consistent input conventions. Note, however, that the type of marking regime is still inferred automatically from whatever (diacritic) annotations are found in the input. To give an example, an input such as `ya zóó` will get interpreted as *yà zō*.

The sporadic marking regime, by contrast is most useful for interactive user input. Under this regime, the user can specify individual tones, yet no inference is drawn as to the interpretation of unmarked tones. Thus, the user may specify certain critical tones, e.g., for disambiguation, without being forced to consistently specify all occurrences of this tone throughout the input. Taking the input `Allah yà gafarta malam` as an example, under a sporadic marking regime, the user can specify the tone of the subjunctive marker without having to carefully specify all other low tones, as in `Allàh yà gaafàrtà maalàm`.

Although unrelated to the treatment of suprasegmental phonology proper, the grammar recognises one further level of robustness towards non-standard input, namely absence of hooked letters: while the grammar readily accepts letters postfixed with an apostrophe as equivalent to hooked letters, it also caters for an ultra-robust mode where marking of glottalised consonants is not required at all.

As opposed to parsing, which provides various levels of underspecification, the generator standardly produces fully tone and length marked output using the convention employed in the two recent reference grammars of Hausa (Newman, 2000; Jaggar, 2001). Technically, diacritic marking in generation is achieved by means of a regular expression substitutions that translate the grammar-internal representations into tone and length marked surface strings.

The systematic treatment of tone and length as implemented in the grammar, together with its robustness towards suprasegmental marking in the parser's input provide a solid basis for tone reconstruction: once we can identify the correct reading from the set of available analyses, e.g., by means of a probabilistic model,[3] we can regenerate a suprasegmentally fully specified surface string. Compared to dedicated diacritic reconstruction approaches, as proposed for African languages by, e.g., De Pauw, Wagacha and De Schryver (2007), the grammar-based approach has the advantage of tying the specific problem of diacritic reconstruction to the more general issue of syntactic disambiguation. Since Hausa standard orthography input is devoid of suprasegmental information, this added ambiguity is part of the parsing problem anyway. With a grammar that specifies not only lexical, but also local and non-local grammatical constraints on tone and length, statistical disambiguation will actually be supported by symbolic constraints.

3. Inflectional morphology

While verbal inflection including TAM marking and subject agreement is mostly expressed by syntactically independent markers rather than morphologically bound forms, the system of Hausa plural inflections is particularly rich. In HaG, a set of 36 morphological rules models the 14 nominal plural classes identified in Newman (2000), including their subclasses (up to 5). For testing, the grammar's lexicon contains entries for each of these classes, including exhaustive listing for some unproductive classes.

The complexity of Hausa plural inflection is not only due to the sheer number of different paradigms and the fact that some lexemes can subscribe to more than one of these paradigms, but also by the richness of formal devices employed by the plural formation processes. Thus, alongside common or garden suffixation, we find a plethora of non-concatenative processes including gemination (e.g., *da**m**ì* ↦ *dà**mm**ai*), root consonant reduplication (e.g., *tàmbay**à*** ↦ *tamba**y**ōy**ī***), different forms of partial reduplication (e.g.,*ɓē**rā*** ↦ *ɓē**ràrra**kī*), as well as total reduplication (e.g., ***nâs*** ↦ ***nâs nâs***).

With the exception of total reduplication, both concatenative and non-concatenative are implemented by means of LKB's built-in string unification formalism. Total reduplication, however, which formally goes beyond the power of LKB's orthographemic component, is modelled in syntax by means of a

[3] The current version of HaG already comes with a smallish statistical parse selection model, trained on the grammar's test suite, using the technology developed by Toutanova, Manning, Shieber, Flickinger and Oepen (2002). Similarly, for realisation ranking we build on the proposal by Velldal and Oepen (2005)

semantically non-compositional binary rule.

A recurring issue of Hausa plural formation is what Newman (2000) calls tone-integrating suffixes, i.e. morphological processes by which a suppletive tonal melody is assigned holistically to the entire derived form. Well-known examples of tone suppletion including the highly productive all-H plural pattern I (*tàmbayà̀* 'question' ↦ tambayōyī) and the equally productive L-H (*àlmùbazzàrī* 'spend-thrift' ↦ *àlmùbàzzàrai*), with right to left automatic spreading. As we have just seen, the tonal representation of the base is completely overwritten by plural formation, whereas segmental material and length information of the base is largely preserved. Since all three levels are already represented separately in HaG, drawing on the basic insight of Autosegmental Phonology, the only remaining issue to be addressed in the light of holistic melody assignment with automatic spreading is how to specify tonal constraints independently of the number of tone-bearing units. To this end, HaG employs typed list constraints, such as the ones depicted in Figure 4: as stated by the excerpt from the type hierarchy, an all-H list *h*-list* can be either an empty list *h*-empty-list*, or else a non-empty list *h*-non-empty-list*. The sub-type *h*-non-empty-list* constrains its first list element (FIRST) to be *high*, and the remainder of the list to be again of type *h*-list* (possibly empty). If the REST features contains an element, e.g. for a two-elementary list, the type *h*-list* will be specialised to *h*-non-empty-list* enforcing all its associated constraints.[4] As a consequence, lists of this type recursively state that all its members (however many) will be required to be *high*.

HaG currently recognises 15 distinct tonal melodies, with L-H-L-H as the most complex pattern. This list is not necessarily complete and may grow, to some limited extent, as the lexicon is extended.

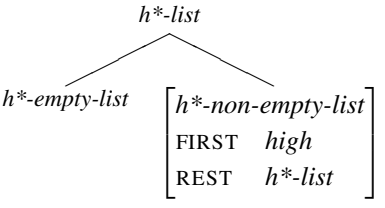

Figure 4: Implementing tone spreading as typed list constraints

With an implementation of automatic spreading in place, morphological rules of Hausa plural formation can now invoke these list constraints in order to model suppletive tone assignments (cf. Figure 5).

$$\begin{bmatrix} \text{SUPRA} & \begin{bmatrix} \text{TONE} & \textit{h*-list} \\ \text{LEN} & \left\langle \textit{long, long} \mid \boxed{1} \right\rangle \end{bmatrix} \\ \text{DTR} & \begin{bmatrix} \text{MCLASS} & \textit{n-pl-1} \\ \text{SUPRA} & \left[\text{LEN} \quad \left\langle \textit{len} \mid \boxed{1} \right\rangle \right] \end{bmatrix} \end{bmatrix}$$

Figure 5: Holistic melody assignment by morphological rule (noun plural I)

As depicted by the representation of the morphological rule for the regular noun plural class I in Figure 5, tonal information is entirely represented by means of a list constraint. Application of the morphological rule is constrained according to morphological class membership of the base (in DTR). However, the tonal make-up of the derived form does not make any reference to the tonal properties of the base, which is characteristic of suppletive melody assignment. Length restrictions, by contrast, are much more conservative in their mode of operation: in the example given here, all length constraints

[4] Since the feature FIRST is not appropriate for either the empty list, or for the super-type *h*-list* it follows by virtue of the logic of typed feature structures (Carpenter, 1992; Krieger, 1996; Copestake, 2002) that whenever a list of type *h*-list* has a member, its type will automatically get specialised to *h*-non-empty-list*, imposing all the constraints that this type imposes.

are shared with those of the base (1), with the exception of the final syllable of the base, which may be long or short, as indicated by the supertype *len*. Instead, two syllable constraints corresponding to the *–ō–ī* plural suffix are added to the front of the list. (Recall that tone and length lists are encoded in reverse order.)

