Information Structure & Peripheries in Zaar

Bernard Caron

▶ **To cite this version:**

Bernard Caron. Information Structure & Peripheries in Zaar. 2016. hal-01359787

Information Structure & Peripheries in Zaar[1]

Bernard CARON

Llacan (Université Sorbonne Paris Cité, CNRS)

ABSTRACT

In this paper analysing peripheries in relation with syntax and information structure in Zaar, a Chadic language spoken in Nigeria, we have argued for a minimal annotation representing in a simple and concise way the interface between information structure and syntax The article uses the concept of macrosyntax, based on illocutionary units, for this new level of annotation using existing morphosyntactic tiers in Elan. The annotation system has been chosen, and a corresponding set of tags developed, bearing in mind that they should be as theory-neutral as possible in order to implement a genuine bottom-up heuristic methodology. The main asset of this system of annotation lies in the notion of stacks it uses to account for disfluencies, discontinuities and ellipses, and represent the oral discursive flow. With the corresponding annotation script, a pilot 90 min (15,000 words) corpus has been annotated to run a preliminary study of peripheries. We have argued that, although topics and frame-setters share the same intonation pattern, their syntactic properties call for a specific syntactic representation within the frame of Universal Dependency Grammar.

This article argues that morphosyntactic glossing of oral corpora is not sufficient for languages with little morphology. A minimal annotation system must be introduced to represent in a simple and concise way the interface between information structure and syntax. The article uses the concept of macrosyntax, based on illocutionary units, for this new level of annotation using existing morphosyntactic tiers in Elan. With the corresponding annotation script, a pilot 90 min (15,000 words) corpus has been annotated for Zaar, a Chadic language spoken in Nigeria and a preliminary study of peripheries in this language has been done on this annotated corpus.

The article is organized as follows: the first section presents the argument for macrosyntactic glossing; the second section introduces the annotation system; the third section presents a typology of peripheries in Zaar; and the final section characterizes the contrast between left-dislocation, and clefting using a microsyntactic representation developed within the dependency framework.

1   The case for macrosyntax

The distinction between micro- and macrosyntax was first proposed by (1990)Blanche-Benveniste *et al.* (1990), Berrendonner (1990), and Cresti (2000) (but see also (Andersen & Nølke 2002) for an overview). These studies put forward macrosyntax as a level of linguistic description capable of accounting for a number of cohesion mechanisms that are particularly frequent in spontaneous spoken language – especially in spoken French and Italian– which cannot be simply regarded as microsyntactic government phenomena, such as, for example, the "paratactic" construction in (1):

(1)   [ceux qui sont en location] [la moyenne] [c'est environ trois ans]

[those who are on a lease] [the average] [it's about three years]

'Those who are on a lease stay three years on average' (Rhaps-D0004 CFPP2000)

The same type of phenomenon is frequent in Zaar too, e.g. in (2):

(2)   [yàːʃí mə́nːːː malâːri máːːː] [lə̆pm zaːr máː] [tá tûːr gyáː dũ̂ː] [tá tûːr gyáː náɣat][2]

[The people… of Malar too...] [at the Angas festival too] [they would brew

beer] [they would cook food]

'The people of Malar, they would brew beer and cook food at the Angas

festival too.' (Cal_Har_045)

While the different macrosyntactic models acknowledge that sequences such as (1) and (2)

have to be considered as forming a cohesive unit at some level of linguistic description, they

diverge slightly as far as the characterization of the nature of this cohesion is concerned.

Macrosyntactic models characterize some major linguistic units that go beyond government

proper and are usually described in the literature from a pragmatic perspective that focuses on

their illocutionary or rhetorical values. Macrosyntax, instead, focuses on the span and the

---

[2]   Zaar is transcribed using the International Phonetic Alphabet, except for /j/ which is transcribed /y/. Vocalic phonemic length is marked after the vowel by single colon (ː). Phonetic length (in fillers, emphasis, etc.) is marked with three colons (ːːː). Phonemic tone is marked with diacritics: á, à, â and ǎ for High, Low, Falling and Rising respectively. Mid tone is left unmarked. The following abbreviations are used in morphosyntactic transcriptions: 1, 2, 3, Person; OBJ, Object; ADV, Adverb; AOR, Aorist; ASP, Aspect; BEN, Benefactive; COMP, Complementiser; COND, Conditional; CONJ, Conjunction; COP, Copula; CPL, Completive; DEF, Definite; DEICT, Deictic; DET, Determiner; DIST, Distal; DM, Discourse Marker; EXCL, Exclamative; FCT, Factual; FILL, filler; FUT, future; ICPL, Incompletive; IDP, Independent; INCH, Inchoative; ITER, Iterative; LOC, Locative; N, Noun; NEG, Negation; NMLZ, Nominalizer; OBJ, Object; PFV, Perfective; PL, Plural; PN, Person-Number; POS, Possessive; POSL, Possessive link; PRO, Pronoun; PROX, Proximate; PTCL, partiCle; QLT, Quality; QUEST, Question; REL, Relative marker; REM, Remote; RES, Resultative; SBJV, subjunctive; SG, singular; SPCF, specifier; SYNT, syntactic; TAM, tense-aspect-mode; TOP, topic; V, verb; VRT, Virtual.

form of macrosyntactic units, using syntactic and distributional criteria (such as suppressions, insertions, commutations) to identify and delimit them. For all the macrosyntactic models, the main identifying criterion of a macrosyntactic unit is the possibility that this unit has to constitute an autonomous utterance.

Since the practical objective is to create a corpus that allows us to study the interface between prosody and syntax, we need to clearly separate these two levels of analysis. Following the methodology first used in Rhapsodie, I have decided not to rely on prosodic criteria to define macrosyntactic units. Therefore I do not follow the prosodic definition of macrosyntactic units proposed by Berrendonner (2011) who describes the maximal extension of a macrosyntactic unit in terms of the presence of a conclusive intoneme; nor could I strictly follow the Florence school's approach (Cresti & Moneglia 2005) that characterizes macrosyntactic units as sequences of prosodic, rather than syntactic, units.

