Pooling resources: Afranaph and SSWL/Terraling

This presentation addresses the question how to best combine the resources of the Afranaph project and the SSWL/Terraling database project. It will make a concrete proposal for a sister project on documenting the internal structure of the noun phrase in African languages, which will eventually be part of a broader SSWl project, and propose the develop a detailed questionnaire mixing SSWL and afranaph methodology, and allowing automated extraction of the data.

SSWL (The Syntax of the World's languages http://sswl.railsplayground.net/) is a a community based, open ended, expert crowdsourced database, that, in the long run, aims to document the syntactic properties of the world's languages, at a level of garnularity not previously undertaken. SSWL was conceived of and developed by Chris Collins and Richard Kayne, in collaboration with Dennis Shasha, professor of computer science at NYU and architect of the database. Shasha's research interests include computational biology, large data and pattern matching, and machine learning. SSWL comes with flexible and powerful search and mapping tools set up to allow pursuing any research questions. Terraling (http://terraling.com) is the next generation of the project (same database, better platform). It is a single database that can be partitioned into groups, one of which will host SSWL. It provides a flexible platform for linguists, which can tailor their projects as they desire, (with an option to make the group private or public), giving full access to the search and mapping tools and to other groups (groups can be combined). In my presentation, I would like to show how we can use this platform for the Afranaph project, while keeping the integrity of the afranaph and existing database intact. Concretely, I propose a sister project on the systematic documentation targeting the internal structure of noun phrase in the languages of Africa, and the development of highly structured step-by-step questionnaire to generate the data for the project, using lessons from SSWL's property definitions, and lessons from working with native speakers from the aphranaph project. The questionnaire must technically be set up in such a way as to allow extraction of the data into a group of the terraling database, after data have been verified. Entry as a separate page of the terraling project (and perhaps also in the afranaph database) will give access to the use of the powerful mathematical search tools of SSWL/Terraling which allow exploring and testing theoretical predictions about expected patterns, gaps, and correlations. This sister project complements existing Afranaph sister projects, and is a self-contained part of the broader SSWL documentation project of the internal structure of the noun phrase cross linguistically.

The theoretically driven semi-automated questionnaires should also directly serve as a template to generate theoretically inspired descriptive papers, theses, ad could be more broadly adapted to other projects.

Besides the project's inherent interest (the systematic documentation of specific properties of the structure of the noun phrase in the languages of Africa), the specific research questions of the project derive from the theoretical

research around modeling Greenberg's (1966) Universal 20 (cf Cinque 2005), which in turn directly connect to my own research over the past 15 years (Koopman & Szabolcsi 2000). The project aims to answer broad and more specific questions listed below.

- Is there a unique universal order of merge within the Noun Phrase from which all surface orders can be derived? One such proposed (small) fragment of the noun phrase can be found in (1) (cf Cinque, 2005) (brackets omitted) :

  (1) $RC_{nonrestr}$.. Q_{univ} ..Dem ..Num_{ord} RC_{restr}.. Num_{card} .. A .. NP

- A more specific question: how do other parts of the noun phrase (plural marking, adjectival hierarchies, compounding, inalienable and alienable possession, encoding of (in)definiteness fit into this fragment?

I will pursue answers to these questions guided by very specific theoretical predictions, which will structure both the questionnaire and the research. These predictions derive from Cinque (2005) modeling of the observed data patterns known as Greenberg's U20 (the order of demonstratives, numerals, A and N is invariant before the N, but varies postnominally). As Cinque proposes, the attested and unattested orders in the domain of U20 (and many other domains since) can be explained if these orders are derived from an invariant right branching hierarchical structure, with all surface patterns derived by leftward movement of a constituent containing the Noun, which may pied-pipe dependents on its way up the nominal spine or not. [1] Languages vary as to how high this NP constituent moves up in the nominal spine. This proposal predict that prenominally, the order of merge will be invariant (there is no movement which induces reordering, while post nominally, much greater variability, is expected, because of the movement of the N and possible pied-piping.

An immediate question is whether this left-right asymmetry is indeed confirmed in African languages. Here, languages in and to the West of Cameroon/Nigeria will play a particularly important role, as we find languages with prenominal adjectives, numerals, noun class suffixes, Gen N orders, head final compounds, etc. The afranaph community will be an invaluable asset for this project. Systematic data from these languages should also provide important information about historical change. Since many African languages are known to be "heavy pied-piping" languages (cf Nkemnji 1996) we can expect privileged insights into its formal properties from this particular project.

As a concrete example of how we plan to investigate more specific theoretical questions listed above, consider the question how nominal plural marking fits into the hierarchical fragment given in (1)b. If we assume that nominal

---

[1]The issue is not whether head movement exists as a theoretical option, but whether we can model the comparative picture if head movement is an instance of movement of a tiny phrase which does not trigger pied-piping

morphology spells out the semantic plural head, we can ask the question if plural is merged below direct modification As, above direct modification A, or somewhere between direct modification As.

(2) Num_{card} $\boxed{?\text{PL}}$ .. A .. $\boxed{?\text{PL}}$ NP

Does UG provide a unique solutions? Or are all options attested? Here is a preliminary case study that shows how to pursue this question theoretically.

There is considerable evidence that English plural is merged higher than A N, i.e. the order of merge is $Pl > A* > N$. If this is the only order of merge UG allows (certainly the most restrictive option), the Cinquean program makes the following predictions about possible and excluded surface orders.