In general, length alternations in Hausa (and therefore in HaG) are always local, manipulating right-peripheral vowels, either by lengthening or shortening, or else by suffixation of new vowels (and therefore new length specifications). Tonal operations may also be of this nature, but the typed-list approach permits the expression of holistic tone assignment with right-to-left spreading, alongside local tone alternations at the right edge (substitution, suffixation).

Before we close this discussion, I would like to briefly address a few limitations imposed by the underlying formalism: first, owing to the fact that input strings are handled by the processing platform in an entirely different way from the rest of the grammar's constraints, there is no *direct* way to implement the interdependence between syllable (CV(C)) structure and vowel length, e.g. vowel shortening in closed syllables. At present, this has to be taken care of explicitly by the grammar writer. As a partial solution, however, it is possible to invoke regular expressions in the input chart mapping to impose these phonotactic constraints. Second, and related to this is the issue of synchronising suffixation of both length and (non-holistic) tone specifications in morphology. Grammar-internally, this problem is mitigated, though, by providing a set of rule types that defines the set of well-formed combinations. Note, however, that none of these limitations have any repercussions on the possibility to model the morphosyntax of tone and length in Hausa (or any other language): it merely means that grammar writers may end up having to stipulate some of the more systematic phonological dependencies in a case by case fashion, a limitation that is defensible in a grammar that focuses on syntax and semantics.

In sum, the adoption of tone list constraints has proven to be highly successful at connecting inflectional processes to suprasegmental alternations. I hope that the current implementation can serve as a model for, or even be reused in future implementations of other African tone languages.

4. Morphosyntax

4.1. Direct object marking

Verbs in Hausa undergo inflectional marking according to their local syntactic environment (Parsons, 1960), signalling whether or not a direct object is present *in situ*. Traditionally, three syntactic environments are distinguished: the A-frame, subsuming intransitive uses, pro-dropped objects, and extracted (=non-local) objects, the B-frame, characterised by an adjacent (weak) pronominal direct object, and the C-frame, denoting environments with a VP-local (=*in situ*) NP direct object. The forms corresponding to these frames are named accordingly: A-form, B-form and C-form. Marking of these forms varies across verb paradigms (called grades in the Hausa literature): while in some paradigms (grades 1 and 4), the C-frame is distinguished from both A- and B-frames by means of final vowel shortening, grade 2 additionally shows alternation of the final vowel (cf. 1).[5]

(1) a. nā sàyi gōrò
1.SG.COMPL buy.C cola.nut
'I bought cola nut.'

b. nā sàyē -shì
1.SG.COMPL buy.B -3.SG.M.ACC
'I bought it.'

c. gōrò na sàyā
cola.nut 1.SG.COMPL buy.A
'It's cola nut I bought.'

[5] Note that direct object marking only marks presence vs. absence of a local direct object together with its mode of realisation (object NP vs. pronominal affix), but no other properties of its dependent: more specifically, it is not a marker of object agreement.

As to the theoretical status of direct object marking, two competing views can be found in the literature: while Hayes (1990) claims that Final Vowel Shortening (i.e. frame C forms) is an allomorphic process triggered by an adjacent NP at the syntax phonology interface, I have argued extensively in previous work (Crysmann, 2005) that direct object marking is better understood as an inflectional process pertaining to valency rather than surface adjacency. As witnessed by the data below, surface adjacency of an NP constituent is neither a necessary nor a sufficient criterion, since on the one hand structural direct objects can trigger pre-object forms across intervening discourse particles, and on the other, verb-adjacent oblique NP complements fail to trigger direct object marking altogether.[6]

(2) tā tàmbàyi kùwa mȁ̂târ
3.SG.F.COMPL ask.C moreover woman.DEF
'She asked moreover the woman.' (Newman, 2000)

(3) Kànde cȅ Bintà ta ɗarȁ̂ tsāwō
Kande FOC Binta 3.SG.F.COMPL slightly.exceed.A height
'It's Kande who is slightly taller than Binta.' (Newman, 2000)

Furthermore, the surface-phonological approach advocated for by Hayes (1990) is also highly limited in its scope: capitalising on a single exponent of direct object marking, namely final vowel shortening, Hayes assumes a basic bi-partition between C-frame and non-C-frames. However, once grades other than grade 1 are taken into consideration, it becomes immediately apparent that the distinction is at least tri-partite, featuring different patterns of syncretism, depending on the paradigm: while verbs in grade 2 provide distinct forms for all three basic frames (A vs. B vs. C), forms in grade 6 are fully syncretic (A/B/C). Partially neutralised grades include grade 1 (A/B vs. C) and grade 2 imperatives (A/C vs. B). Given the diversity of neutralisation patterns, it appears quite arbitrary to single out final vowel shortening.

Marking for local direct objects is by no means limited to verbal paradigms, but appears to be a more general property of Hausa morphosyntax. Hausa continuative aspect markers, as opposed to all other tense, aspect mood markers select a gerund, instead of a standard verb. Morphologically, Hausa gerunds are subdivided into *`wā* and non-`wā verbal nouns. While the former mark the A-form by the characteristic floating tone-initial suffix *`wā*, but use standard verbal B- and C-forms, the latter mark the A-form by an allomorphic variant of the verb base, but affix the "genitive" linker *-n/-r* in the B and C-forms. Note that choice of the masculine vs. feminine form of the linker is determined by the inherent gender of the verbal noun, not that of its direct object. The nominal nature of non-`wā verbal nouns is further underlined by the property of taking pronominal objects from the possessive or genitive set, instead of the accusative.

Despite the fact that we are actually dealing with a clearly nominal form, we still find that presence of a local realisation of a direct object is signalled morphologically.

(4) a. Inȁ sàye-n rȉ̄gā
1.SG.CONTINUATIVE buy.VN.M-LINKER.C.M gown.F
'I am/was buying a/the gown.'

b. Inȁ sàye-n -sà
1.SG.CONTINUATIVE buy.VN.M-LINKER.B.M -3.SG.M.GEN
'I am/was buying it.'

c. rȉ̄gā nakȅ sàyē
gown 1.SG.CONTINUATIVE buy.VN.A
'It's a/the gown I am/was buying.'

[6] The oblique status of the standard of comparison in (3) is supported by independent evidence: while direct objects can passivise (in grade 7), the standard of comparison cannot (Newman, 2000, pp. 685–686). Furthermore, if both objects are realised in their canonical post-verbal position, the standard of comparison cannot intervene between the direct object and the governing verb.

While the syntactic conditions are essentially the same as for verbs, verbal nouns differ in the choice of exponent, i.e. a gender-differentiated "linker", and in the pattern of neutralisation (A vs. B/C).