Rather, I consider that macrosyntax describes the whole set of relations holding between the microsyntactic units that make up one and only one illocutionary act, although microsyntax can sometimes go beyond macrosyntactic units. This definition combines the syntactic model proposed by the Aix model (Blanche-Benveniste *et al*. 1990), according to which the minimal units that compose a macrosyntactic unit are syntactic in nature, and the pragmatic model developed by the Florence model (Cresti & Moneglia 2005), according to which the maximal extension of a macrosyntactic unit is defined in terms of illocution.

Such a choice led us to call the maximal macrosyntactic units *Illocutionary Units* and to provide, in our work, an account and an annotation for the syntactic rather than the prosodic units that compose them.

1.1   The Rhapsodie framework

The macrosyntactic framework developed in the ANR Rhapsodie project (*Corpus de français parlé annoté pour la prosodie et la syntaxe* (Lacheret, Pietrandrea & Tchobanov 2014)) has proved to be particularly efficient in dealing with the specificities of oral corpora, e.g. piles stacking, disfluencies, repetitions, discourse markers, overlaps, co-enunciation, false starts, self-repairs and truncations. This method is data-driven, inductive (the relevant units are identified through annotation) and modular. Rhapsodie annotates two levels of syntactic cohesion: microsyntax, i.e., syntactic cohesion guaranteed by government and macrosyntax, i.e. syntactic cohesion guaranteed by illocutionary dependency.

The macrosyntactic level describes the whole set of relations holding between all the segments that make up one and only one illocutionary act. A macrosyntactic punctuation will mark macro-syntactic boundaries (i.e. illocutionary units and their main components: nuclei, pre nuclei and post nuclei, including discourse markers) and limits between pile layers (disfluencies, reformulation, coordination).

Each text is segmented into a string of illocutionary units (henceforth IlU); each IlU is composed of 3 kinds of components: a nucleus (obligatory), pre-nuclei (optional) and post nuclei (optional); see below: (1) and (2), where '<' follows a pre-nucleus and precedes a nucleus or another pre-nucleus; '>' precedes a post-nucleus and follows a nucleus or a previous post-nucleus; and '//' indicates the right boundary of a IlU (nuclei are in bold).

[context : two adolescent girls talk about clothes, boyfriends and (a little) about school]

(3)   *ndà:də̂m má: < **má ɫə́ yéltə̀** > 'áy' //*
        'Ndadem too < *I will go and see him* > eh. //' (Girls_A_005)

(4)   *fé:lêks < kyâ:n má: < **ká: rigá kə yisə́n tí:** > 'éy' //*
        'Felix < you too < *you know him* > eh. //' (Girls_B_092)

(5)   *sə̀ké:dî < **à: ná:y ɗàrí lim** // kó: //*
        'A skirt < *it reaches six hundred* // doesn't it? //' (Girls_A_30)

5

Heuristically, in order to identify Illocutionary Units (IlU) and their constituents (IlC), annotators rely on intonation cues perceived while listening to the data that is annotated. Perceptively relevant prosodic cues enable them to identify terminal and non-terminal breaks, the former constituting the IlU limits. The is defined as follows by Cresti & Moneglia (2005:17)

> *"The various kinds of prosodic breaks are conceptualised and defined as follows:*
>
> a. Prosodic break*: perceptively relevant prosodic variation in the speech continuum such as to cause the parsing of the continuum into discrete prosodic units.*
>
> b. Terminal prosodic break*: given a sequence of one or more prosodic units, a prosodic break is considered terminal if a competent speaker assigns to it, according to his perception, the quality of concluding the sequence.*
>
> c. Non-terminal prosodic break*: given a sequence of one or more prosodic units, a prosodic break is considered non-terminal if a competent speaker assigns to it, according to his perception, the quality of being non-conclusive."*

The basis prosodic distinction in Zaar is between pre-nucleus units whose boundary is characterized by a level intoneme followed by an initial step-up (pitch reset) at the onset of the following unit; and final prosodic breaks signalling the end of an IlU by a fall. The final fall can be replaced by or combined with other intonemes (e.g. rise and high-rise) in case of emphasis or exclamation. (Caron 2015:17)

1.2   Nature of the nucleus

Nuclei are usually governed by a tensed verb or a copula. But not always. See (6) where a noun (*la:,* 'work') is governing the nucleus, and (7) where the whole nucleus of the second IlU is an exclamation (*kây*, 'hey').

(6)   *gí: < ŋa: **la: ɓastə //***

gí: ŋa: la: ɓas =tə

DIST   small   work   at   3SG
'That's easy for him.' (*lit*. 'that < small *work* at him //') (Girls_B_094)

(7)   *kúní: à: məs ɓasəm sò:séy // myâ:n kúmá < **kâ:y //***

kúni -í: à: məs ɓas =mə   sò:séy myá:ni kúmá kâ:y

boy -DIST 3SG.PFV die at   1SG.OBJ quite 1SG   also   EXCL
'That boy is dying for me. Myself, *I don't care*!'
(*lit.* 'as for me < *hey!'*) (Girls_B_087)

1.3   Piles / stacking

The concept of stacking ('piles' in French), which introduces the notion of paradigm in syntax, was introduced by the Aix School (Blanche-Benveniste et al. 1990). Stacks are the multiple realization of one and the same structural position, which occurs in continuous speech in various types of segments, especially syntactic disfluencies. See (8) below (disfluencies in bold) where the same utterance is annotated (a) in the CorpAfroAs format (Mettouchi, Vanhove & Caubet 2015), (b) in the Aix format:

(8)   a   *tô: **kə́### kə́** dû **te: ɣə́** / **te:** gə̀ʃi tsə́n ŋ //*
                'Well you... *you would* beat it *towards er... downhill* like this.'
                (Bury_Har_052)

         tô: kə́   ##   kə́   dû   te: ə́

         DM   2PL.AOR   ##   2PL.AOR   beat   around   FILL

teː	gɜ̀ʃɪ	tsɜ́n	ŋ
around	downhill	like_this	COP2

b   *tôː*   **kɜ́**
            kɜ́   *dû*   **teː ɣɜ́**
                            teː gɜ̀ʃɪ   *tsɜ́n ŋ //*

     'well   *you would*
             you would   beat (it)   *toward er*
                          downhill   like this //'

Rhapsodie proposes a complete annotation and a functional tagging of stacks (Kahane & Pietrandrea 2012). The format is: { -- | -- }. See (8c) for the same unit thus annotated:

(8)   c   *tôː: <{kɜ́ | kɜ́} dû {teː ɣɜ́ | teː gɜ̀ʃɪ} tsɜ́n ŋ //*

           'Well < *{you'd | you'd }* beat (it) *{ toward er | downhill }* like this //'

Although extremely frequent in spoken language, this cohesion device, which can be regarded as a particular type of micro-syntactic relation, is often disregarded in corpus annotation. By extensively annotating and tagging stacking phenomena this annotating script aims at giving an exhaustive micro-syntactic annotation of all the data, including disfluencies, repetitions, and reformulations generally considered as performance errors and not analysed in spoken language treebanks.

Apart from disfluencies and reformulations, stacks note a micro-syntactic relation where text segments occupy the same position in the dependency structure. They appear within IlUs, e.g. in coordination, as in example (9) below. The sign "^", e.g. in ^kóː --- ^kóː, '^or --- ^or' identifies words working as conjunctions for illocutionary units.

(9)   a   *wáːni {àː fîn & | àː ŋgap ŋaː gɜ̀dî wáːni maráy | ^kóː àː ʧɪ mɜ̂ːr | ^kóː*

               á ʃɪ maːndɜ} //

               'So-and-So {has done & | has caught So-and-So's daughter and spoilt

               her | ^or has stolen | ^or has gone into a fight}' // (Rel_Har_188)

Following (Blanche-Benveniste et al. 1990) this stack can be represented as in (9b) below

(9)　b　　*wá:ni*　　　*à: fìn &*
　　　　　　　　　　　à: ŋgap ŋa: gɔ̀dì wá:ni maráy
　　　　　　　　　　　^kó: à: ʧì mɔ̂:r
　　　　　　　　　　　^kó: á fì ma:ndə

　　　　　'So-and-So　has done &
　　　　　　　　　　　has caught So-and-So's daughter and spoilt her
　　　　　　　　　　　^or has stolen
　　　　　　　　　　　^or has gone into a fight'

1.4　Non-alignment of Illocutionary and Government Units

Illocutionary and government units are not necessarily aligned, and, as a consequence, the annotation system does not consider IlU or turn-taking as boundaries of macrosyntactic dependencies. The same annotation device is modified to mark stackings across IlU boundaries as follows : { --- |}　{| --- }.

1.4.1　Stacking across IlU boundaries

Example (10) illustrates stacking across IlU boundaries when illocutionary and government units are not aligned.

(10)　a　　tô: mɔ́ ŋgyǎ:r gya: gà:l　　　ɓét ɗaŋ //
　　　　　　　　　　　　　^kó: gèri
　　　　　　　　　　　　　^kó: ma:t //

　　　b　　*"tô:" mɔ́ ŋgyǎ:r {gya: gà:l |} ɓét ɗaŋ //+ {| ^kó: gèri | ^kó: ma:t} //*
　　　　　　'"Well" we slaughter plenty {cows |} too //+ {| ^or hens | ^or goats}. //'
　　　　　　(Cal_Sdy_ 032)

In this example, the first IlU finishes with the adverbial adjunct *ɓét ɗaŋ,* 'plenty too' and the end of the unit is marked with a terminal prosodic break. Then, as an afterthought, two nouns are added, forming a discontinuous chain of three coordinated direct objects (*gyá: gà:l,* 'cows;

9

gèri, 'hens' and *ma:t*, 'goat') of the verb *ngyá:r* 'slaughter'. The afterthought forms a second IIU starting with a pitch reset and finishing with its own terminal prosodic break. The stacking links the direct objects across the IIU boundary.

1.4.2 Stacking across turn-taking

Stacking through coordination can occur across turn-taking and result in elliptic structures. Instead of considering those as either incomplete structures or structures where most of the elements have been omitted, they are considered as a special case of micro-syntactic relations, i.e. coordination across turn-taking.

It is illustrated in example (11) below, where the nouns *gàt* 'woman' in (11a, b, and d), and *ŋa: gàt* 'girl' in (11c) are all part of the same stack, and share the same syntactic relation as part the adverbial adjunct[3] of the nucleus of (11a) : *tá gì: tà gòs dò:*, 'they will bury her where?'

(11) a   [S1] *"tô" { gàt]} kən >+ yá: mâs kúmá <+ tá gì: tà gòs dò: //*

            tô: gàt   kən   yá: məs kúmá

            DM   woman COP2 3SG.COND die   too

            tá   gi: tə   gòs   dô:

            3PL.FUT bury 3S.OBJ 3SG.POS   where

            '"Well" if it is *{ a woman]}* >+ that dies <+ they will bury her where? //'

  b   [S2] *{[gàdà:]} //*

            gàt   a:

            woman QUEST

            '{[A woman?]} //'

---

[3]   The clausal adverbial adjunct (*gàt kən yá: mâs*, 'if it is a girl that dies') is a conditional with a cleft embedded.