Expected orders under a Merge hierarchy $Pl = 1 > A* = 2 > N = 3$:
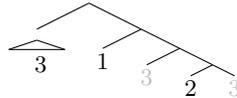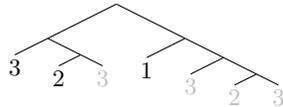
a.   1 2 3 (no reordering)⇝ surface order Pl A N)

```
    /\
   1  /\
     2  3
```
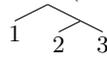
b.   132: Leftward movement of 3(=N) past 2 ⇝ surface order Pl N A

```
    /\
   1  /\
     3  /\
        2  3
```
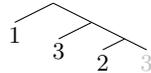
c.   312: leftward movement of 3 (past 1) ⇝ surface order N Pl A

```
      /\
     /\   \
    3   1  /\
          3  /\
            2  3
```

d.   321: Leftward movement 32 ("pied piping past 1 ⇝ surface order N A Pl)

```
        /\
       /\    \
      3 /\ 3   1  /\
        2          3  /\
                      2  3
```

e.   231: Leftward movement of 23 past 1 ⇝ surface order A N Pl)

```
      /\
     /\  \
    2  3  1  /\
             2 3
```
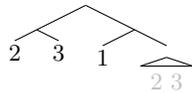
Expected gap the 213 order (i.e. the order A pl N, with plural dependent on the N) is excluded by the theory; this order can only arise by moving 2 (the A) without the N =3 to the left of Plural. Do we find such cases or not? Are they superficial counterexamples or not? Superficially, this pattern is attested in Nweh (Nkemnji 1996), who shows however that the Plural/ class marker depends on the presence of a silent N that occurs with the A.

The following table summarizes the patterns that are expected to occur, and list potential languages that potentially illustrate these particular patterns. The boxed african languages, Tuki and Nawdm are discussed below, with current data from SSWL.

Patterns expected to occur ✓; and predicted gaps 0)

123	Pl A N	✓	Tuki? Shupamem?
132	Pl N A	✓	Tuki
312	N Pl A	✓	Romance?
321	N A Pl	✓	Nawdm (Gur) , Vata
231	A N Pl	✓	English..
213	A Pl N 0?		

Tuki and Nawdm

Tuki (Tokombo) (SSWL data from Edmond Biloa) shows a somewhat unexpected pattern of plural noun class distribution. Noun class morphemes precede both prenominal adjectives and the noun. Postnominal adjectives are *invariant*, and do not show any plural marking, but postnominal numerals do.

Let us assume that the order of Merge is $Pl > A* > N$, as in English. Post nominal adjectives (i.e. color adjectives which are low in the adjectival hierarchy) must involve movement of the noun up in the adjectival hierarchy, leading to (3). Thus, the plural preceding pronominal A could simply be a spell out of the semantic plural (instead of an "agreement" with the N).

(3) $Pl > A > N > A > N$.

The N is also preceded by a plural: which at present could be analyzed in various perhaps equivalent ways (probing, selection by Pl of a u Pl (or noun class/gender) head, which in turn selects for N, agreement with a silent subject of the nominal predicate as in Koopman 05)). Whatever the right formal account, the merge structure in (3) provides an excellent start for an explanation for why post nominal adjectives are not marked for plural: plural cannot see further than the surface position of the N. Postnominal numerals must agree. This will follow from the fact that Numerals are merged above Pl:) $Num > Pl > A > N > A > N$, and the plural constituent pied-pipes to the left of Num, triggering the "upwards" (i.e. Spec Head) agreement so characteristic of the languages in the region. Given this structure, if the broad lines of this analysis are on the right track, we make further predictions. We expect to find languages without plural on the N, but with Pl preceding prenominal adjectives (cf (3)). This pattern is perhaps found in Shupamem. This state of affairs is in fact found in Nawdm (Gur), after reordering. In Nawdm, adjectives follow the noun, but *precede* the noun class marker.

Adjectives line up according to the order of merge of adjectives ($size > color$), which in the Cinquan theory can only be analyzed as a case in which the N(P) moves to the left of the adjectives, stranding the lower adjective:.

(4)   dʒèd bóbók hôn꞉ dé
        chair tall black CM(de)
        the tall black chairs.

Plural merges with this constituent which pied pipes to merge with Pl. As in Tuki adjectives will not agree because they are never local to Pl. As in Tuki, numerals must agree for plural.
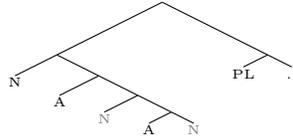
Thus, Nawdm and Tuki (and English) could have an identical order of Merge, and differences between the languages depend on how high the NP moves up into the nominal spine, and cyclic derivations, which shield the adjective from agreement. Whether this analysis can be independently supported, and alternatives ruled out, will of course depend on what the empirical picture turns out to be. This much is sure however, in order to answer such questions, we need both fine-grained data, and ways to generate them, databases to store the properties, powerful extraction and visualization tools to verify predictions, correlations, and a local community which collaborates towards the common goal of scientific progress. Pooling resources between Afranpah and SSWL/Terraling could be an important step forward.

Selected References

Biloa,Edmond, Tuki (Tokombo) http://sswl.railsplayground.net/ accessed on 5/25/2013. Albro, Daniel (1998). Cinque, G. 2005. Deriving Greenberg's universal 20 and its exceptions, *Linguistic inquiry*, 36. 3. 315-332.

Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of language*, ed. by Joseph Greenberg, 73113. Cambridge, Mass.: MIT Press.

Kayne, Richard S. 1994. The Antisymmetry of Syntax. Cambridge, MA: MIT Press.

Koopman, Hilda (2005) On the parallelism of DPs and clauses: evidence from Kisongo Maasai. in Carnie, 2005.

Koopman, Hilda and Anna Szabolcsi 2000. Verbal Complexes. Cambridge, MA: MIT Press.

SSWL data from Nawdm (Albro, 1998). http://sswl.railsplayground.net/ accessed on 7/25/2013.