The implementation in HaG follows quite closely my previous theoretical work (Crysmann, 2005). First of all, direct object marking is treated as an obligatory inflectional dimension (of verbs and nouns). Different morphological expression is effected by a set of 17 inflection rules that are conditioned on basic category (verb vs. noun), verbal paradigm (grade), and, of course the valence structure.

Reference to the valence lists, where subcategorisation information is represented in HPSG, not only permits to distinguish between transitives and intransitives, but also directly captures the parallelism between intransitives and extracted direct objects as far as direct object marking is concerned. In HPSG, the standard treatment of complement extraction is trace-less (Pollard and Sag, 1994, ch. 9), i.e., subcategorisation requirements for local complements are removed from the head's valence list by means of a lexical rule which places the relevant information onto the non-local feature SLASH instead from where it percolate up the tree until a suitable filler is found. Thus, non-locally realised valents will already be removed from the valence list at the point where the head is inflected for direct object marking.

4.2. Mixed categories

It is a characteristic feature of Hausa that the language makes quite extensive use of the category *noun*: besides common nouns and proper nouns, which typically denote individuals, we also find dynamic or action nouns as well as verbal nouns, all of which denote events. Furthermore, it has been claimed repeatedly that a distinct category of adjectives is difficult to established on morphological grounds (see Newman, 2000 for discussion). Likewise, a great part of the inventory of functional prepositions display clearly nominal properties, leading Wolff (1993) to coin the term of "prepositional nouns".

While we clearly need to functionally distinguish adjectives, prepositions, and gerunds from ordinary nouns, the striking morphological parallelism militates in favour of an analysis that assigns all these items to the same basic category, i.e. *noun*. Building on previous work (Crysmann, 2011, to appear), I shall argue that a feature-based syntactic theory, such as HPSG, lends itself quite naturally to a treatment of mixed categories. Furthermore, once functionally diverse elements are conflated into a single category, it is imperative to rule out overgeneration. As we shall argue, a computational grammar can provide the necessary testing ground to establish this.

4.2.1. Adjectives

The relative unease regarding the postulation of a separate category of adjectives in Hausa relates mainly to the fact that adjectives tend to draw on the same morphological paradigms that are independently used for nouns. Furthermore, when used pre-nominally, adjectives are obligatorily inflected with the gender-differentiated "genitive" linker *-n/-r*, another property shared with nouns.

Functionally, Hausa adjectives are clearly modifiers, predicating a property of an individual. Similar to other modifiers in this head-initial language, all adjectives may appear in post-head position. Given their semantic and word order properties on the one side, and their morphosyntactic ones on the other, we can conclude that Hausa adjectives are best analysed as inherently modifying nouns. In HPSG, modifier status is represented by a selectional feature MOD by which an adjunct selects the properties of a head it combines with. Thus, the difference between ordinary nouns and inherently modifying nouns (vulgo: adjectives) can be represented by reference to the MOD feature.

As detailed in Figure 6, an adjective like *ƙàramī* 'small' selects the nominal head it modifies via the feature MOD. The semantic effect of modification is represented by identifying the first argument (ARG1) of the *small* relation with the referential INDEX of the head noun, selected via MOD.

In addition to the canonical post-head position, adjectives can optionally appear pre-nominally. In this case, the adjective obligatorily gets inflected with the linker *-n/-r*, which is, by the way, illicit with post-nominal adjectives. Since the presence vs. absence of the linker can be fruitfully analysed as the nominal counterpart of direct object marking and since true adjuncts typically follow their heads, rather than precede them, I have suggested (Crysmann, 2011) that pre-nominal adjectives are best understood

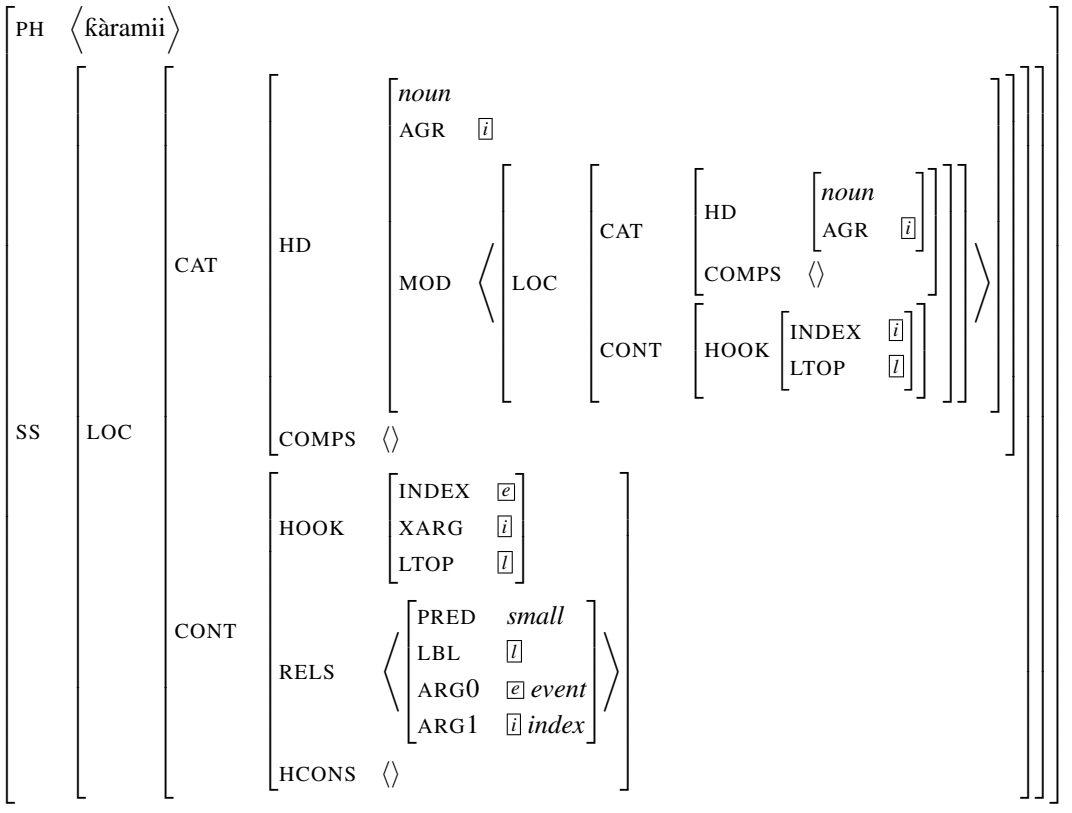

Figure 6: Lexical representation of adjectives

as heads that take as a complement the noun they modify. In essence, the adjective, which is a semantic modifier and a syntactic adjunct selecting the nominal head is turned into a head selecting the modified noun as a complement by means of a type-shifting lexical rule.

Since the semantic effect of modification is already fixed at the lexical level, type shifting only affects the mode of syntactic combination, i.e. the reversal of head–non-head (head–complement vs. head–adjunct) relations. If the semantics are the same for pre-head and post-head adjectives, HaG will actually generate both variants given a single input semantics, as shown in the screen shot in Figure 8.