c [S1] *{| ^kó: ŋa: gɔ̀t }* //

     kó: ŋa: gɔ̀t

     or young woman

     '{| ^Or a girl. |} //'

d [S2] *ŋa: gɔ̀t < tá gì: ʃĭ ɓɔ́ʣɔ̀ŋ > kápwâ:sɔ̀ŋ // [...]*

     ŋa: gɔ̀t tá   gi: ʃĭ   ɓɔ́ʣɔ̀ŋ káp wá:sɔ̀ŋ

     young woman 3PL.FUT bury 3PL.OBJ outside all 3PL.POS

     'Girls < they would bury them outside > all of them. // [...]'

e [S1] *{| ^tɔ́ gɔ̀t ɓét |}* kó: //

     tɔ́ gɔ̀t   ɓét kó:

     with woman all   or

     {| ^*And women in general* |} or what? //

f [S2] "m̀:" ^tɔ́ gɔ̀t ɓét < tá gì: ʃĭ dân //

     m̀: tɔ́ gɔ̀t ɓét tá   gi: ʃĭ   dáni

     er with woman all 3PL.FUT bury 3PL.OBJ there

     '*"Yes"* women in general < they would bury them there.//'

     (Bury_Sdy_20)

The elements coordinated across the turn-taking continue the microsyntactic construction:

'{ gɔ̀t | ^kó: ŋa: gɔ̀t | ^tɔ́ gɔ̀t ɓét } kən >+ yá: mâs [...], ' if it is { women | ^or girls | ^and women in general } that die [...]'). The noun in S2's echo-question ({| *gɔ̀dâ: |}*,

'{| women |}?', where the final –*a:* marks the interrogative sentence modality) inherits the

same function as the coordinated elements in S1's turns : **gɔ̀t kən** *>+ yá: mâs a: [...],* 'if *it is*

women that die [...][4]?'.

---

[4] The whole passage (11a to f) is 2mn50s long. [...] in (11d) stands for 2 IlUs giving more precise information about the distance between the grave and the family compound, which have not been reproduced here as irrelevant for the point discussed.

1.4.3   Microsyntactic dependencies across IlU

As we can see, macrosyntactic constituent units which reflect and convey information structure are not necessarily congruent with microsyntactic structures. When microsyntactic dependencies override IlU boundaries, a sign "+" is added to these boundaries: "//+" for final IlU boundaries; "<+" and ">+" for non-final. An illustration is given in (12) below, an utterance with a cleft structure, where the ">+" sign shows that the post-nucleus unit is in a dependency relationship with an element of the nucleus, and the "//+ sign" shows that *yâːn nə myâːn,* 'if it's me' is a clausal adjunct added to the nucleus as an afterthought:

(12)   *"tôː" dzàŋ gíː ɣəŋ >+ tá fĩ mátaŋgáy //+ yâːn nə myâːn //*

   tôː   dzàŋ   gíː   kən   tá   fĩ   mátaŋ   káy
   DM   day   DIST   COP2   3PL.FUT   do   ritual_flogging   LOC

   yâːn   nə   myáːni
   if   COP1   1SG.IDP

   '"Well" *it's that day>+* they will do matang //+ if it is me. //' (Bury_Ha_201)

1.4.4   Embedding

Illocutionary units are embedded in cases reported speech and asides, or parentheses.

- *Reported speech* (both direct and indirect).

The reported speech is governed by the introducing particle *tu* : [--- //]

(13)   *á wû **tu** [yáːwón <+ wò ɬə duʃĩ wàya mə́n ɗa //] //*

   á   wul   tu   yáːwón   wò   ɬə   ɗu   =ʃĩ
   3SG.AOR   say   comp   today   3SG.FUT   go   beat   3PL.OBJ

   wə́ya   mə́n   ɗa
   phone   BEN   again

   'He said [today <+ he will phone them again. //] //' (Girls_A_097)

- *Evidentiality*

Evidentiality-introducing verbs (see, think, etc.) share the same structure as reported speech. The illocutionary particle *–o:* on the embedded predication below is a hint that the reported speech functions as an illocutionary unit.

(14)   *myá: **yel ku tu** [gyò: ɣəndá >+ wò mop de:déyo: //] //*

myá:	*yel*	*kutu*	*gyò:*	*kəndá*	*wò*	*mop*	*de:dé:*	*-o:*
1sg.icpl	see	as_if	which	=COP2	3SG.FUT	surpass	correct	FCT

'I *wonder* [which one is it >+ that will be best? //] //' (Girls_B_016)

This structure can be extended to the same verbs without the introducing particle *tu*:

15   *tsə́tŋg�̂n dən má: < **myá: yel** [nə lă: bàptàk ɓasmí //] //*

tsə́tn	*-kə́nì*	*dən*	*má:*	*myá:*	*yel*
sit	-NMLZ	house	even	1SG.ICPL	see

nə	*la:*	*kə́*	*bàptàk*	*ɓas*	*=mí*
COP1	work	POSL	useless	by	1PL.OBJ

'Sitting home too < *I see* [it is useless for us .//] //' (Girls_B_035)

- *Parentheses* : (--- //)

(16)   *{^yâ:n nə múr & | yâ:n nə {mə́m & | mə́m vàr} **(wà:tò: mur ɣə́ dən //)** <+ "tô:" tá gì: tə̀ gìp dən ŋgá vì:n wón //+ ^ɗaŋ gyá: ndú:ri ʧiɣə́y //*

'{^if it's a man & | if it's {an old & | an old man} *(that is to say somebody of the house //)* <+ "well" they will bury him inside the house in a hut //+ ^like in a storeroom like this. //' (Bury_Har_034)

1.4.5   Parallel IlUs

Some IlUs are linked by lexico-structural similarities that bind them beyond mere paratax. This is noted by //=.