Note further that the type shifting approach not only relates the surface order of pre-nominal adjectives to general word order properties of this head-initial language, but it also provides a direct account of the obligatory presence of the linker: if pre-nominal adjectives are indeed nouns that select the nominal head they modify as their first complement, they actually match the structural description for direct object marking.

4.2.2. Prepositions

A major subset of Hausa prepositions, the so-called "genitive prepositions" (Newman, 2000) or "prepositional nouns" (Wolff, 1993) behave morphosyntactically like nouns, taking the gender-differentiated linker *-n/-r*. Moreover, if their internal complement is pronominal, the affixal pronouns used are those from the genitive or possessive set.

(5) a. ciki -n gàrī
in.M -LINKER.M town
'in (the) town'

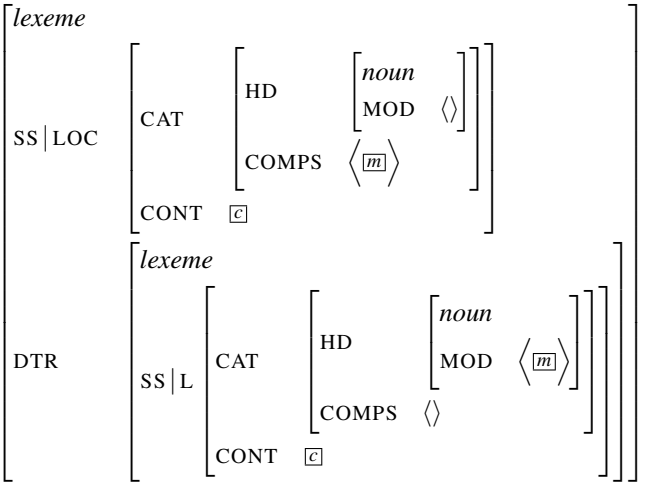

Figure 7: Lexical rule for pre-nominal adjectives

b. kàma-r jūji -n dūniyà̄
like.F-LINKER.F rubbish.heap.M -LINKER.M world
'like the filth of the earth'

Prepositional nouns, in contrast to true prepositions, permit extraction of their internal complement, giving rise to preposition stranding. If the complement of the prepositional noun is realised non-locally, however, use of the linker is illicit.

(6) a. àkwà̄tì dà yakè̄ ciki (*-n)
box REL 3.SG.M.CONTINUATIVE in
'the box it is in'

b. yanà̄ ciki (*-n)
3.SG.M.CONTINUATIVE in -linker.m
'He's inside.'

Likewise, if the prepositional noun is used intransitively, the linker cannot be used either. Thus, in addition to the morphological similarity between "prepositional nouns" and other nominal expression, we find that the structural conditions on direct object marking are entirely parallel to those observable for, e.g. verbal nouns.

Similar to the mixed category treatment of adjectives described in the previous section, I have argued in Crysmann (to appear) that Wolff's intuition of a mixed category can receive a formal interpretation which relates its nominal morphosyntactic properties to its categorial properties (*noun*), but captures the syntacto-semantic similarity to true prepositions in terms of semantic representation and syntactic combinatorial features (MOD).

As depicted by the sample lexical entry for *kàrƙashī* 'under' in Figure 9, the modifier-properties are represented by linking the external argument (ARG1) of the *under* relation to the head selected via MOD, and the internal argument (ARG2) to the NP complement, selected via COMPS.

The lexical representation of mixed category prepositional nouns is thus highly reminiscent of the one we suggested earlier for adjectives, the only difference being that prepositions lexically subcategorise for an internal complement, whereas adjectives generally do not do so. This difference in valency will also prove crucial for distinguishing these two categories: in essence, HaG's type shifting lexical rule requires the modifying noun it operates on to have an empty COMPS list, a requirement which is sufficient to block application of this lexical rule to prepositions.

With respect to direct object marking, inflection with the linker follows again directly from the lexical description given in Figure 9: unless the valency is lexically removed (intransitive use or extraction lexical rule), direct object marking is triggered, with exponence determined by basic category (*noun*).

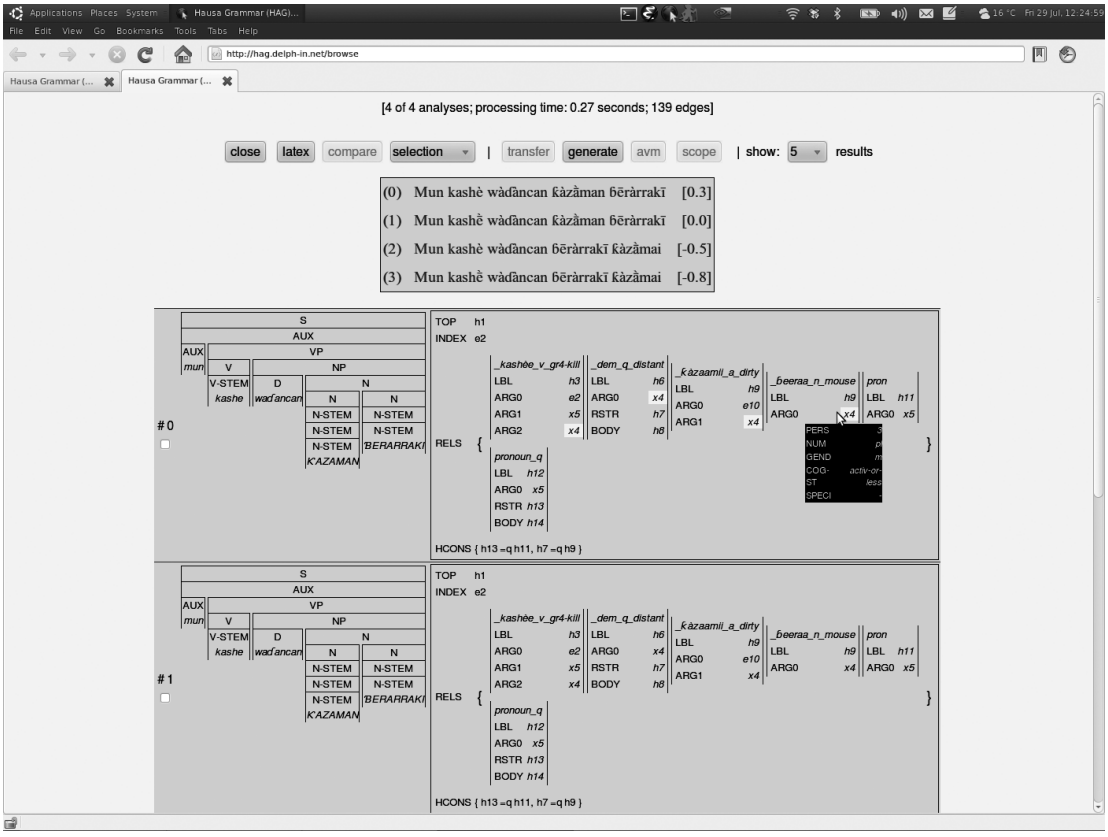

Figure 8: HaG generator output

To summarise, morphosyntactic similarity between nouns, verbal nouns, adjectives and prepositional nouns is modelled in HaG by means of subsuming these distinct descriptive categories under the unique basic category noun. Functional differences between these traditional categories are captured instead at the level of syntactic combinatory potential (valency) and inherent semantics (modifier vs. non-modifier).

4.3. Bound pronouns

Besides a set of independent pronouns, which is mainly used for obliques, pronominal arguments in prominent grammatical functions are realised by different sets of bound pronouns.

4.3.1. Pronominal subjects

As far as subject pronouns are concerned, there appears to be consent that these are best understood as agreement markers integrated with the system of TAM markers (Tuller, 1986; Newman, 2000; Jaggar, 2001). Accordingly, Tuller (1986) explicitly classifies Hausa as a null subject language. HaG essentially follows this line of analysis: TAM markers are treated as raising auxiliaries, i.e. auxiliaries which subcategorise for a VP complement, inheriting the VP's subject valency. As a consequence, agreement between the TAM marker and its subject is expressed as a lexical constraint. In order to license null subjects, the grammar provides a unary rule alongside the binary subject-head rule, in order to discharge subject valencies in the absence of an overt subject sister constituent.

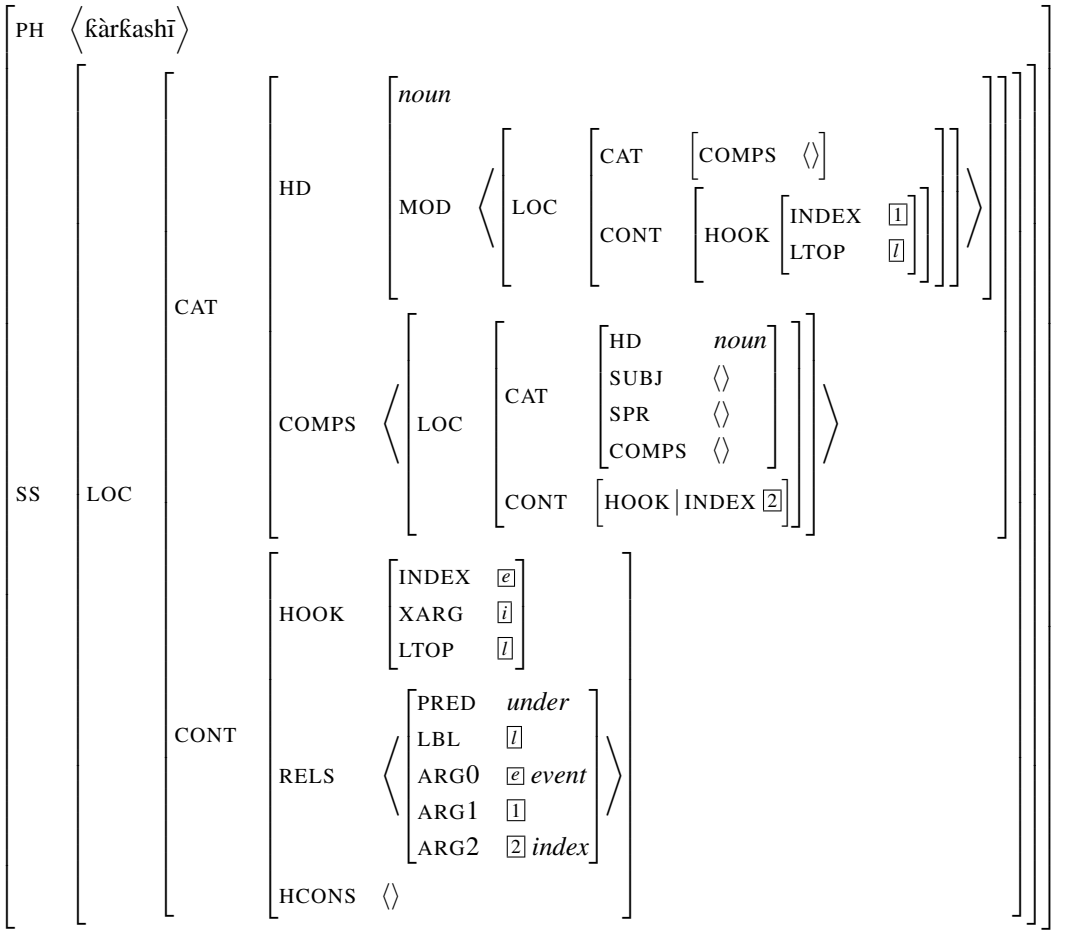

Figure 9: Lexical representation of prepositional nouns

4.3.2. Pronominal objects

Regarding direct object pronouns, Hausaists agree on the bound nature of these markers. In contrast to the traditional view according to which direct object pronouns constitute a single paradigm of polar tones, Newman (2000) and Jaggar (2001) argue for two distinct sets of direct object pronouns, distinguishing a "strong" high tone and a "weak" low tone paradigm. Choice between these two paradigms is considered to be morphologically conditioned, e.g., on the basis of verb grades. Neither high tone nor low tone direct objects can "be focused, conjoined, or contrastively stressed" (Jaggar, 2001, p. 406), all properties which suggest either clitic or pronominal affix status. As choice between pronominal paradigms is determined by the verb's paradigm membership, which is a purely morphological rather than morphosyntactic property, the principle of Lexical Integrity strongly suggests that these pronominal elements are indeed morphologically, not just phonologically bound elements.[7] Pronominal affix status is also supported by several other classic criteria (Zwicky and Pullum, 1983). First, the set of hosts to which these markers can attach is a circumscribed set of lexically determined elements (Criterion A). Furthermore both sets of direct object pronouns observe a strict adjacency condition regarding their hosts, a property which distinguishes them from independent pronouns and NPs in general: e.g., modal particles, which can otherwise be freely interspersed between syntactic words, and, for that matter between verbs and their complements, cannot intervene between the host and a direct object pronominal. Finally, since affixation, as compared to cliticisation is the typologically unmarked option

[7] See, e.g. Anderson (1992) for a discussion on morphological vs. morpho-syntactic features in the context of Lexical Integrity.

(Zwicky, 1985), I shall analyse Hausa direct object markers as morphologically bound affixes, rather than prosodic-ally attached clitics. This approach also concurs with the observation that Hausa is a predominantly head-marking, as opposed to argument marking language and that case distinctions are otherwise unattested in Hausa noun syntax.