(17)   *má ʧî kádî hή // kyáː ʧí kádî <+ ɣa tu ŋaː gɜ̀t dɑ̃ŋ //* **ka ʧi ʒaːki hή //= ka ʧi**
pərʃi hή //

'We don't eat dog // if you eat dog <+ you will not find a wife // *you don't eat*
donkey //= you don't eat horse. //' (Rel_Har_136)

The previous section has introduced a punctuation system which identifies the boundaries of
macrosyntactic units together with information concerning the congruence of these
boundaries with microsyntactic dependency. To make the corpus available for syntactic and
information structure queries, and for the study of their interface, further specific annotations
are necessary, i.e. Information Structure (IS) and microsyntactic tagging. In the last section,
an example of syntactic annotation within the dependency framework will show how the
contrast between clefting and left-dislocation can be represented. The next section will deal
with Information Structure.

2   Peripheries and IS tagging

2.1   IS tagging

Information Structure tagging has been done with a new module of ElanCorpA (Chanard
2014) that is being developed by M. Aouini & C. Chanard at Llacan, as part of the Cortypo[5]
programme. This module is a new type of annotation, based on the annotation tiers that
already exist in the CorAfroAs / Cortypo format. This new functionality in Elan is meant to
create annotations on a dependent tier that cover non-contiguous annotations of the parent

---

[5]   The Cortypo programme currently directed by A. Mettouchi (http://cortypo.huma-num.fr/index_fr.html)
is a follow-up of the CorpAfroAs programme (Mettouchi, Vanhove & Caubet 2012).

tier. For a given annotated file in the 'classical' Elan format, extra annotations can be created as new lines in two sets of tables: Groups and Links. Individual groups and links in the table can then be selected and highlighted in the annotation tiers, where the corresponding passage in the sound file can be played. The file can be searched, with multiple criteria including tier annotations, table, and distances in terms of alignment, annotation and time span. These tables can be sorted by types, names or annotations, which has a great heuristic value and opens new possibilities for structural annotations (whether informational or syntactic) in Elan. Figure (1) shows a screenshot of Example (12) annotated for SI with Elan and the Links and Groups module.

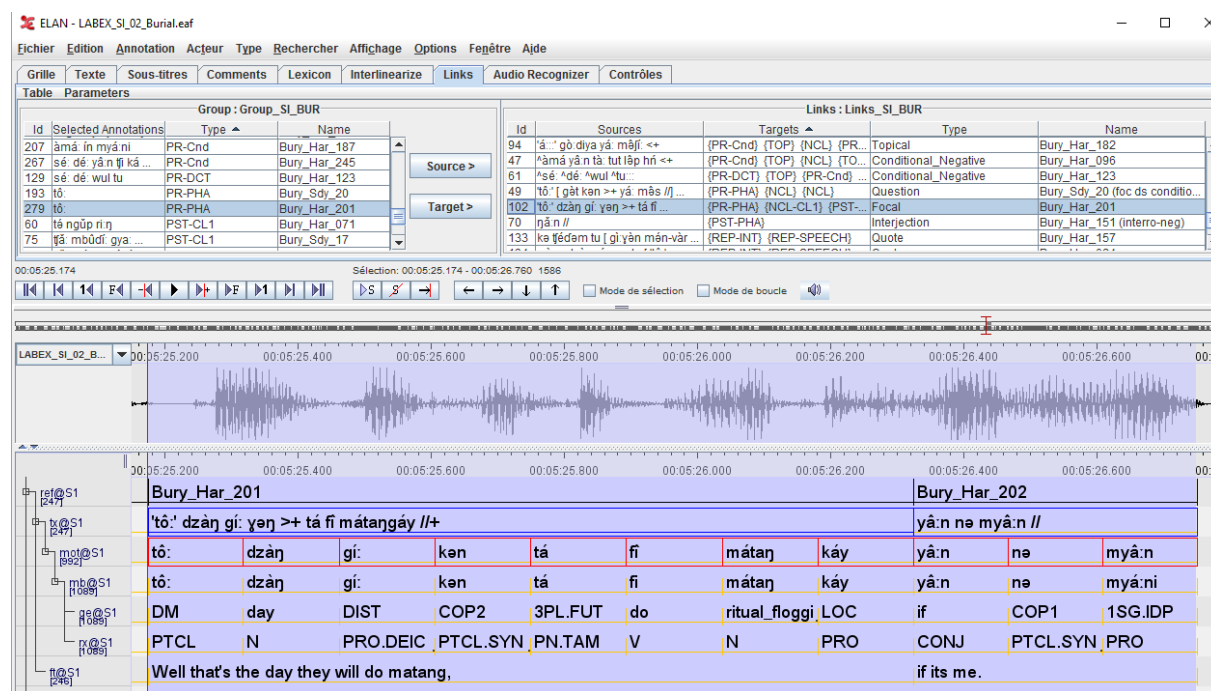*Figure 1 : Screenshot of Ex.(12)*

In the first table (called Groups, top left of the screen), to create a group, the annotator selects a set of annotations in any of the existing tiers, gives this group a name and a type that can be selected in a controlled vocabulary. These sets consist of a single or several annotations that can be selected from one or several tiers, and can be discontinuous.

For this work, I used the Groups table to identify sets of words that make up Illocutionary Constituents (IlC: Nuclei, Pre-nuclei, Post-nuclei and In-nuclei), and tag them with their function (type) and reference number (name). (See Figure (2)).

	Group : Group_SI_BUR		
Id	Selected Annotations	Type ▲	Name
207	àmáː ín myáːni	PR-Cnd	Bury_Har_187
267	séː déː yâːn tʃi káʔ ...	PR-Cnd	Bury_Har_245
129	séː déː wul tu	PR-DCT	Bury_Har_123
193	tôː	PR-PHA	Bury_Sdy_20
279	tôː	PR-PHA	Bury_Har_201
60	táʔ ngŭp riːŋ	PST-CL1	Bury_Har_071
75	tʃ̃ãː mbûđíː gyaː ...	PST-CL1	Bury_Sdy_17

Figure 2. Groups table sorted by Type
with the phatic pre-nucleus of Ex. (12) selected

In the second table (called Links, top right of the screen), the annotator creates links between two sets of annotations built on the same principle as groups. One set is called the Source, and the other set is called the Targets. The links created are given a Name and a Type in the same way as for groups. The sources or the targets can also be taken from the Groups table. In this case, the sets selected from the Groups table can be viewed either by showing the annotations in the tiers, or the types and names given to the groups in the Groups table.