Consequently, HaG treats direct object pronominals as affixes which are introduced by means of inflectional rules. Since the bound status of direct object pronominals is not reflected in the orthography, the grammar makes use of the token mapping functionality (Adolphs et al., 2008) provided by the Pet and ace parsers to introduce an *additional* hypothesis into the parse chart, where token boundaries between pronominal affix candidates and the host are removed.[8]

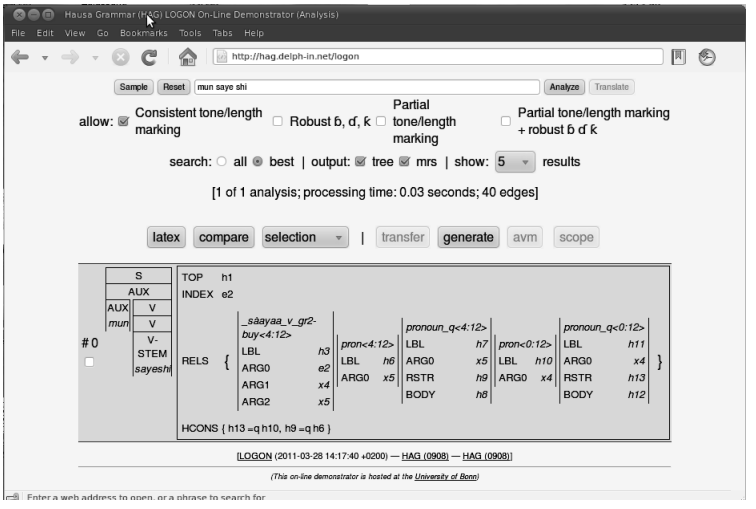

Figure 10: Tokenisation of pronominal affixes

In essence, the possibility of input chart mapping paves the way for a unified treatment of both nominal and verbal bound object pronominals independent of accidental orthographic conventions.

4.3.3. Pronominal indirect objects

The analysis of indirect object pronouns in Hausa is intimately linked to that of the indirect object marker *wà/mà*. Following Abdoulaye (1992), HaG treats *wà/mà* as an applicative verb that forms a verbal cluster with the main verb. Consequently, indirect object pronouns are analysed in terms of bound pronominal affixation to the applicative verb. The verb cluster analysis of the indirect object marker, as suggested by Abdoulaye (1992) not only provides an ingenious account of obligatory stranding with indirect object extraction, but it also predicts the lengthening of the final vowel of the marker in these contexts which can be observed in several dialects (see, e.g., Abdoulaye, 1992; Newman, 2000).

5. Unbounded dependencies

The range of constructions that feature unbounded dependencies (extraction) in Hausa is highly similar to that also found in European languages, such as English, and Afroasiatic languages, such as Hebrew or Arabic: typically, they involve wh-question formation and relativisation. In addition to these, Hausa also employs extraction for focus marking (*ex situ* focus), although it is generally recognised that Hausa also permits realisation of focused material *in situ* (Green and Jaggar, 2001; Hartmann and Zimmermann, 2007).

[8] As direct object pronominals are orthographically identical to TAM markers, it is imperative to be able to provide the parser with two alternative tokenisation hypotheses, i.e. one with the putative object marker suffixed to the preceding word, and another one, where the putative TAM marker is represented as a separate token.

Following the standard approach to long-distance dependencies in HPSG (Pollard and Sag, 1994; Sag, 1997; Ginzburg and Sag, 2001), argument extraction in HaG is effected by means of lexical rules which remove a head's valency and introduce its LOCAL value into the head's non-local feature SLASH. For adjuncts, SLASH introduction is performed by a unary syntactic rule, following Levine (2003). Once introduced into SLASH, non-local features are percolated until a filler is found. At the filler site, the non-local dependency is discharged, unifying the element in SLASH with the local value of the filler.

A peculiarity of Hausa is that the language marks binding of long distance dependencies by a filler (Tuller, 1986; Davis, 1986; Crysmann, 2005): in the completive and continuative, distinct paradigms are used depending on whether the clause contains a filler or not. In HaG, this alternation is implemented by means of a feature specification on filler-head structures and root nodes, constraining the TAM value of the head daughter.

A property that Hausa shares with distantly related Arabic and Hebrew (Sells, 1984) is the availability of resumptive pronoun strategies alongside gap strategies (Tuller, 1986). While argument extraction by means of filler–gap dependencies appears restricted to the arguments of verbs and nouns, complements of true prepositions can only be extracted, if a resumptive element is placed in situ (Newman, 2000; Jaggar, 2001). Genitive prepositions, however, permit both resumptive and gap strategies (cf. the examples in Figure 11): under the hypothesis made above that these elements are indeed nouns (Crysmann, to appear), we expect the internal argument of these prepositional nouns to extract, just like those of verbal nouns or dynamic nouns. This categorial restriction on the distribution of gaps is captured in HaG by an appropriate constraint on the complement extraction lexical rule.

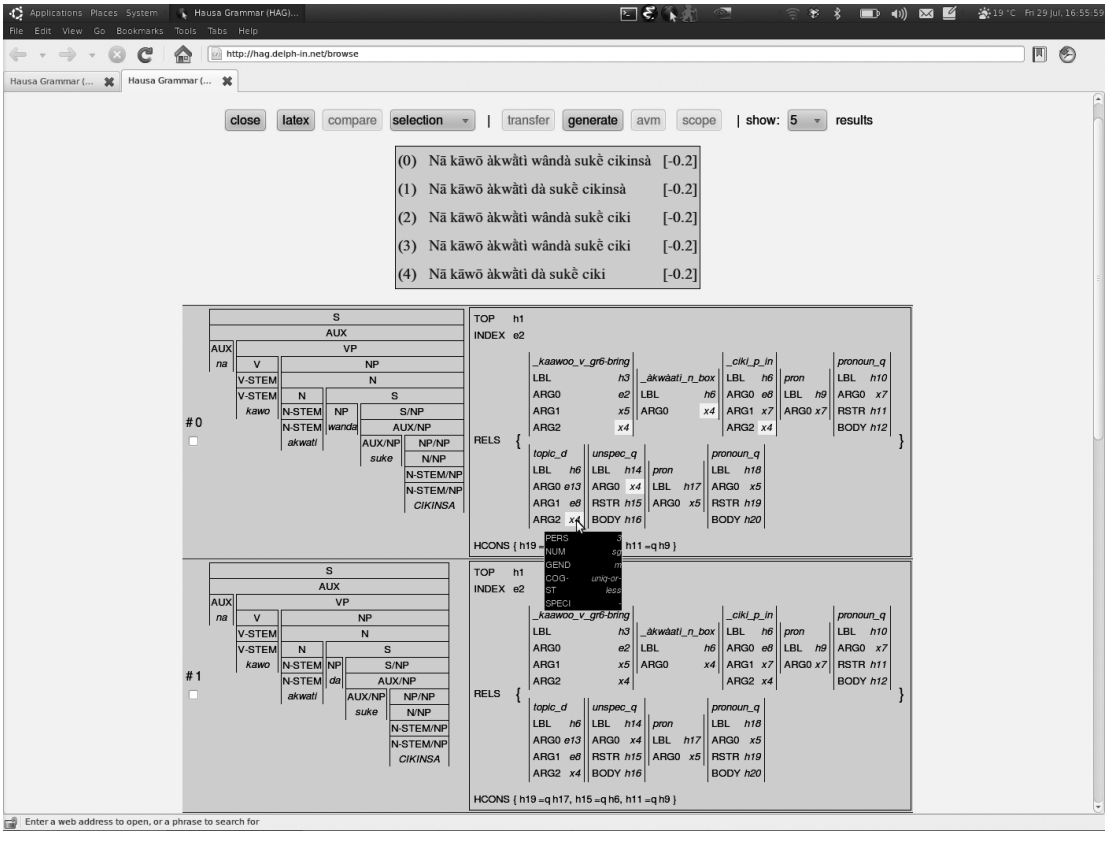

Figure 11: HaG generator output

As for resumptives, categorial restrictions do not appear to apply. In HaG, unbounded dependencies can be launched optionally, whenever a pronominal element is introduced. As illustrated in Figure 11,

HaG treats resumptive elements as true bound variables, assigning them the same semantics as in filler–gap dependencies. Consequently, the HaG generator may produce resumptive and filler–gap variants for a single semantic input structure.