For this work, I have used the Links table to tag the Illocutionary Units. For better readability and convenience sake, the table shows the full text on the text tier as the source of the links and the IlC (groups) tagged in the Groups table as targets. I have used the "type" column of the Groups table for a temporary, rule of thumb functional tagging of IlU, indicating whether they contain e.g. questions, conditionals, rhetorical devices such as parallel IlUs, etc. (See Figure (3))

	Links : Links_SI_BUR			
Id	Sources	Targets ▲	Type	Name
94	'áːːː' gòːdiya yáː mêʃíː <+	{PR-Cnd} {TOP} {NCL} {PR...	Topical	Bury_Har_182
47	˄àmá yâːn tàː tut lêp hń <+	{PR-Cnd} {TOP} {NCL} {TO...	Conditional_Negative	Bury_Har_096
61	˄séː ˄déː ˄wul ˄tuːːː	{PR-DCT} {TOP} {PR-Cnd} ...	Conditional_Negative	Bury_Har_123
49	'tôː' [gàt kən >+ yáː mâs //] ...	{PR-PHA} {NCL} {NCL}	Question	Bury_Sdy_20 (foc ds conditio...
102	'tôː' dzàŋ gíː yəŋ >+ táʔ fíː ...	{PR-PHA} {NCL-CL1} {PST-...	Focal	Bury_Har_201
70	ŋãːn //	{PST-PHA}	Interjection	Bury_Har_151 (interro-neg)
133	kə tʃéđəm tu [gìːyàn mén-vàr ...	{REP-INT} {REP-SPEECH}	Quote	Bury_Har_157

Figure 3. Links sorted by types showing Ex (12) in IlU classification

The corresponding annotation can be selected and viewed in the tiers below and the corresponding sound segment can be played via the media player in Elan. (See Figure (4))
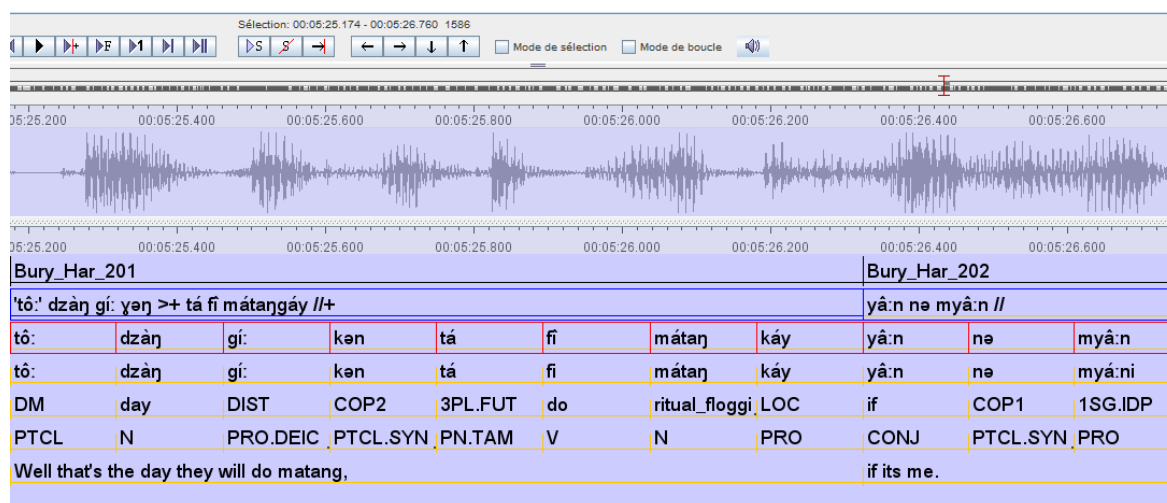
Figure 4. Annotation and sound file segment corresponding to Ex.(12)

The labels tagging the macrosyntactic constituents are both structural and functional. They can be divided along two lines : (i) nucleus and pre-, in- and post-nucleus; (ii) aligned vs. non-aligned on the other hand.

	Aligned constituents	*Non-aligned constituents*
Pre-Nucleus	PR-ALL : Allocutive, Vocative PR-DCT: Discourse connector PR-EXP: Expressive PR-PHA: Phatic PR-TOP: Left-edged Topic	PR-Adv: Left-dislocated adverbial adjunct PR –Cls: Left-dislocated clausal adjunct PR –Cnd: Left-dislocated conditional adjunct PR-CL2: Pre-nucleus section of Pseudo-clefts
Nucleus	NCL	NCL-CL1 (Nucleus of it-Clefts) NCL-CL2 (Nucleus of Pseudo-clefts)
Post-Nucleus	PST-ALL: Allocutive, Vocative PST-DCT: Discourse connector PST-EXP: Expressive PST-TOP: Right-edged Topic	APX: Nucleus Appendix (Afterthought) PST-CL1: Post-nucleus section of it-Clefts
In-Nucleus	GFT: Graft ; PAR : Parenthesis	

Table: Group Types tagging macrosyntactic constituents

Using this tagset, I have been able to test a tentative typology of peripheries (pre- and post-nucleus units) on 11 annotated files (90 minutes, 15 000 words). I was able to extract the list of illocutionary constituents, and check the consistency of the annotation. The aim of this type of extraction is to look for regularities in the marking of the units, in syntax, morphology, intonation and reference tracking, i.e. do a basic bottom-up research. It is clear that the relevancy of the results is dependent on the tagging, which is based on my intuition and understanding of the language. Of course, this bottom-up stance is not devoid of any theoretical bias, but the exhaustivity of the annotation will (and already has) lead me to a revision of my analyses and some of the labels used for tagging. This labile process must strike a balance between rapidity of annotation (a process which can be very time consuming) and how fine-grained our analysis needs to be. To be fully labile, the tagging system must anticipate the need for regular revisions, e.g. automatic conversion and collapsing of categories.