In the current implementation, HaG does not yet license relativisation out of wh–islands: As discussed in detail by Tuller (1986) long relativisation out of wh–islands is indeed possible in Hausa, provided a pronominal is used at the gap-site. In contrast to all the other constructions discussed in this paper, which can be considered stable, the current treatment of resumption will be replaced in the near future by the unified approach to resumption and extraction developed in Crysmann (2012).

6. Conclusion and outlook

The present paper documents the major properties of HaG, an emerging computational grammar of Hausa. While development of the grammar is still ongoing, the analysis of the core phenomena as discussed here is essentially stable and can therefore serve for future reference. With an implementation of core syntax and morphology in place, future development will be focused on increasing the lexical coverage of the grammar, enabling us to evaluate the grammar on the basis of Hausa text corpora.

Before closing, I would briefly like to discuss to what extent the present grammar resource can be of interest to descriptive, theoretical, and computational linguists working on African languages. Finally, I shall point out a few long-term goals in the development of this grammar.

One of the main distinctive features of this grammar is the integration of morpho-syntax with grammatical and lexical tone, a property which is crucial for many African languages. To the best of my knowledge, this is the first-ever implemented grammar that systematically integrates these levels of description. Given the fact that HaG is based on the LinGo grammar Matrix (Bender et al., 2002), a starter kit for grammar development targeted at descriptive linguists and typologists alike, the implementation of suprasegmental phonology in HaG could be reused, tested and expanded to cover other tone languages, ultimately leading to a more refined computational model of grammatical tone in general.

In the area of modelling Hausa syntax and morphology, implementation has proven to be an indispensable tool for theoretic modelling, enabling us to systematically test the empirical repercussion of underspecified and mixed categories: the reversibility of the grammar, in particular, was instrumental for pushing towards a maximally generalised model of categorial identity, while at the same time controlling for over-generation. As a net effect, it was possible not only to give a formal interpretation of intuitions expressed in descriptive grammars, but also to push these intuitions towards their logical conclusion.

The grammar already comes with a test suite that is parsed with the current grammar and manually disambiguated, providing full syntactic and semantic annotation of Hausa core phenomena. As the grammar grows, the empirical coverage of this test suite will be expanded. Once we have expanded the lexicon of the grammar, we shall provide syntactically and semantically annotated and disambiguated treebanks of real-world Hausa texts that can be used for automatic disambiguation and corpus-based research alike.

Grammar development in HaG has so far been deliberately focused on grammatical rather than lexical coverage. A major research question will be how to bootstrap the lexicon from a grammar with high constructional coverage using current technology in (semi-)automatic deep lexical acquisition (e.g. Zhang and Kordoni, 2006). The valency information thus acquired will not only be available to the open-source grammar, but can, of course, be used independently.

Once lexical coverage has been extended, the grammar shall be put to use in a number of different application scenarios: first, a tutorial system for learners of Hausa that can generate grammatical exercises and check and correct students responses, based on the approaches developed in Crysmann, Bertomeu, Adolphs, Flickinger and Klüwer (2008). Second, the tone and length specifications could be reconstructed for standard orthography *boko* text, drawing on a combination of grammar-based symbolic constraints and statistical information; besides being useful for students of Hausa, this application has, of course, the further potential of text-to-speech synthesis. Third, the Hausa grammar, which is based on MRS, can be connected straightforwardly to the LOGON machine translation system (Oepen, Dyvik, Lønning, Velldal, Beermann, Carroll, Flickinger, Hellan, Johannessen, Meurer, Nordgård and Rosén, 2004), enabling us to develop formal models of contrastive linguistics, while at the same time implementing a machine-translation system.

To summarise, I have presented the major morphological, syntactic, and semantic properties of HaG, an emerging grammar of Hausa. While the grammar already has some interesting properties of its own, its integration into HPSG and MRS-based processing platforms opens up a broader universe of applications in the near future.

References

Abdoulaye, Mahamane L. 1992. *Aspects of Hausa Morphosyntax in Role and Reference Grammar*. Ph. D.thesis, SUNY Buffalo, NY.

Adolphs, Peter, Oepen, Stephan, Callmeier, Ulrich, Crysmann, Berthold, Flickinger, Dan and Kiefer, Bernd. 2008. Some Fine Points of Hybrid Natural Language Parsing. In *Proceedings of the 6th Conference on Language Resources and Evaluation (LREC 2008), May, Marrakesh*.

Anderson, Stephen R. 1992. *A–Morphous Morphology*. Cambridge Studies in Linguistics, Cambridge: Cambridge University Press.

Bender, Emily M., Flickinger, Dan and Oepen, Stephan. 2002. The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammar. In John Carroll, Nelleke Oostdijk and Richard Sutcliffe (eds.), *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, pages 8–14.

Callmeier, Ulrich. 2000. PET — A Platform for Experimentation with Efficient HPSG Processing Techniques. *Journal of Natural Language Engineering* 6(1), 99–108.

Carpenter, Bob. 1992. *The Logic of Typed Feature Structures with Applications to Unification-based Grammars, Logic Programming and Constraint Resolution*, volume 32 of *Cambridge Tracts in Theoretical Computer Science*. New York: Cambridge University Press.

Carroll, John, Copestake, Ann, Flickinger, Dan and Poznanski, V. 1999. An efficient chart generator for (semi-)lexicalist grammars. In *Proceedings of the 7th European Workshop on Natural Language Generation*, pages 86–95, Toulouse, France.

Carroll, John and Oepen, Stephan. 2000. High efficiency realization for a wide-coverage unification grammar. In *Proceedings of the 2nd International Joint Conference on Natural Language Processing (IJCNLP)*, pages 165–176, Jeju, Korea.

Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. Stanford: CSLI Publications.

Copestake, Ann and Flickinger, Dan. 2000. An open-source grammar development environment and broad-coverage English grammar using HPSG. In *Proceedings of the Second conference on Language Resources and Evaluation (LREC-2000)*, Athens.

Copestake, Ann, Flickinger, Dan, Pollard, Carl and Sag, Ivan. 2005. Minimal Recursion Semantics: an introduction. *Research on Language and Computation* 3(4), 281–332.

Crysmann, Berthold. 2005. An Inflectional Approach to Hausa Final Vowel Shortening. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 2004*, pages 73–112, Kluwer.

Crysmann, Berthold. 2009. Autosegmental Representations in an HPSG for Hausa. In *Proceedings of the ACL-IJCNLP workshop on Grammar Engineering Across Frameworks (GEAF 2009)*, ACL.

Crysmann, Berthold. 2011. A unified account of Hausa genitive constructions. In Philippe de Groote, Markus Egg and Laura Kallmeyer (eds.), *Formal Grammar. 14th International Conference, FG 2009, Bordeaux, France, July 25-26, 2009, Revised Selected Papers*, volume 5591 of *Lecture Notes in Computer Science*, Springer.