The next section is devoted to a typology of the peripheries retrieved in the corpus with the groups table sorted by type. (See Figure (2).)

2.2   *Typology of peripheries*

When micro- and macro-syntactic dependencies are aligned, the boundaries of the nucleus correspond to the microsyntactic dependency unit of the verb/predicate carrying the illocutionary act, and include all the elements governed by this head. All the dialogic units are aligned (*viz* outside the government of the nucleus head). The aligned textual units are: Discursive links (PR-DCT) and Topics (TOP and ANT). As for non-aligned units, the pre-nucleus governed constituents comprise left-dislocated adjuncts (PR-Adv, PR-Cls, PR-Cnd) and the pre-nucleus constituent of pseudo-clefts (PR-CL2). The post-nucleus governed

constituents are the nucleus appendix (APX, e.g. afterthoughts) and the post-nucleus constituent of it-Clefts (PST-CL1).

2.2.1 Aligned peripheries

Aligned peripheries are divided into two classes which are respectively dedicated to different types of information functions: a) the textual construction of the utterance (textual peripheries, e.g. Topic, Appendix, Locutive Introducer); b) its communicative support (dialogic peripheries, e.g. Phatic, Allocutive, Expressive, etc.) (Cresti 1999:15). The only textual periphery that is not governed by the head of the nucleus is the Topic (TOP) and it appears massively in pre-nucleus position: only 2 examples of post-nucleus topics (also called right-edged topics, or antitopics: ANT) are found in the corpus, against 611 cases of TOP. Topics are illustrated below in (18) for left-edged Topics (TOP) and in (19) for right-edged Topics (ANT). Left-edged Topics will be examined in further details in the next section in contrast with Clefts and Frame-setters. Righ-edged topics are characterised by a low tone, flat contour, and follow a non-final prosodic break.

<div align="center">

TOPIC

</div>

(18) ***tsә́tŋgân dә̀n má: <*** *myá: yel [nә lă: bàptàk ɓasmí // //*

tsә́tn	*-kә́nì*	*dә̀n*	*má:*	*myá:*	*yel*
sit	-NMLZ	house	even	1SG.ICPL	see

nә	*la:*	*kә́*	*bàptàk*	*ɓas*	*=mí*
COP1	work	POSL	useless	PREP	1PL.OBJ

'*Sitting home* < I see [it is useless for us. //]. //' (Girls_B_035)

(19) *gò̀pm < kó:dzà̀ŋgyò: <+ mìɣá la: káwêy >* ***myà:ní:::*** ***gút ẕa:r //***

gò̀pm	kó:dzàngyó:	mìká	la:	káwêy
1PL.POS	everyday	1PL.CONT	work	merely

myà:ní   guɗi   za:r

1PL   woman.PL   human

'We < everyday <+ we do nothing but work > *we Zaar women. //*'

(Wom_A_169)

Dialogic constituents are used to establish, maintain or qualify the illocutionary act. They occur before, or after the nucleus. They are surrounded by inverted commas in the transcription. It is possible to distinguish the following types:

- *Phatic* (PR-PHA & PST-PHA), dedicated to control the communicative channel, ensuring its maintenance. They are either fillers (e.g. *er…*, *mm…*), discourse punctuators (*OK*, *well*, Zaar *tô:*), marks of agreement with the speaker (*uh*, Zaar *m:*, *è:*), etc.

- *Allocutive* (PR-ALL & PST-ALL), specifying to whom the message is directed, keeping their attention (Vocative, you know, you see) or introducing evidential modality (I think, etc.).

- *Expressive* (PR-EXP & PST-EXP), giving an emotional strength to the illocutionary act.

- *Connective* (PR-DCT & PST-DCT), linking different parts of the discourse (utterances within a turn, or across turns) maintaining some explicative, causal, temporal or concessive values. Most of them occur in pre-nucleus position.

<div align="center">

PHATIC

</div>

In (20), three cases of phatic units are exemplified: *tô:*, 'well' and *yâwwà:*, 'OK' as PR-PHA, and the TAG *ŋǎ:n*, 'no?' as PST-PHA.

(20)   **"*tô:*"** *< yâ:n < ʧǎ: ʧím tɘ̀ ɣá vì: válti tu [ʃéro //]* > **"*ŋǎ:n*"** *//* **"*yâwwà:*"** *< "***tô:***"*

           < átâ yi ʧík //

tô:	*yâ:n*	*ʧǎ:*	*ʧim*	*tə*		*ká*	*vì:*	*válti*	*tu*
DM	3SG.IDP	3PL.ICPL	call	3S.OBJ		at	speech	muslim	comp

ʃéro	*ŋǎ:n*		*yâwwà:tô:*	*átâ*	*yi*	*ʧík*
flogging	QUEST		ok	DM 3SG.REM	be	thus

'"Well" < this < they call it in Hausa [shoro //] > "no"? //"OK" < "well" <

that's how it used to be. //' (Bury_Har_149)

<center><i>ALLOCUTIVE</i></center>

Vocatives are examples allocutives that can appear either before (PR-ALL, in (21)) or after

the nucleus (PST-ALL, in (22)):

(21) "ká" < **àfó:** < ká: ye yáddiyó:ɗam myá: sú: sú kámʃâk > "kwǎ:" //

ká		*àfó:*	*ká:*		*yel*	*yáddiyó:ɗan*
disapproval		Afo	2SG.CPL		see	how

myá:		*sú:*	*sú*	*kámʃâk*		*kwǎ:*
1SG.ICPL		like	PL	Kamshak		FCT

'"What" < Afo < you saw how I like Kamshak > "anyway". //' (Girls_B_073)

In (22), the speaker is protesting, using a yes/no rhetorical question, ending in a vocative.