Crysmann, Berthold. 2012. Resumption and Island-hood in Hausa. In Philippe de Groote and Mark-Jan Nederhof (eds.), *Formal Grammar. 15th and 16th International Conference on Formal Grammar, FG 2010 Copenhagen, Denmark, August 2010, FG 2011 Lubljana, Slovenia, August 2011*, volume 7395 of *Lecture Notes in Computer Science*, Springer.

Crysmann, Berthold. to appear. On the Categorial Status of Hausa Genitive Prepositions. In Bruce Connell and Nicholas Rolle (eds.), *Proceedings of the 41st Annual Conference on African Linguistics (ACAL 41), Toronto, May 2010*, Somerville, MA: Cascadilla Press.

Crysmann, Berthold, Bertomeu, Núria, Adolphs, Peter, Flickinger, Dan and Klüwer, Tina. 2008. Hybrid processing for grammar and style checking. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 153–160, Manchester, UK: Coling 2008 Organizing Committee.

Davis, Anthony. 1986. Syntactic Binding and Relative Aspect Markers in Hausa. In *Proceedings of the Fifteenth Annual Conference on African Linguistics, Los Angeles, CA, 1984*.

De Pauw, Guy, Wagacha, Peter W and De Schryver, Gilles-Maurice. 2007. Automatic diacritic restoration for resource-scarce languages. In V Matousek and P Mautner (eds.), *LECTURE NOTES IN ARTIFICIAL INTELLIGENCE*, volume 4629, pages 170–179, Springer Verlag Berlin.

Ginzburg, Jonathan and Sag, Ivan. 2001. *Interrogative Investigations: the Form, Meaning and Use of English Interrogatives*. Stanford: CSLI publications.

Goldsmith, John A. 1976. *Autosegmental Phonology*. Ph. D.thesis, MIT.

Green, Melanie and Jaggar, Philip. 2001. Ex-situ and in-situ focus in Hausa. Cognitive Science Research Papers 527. School of Cognitive and Computing Sciences, University of Sussex.

Hartmann, Katharina and Zimmermann, Malte. 2007. In Place — Out of Place? Focus in Hausa. In K. Schwabe and S. Winkler (eds.), *On Information Structure, Meaning and Form: Generalizing Across Languages*, pages 365–403, Amsterdam: Benjamins.

Hayes, Bruce. 1990. Precompiled Phrasal Phonology. In Sharon Inkelas and Draga Zec (eds.), *The Phonology-Syntax Connection*, pages 85–108, University of Chicago Press.

Jaggar, Philip. 2001. *Hausa*. Amsterdam: John Benjamins.

Jungraithmayr, Herrmann, Möhlig, Wilhelm J. G. and Storch, Anne. 2004. *Lehrbuch der Hausa-Sprache*. Köln: Rüdiger Köppe Verlag.

Krieger, Hans-Ulrich. 1996. *TDL — A Type Description Language for Constraint-Based Grammars*, volume 2 of *Saarbrücken Dissertations in Computational Linguistics and Language Technology*. Saarbrücken: DFKI GmbH.

Leben, William. 1973. *Suprasegmental Phonology*. Ph. D.thesis, MIT.

Levine, Robert D. 2003. Adjunct valents: cumulative scoping adverbial constructions and impossible descriptions. In Jongbok Kim and Stephen Wechsler (eds.), *The Proceedings of the 9th International Conference on Head-Driven Phrase Structure Grammar*, pages 209–232, Stanford: CSLI Publications.

Newman, Paul. 2000. *The Hausa Language. An Encyclopedic Reference Grammar*. New Haven, CT: Yale University Press.

Newman, Paul and Ma Newman, Roxana. 1977. *Modern Hausa–English Dictionary*. Ibadan and Zaria, Nigeria: University Press.

Oepen, Stephan and Carroll, John. 2000. Ambiguity packing in constraint-based parsing - practical results. In *Proceedings of the 1st Conference of the North American Chapter of the Association for Computational Linguistics*, pages 162–169, Seattle, WA.

Oepen, Stephan, Dyvik, Helge, Lønning, Jan Tore, Velldal, Erik, Beermann, Dorothee, Carroll, John, Flickinger, Dan, Hellan, Lars, Johannessen, Janne Bondi, Meurer, Paul, Nordgård, Torbjørn and Rosén, Victoria. 2004. Som å kapp-ete med trollet? Towards MRS-Based Norwegian – English Machine Translation. In *Proceedings of the 10th International Conference on Theoretical and Methodological Issues in Machine Translation*, Baltimore, MD.

Parsons, Fred W. 1960. The Verbal System in Hausa. *Afrika und Übersee* 44, 1–36.

Pollard, Carl and Sag, Ivan. 1987. *Information–Based Syntax and Semantics*, volume 1. Stanford: CSLI.

Pollard, Carl and Sag, Ivan. 1994. *Head–Driven Phrase Structure Grammar*. Stanford: CSLI and University of Chicago Press.

Sag, Ivan. 1997. English Relative Clause Constructions. *Journal of Linguistics* 33(2), 431–484.

Sells, Peter. 1984. *Syntax and Semantics of Resumptive Pronouns*. Ph. D.thesis, University of Massachusetts at Amherst.

Toutanova, Kristina, Manning, Christopher D., Shieber, Stuart M., Flickinger, Dan and Oepen, Stephan. 2002. Parse Disambiguation for a Rich HPSG Grammar. In *Proceedings of the First Workshop on Treebanks and Linguistic Theories (TLT2002)*, pages 253–263, Sozopol, Bulgaria.

Tuller, Laurice A. 1986. *Bijective Relations in Universal Grammar and the Syntax of Hausa*. Ph. D.thesis, UCLA, Ann Arbor.

Velldal, Erik and Oepen, Stephan. 2005. Maximum Entropy Models for Realization Ranking. In *Proceedings of the 10th MT-Summit (X), Phuket, Thailand*.

Wolff, Ekkehard. 1993. *Referenzgrammatik des Hausa*. Münster: LIT.

Zhang, Yi and Kordoni, Valia. 2006. Automated deep lexical acquisition for robust open text processing. In *Proceedings of the Fifth International Conference on Language Resourses and Evaluation (LREC 2006), Genoa, Italy*.

Zwicky, Arnold M. 1985. Clitics and Particles. *Language* 61, 283–305.

Zwicky, Arnold M. and Pullum, Geoffrey K. 1983. Cliticization vs. Inflection: English *n't*. *Language* 59, 502–513.

Selected Proceedings of the 42nd Annual Conference on African Linguistics: African Languages in Context

edited by Michael R. Marlo, Nikki B. Adams, Christopher R. Green, Michelle Morrison, and Tristan M. Purvis

Cascadilla Proceedings Project Somerville, MA 2012

Copyright information

ISBN 978-1-57473-453-9 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Crysmann, Berthold. 2012. HaG — A Computational Grammar of Hausa. In *Selected Proceedings of the 42nd Annual Conference on African Linguistics*, ed. Michael R. Marlo et al., 321-337. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2780.