(22) *mə̀ káp ŋgasa: > àfó: //*

mə̀		*kap*	*ngas*	*-a:*	*àfó:*
1SG.SBJV		take	Angas	-VRT	Afo

'We should marry Angas people?! > Afo! //' (Girls_B_104)

In (23) *ká: yisə́ŋ*, 'you know', shows another way of maintaining the communicative channel:

(23) "dón" < **ká: yisə́ŋ** < farko má:<+ ɗaŋ kámʃâk tà wu tu [ʧá: sû:m //] <+ mətá

wultə tu [ba:bù //] //

dón	*ká:*	*yisə́ŋ*	*farko*	*má:*	*ɗan*	*kámʃâk*	*tà*
because	2SG.CPL	know	beginning	even	as	Kamshak	REM

ʧá:	*sú:*	*=mə*	*mətá*	*wul*	*=tə*	*tu*	*ba:bù*
3SG.CPL	love	1SG.OBJ	1SG.REM	say	3S.OBJ	COMP	no

'"Because" < *you know* < in the beginning <+ as Kamchak said [he loved me//] <+ I told him [no way //]. //' (Girls_B_147)

<center>*EXPRESSIVE*</center>

Exclamations in Pre-Nucleus position (e.g. PR-EXP; *kâːy*, 'hey' in (24).

(24)   *[Sp1] sɔ̀kéːdì ʧá: ndará //*

  *[Sp2] "**kâːy**" < ʧá: poləmgáy sòːséy //*

  sɔ̀kêːt -i   ʧá: ndará

  skirt   -INDF   3SG.CPL   be_proper

  kâːy   ʧá: pol   =mə   káy   sòːséy

  eh   3SG.CPL   please   1SG.OBJ   LOC   quite

  [Sp1] 'The skirt is nice. //'

  [Sp2] '"*Hey*" < I really like it. //' (Girls_B_069)

<center>*CONNECTIVE*</center>

In (23), *don*, 'because', is an initial discursive link (PR-CNT) working as a connective.

2.2.2   Non-aligned peripheries

As constituents governed by the head of the nucleus, all non-aligned peripheries are textual.

<center>*PRE-NUCLEUS (<+)*</center>

The pre-nucleus governed constituents are left-dislocated adjuncts (PR-Adv, PR-Cls, PR-Cnd) and the pre-nucleus constituent of pseudo-clefts (PR-CL2).

- PR-Adv, or left-dislocated adverbial adjunct

(25)   *"tòː"< **dzàŋ làːdì** <+ má ɬĭː kində > "báː" //*
  tô: dzàŋ   láːdì   má   ɬə -íː   kində báː:
  well   day   Tuesday   1PL.FUT   go RES   Kində NEG1
  '"Well" < *on Tuesday* <+ we'll go to Kində > "no". //' (Girls_A_001)

- PR-Cls, or left-dislocated clausal adjuncts

(26)   "tò:" < kyà: gì: tí <+ "tô:" < ká ʐə̀:rí: ʧĭp //

tô:	*kyǎ:*	*gi:*	*tə*	*-í:*	*tô:*	*ká*	*ʐə́:r*	*-í:*	*ʧĭp*
DM	2PL.ICPL	bury	3S.OBJ	RES	DM	2PL.FUT	stay	RES	quietly

'"Well" < after you had buried him <+ "well" < you would sit still. //'
(Bury_Har_046)

Correlative conditionals (i.e. conditionals with a temporal meaning: 'if (=when, =each time

that) … then…') are analysed just like ordinary adjuncts:

(27)   **yá: yelmə̌ŋ <+** ʧá: fitə wusúŋə̌ŋ > "éy" //

yá:	*yel*	*=mə*	*hń*	*ʧá:*	*fi*	*=tə*
3SG.COND	see	1SG.OBJ	NEG2	3SG.CPL	do	3S.OBJ

wusúŋ	*hń*	*éy*
be_nice	NEG2	indeed

'If/when he does not see me <+ he is not happy > "hey". //' (Girls_B_077)

- PR-Cnd, or left-dislocated conditionals

(28)   *^yâːn hali ɗa kàm <+ má dìːɓí //*

yâːn	hali	ɗa	kàm	má	dìːp	-i
if	chance	COP3	indeed	1PL.FUT	buy	SPCF

'^If there is a chance <+ we will buy it. //' (Girls_B_056)

- PR-CL2, or pre-nucleus section of Pseudo-clefts

(29)   *^àmá: mə́n yóːɗan ʧǎː fi <+* nə mə́n mársəŋ //

àmá:	*mə́n*	*yóːɗan*	*ʧǎː*	*fi*	*nə*	*mə́n*	*mársəŋ*
but	people	which	3PL.ICPL	do	COP1	people	Marsang

'^But *the people who do it* <+ are the people of Marsang. //' (Cal_Har_010)

Example (30) shows a case of pseudo-cleft where a pro-verb (*fi*, 'do') is relativized in pre-

nucleus position in order to focus the predicate, yielding the structure 'what we will do is…'.

(30) *^dɔ́n tɔ̀ gí: <+ **ŋgɔ́tn wɔ́n ɗa má fì ɗaŋgəní káwêy** <+ sé: mɔ̀ pàŋ kɔ̀*
páŋgɔ̂ŋgéy //

dɔ́n	*tɔ̀*	*gí:*	*ngɔ́tn*	*wɔ́n*	*ɗa*	*má*	*fì*	*ɗaŋgəní*	*káwêy*
because	with	DIST	thing	QLT	REL1	1PL.FUT	do	now	merely

sé:	*mɔ̀*	*paŋ*	*=kə*	*paŋ*	*-kɔ́nì*	*káy*
up_to	1PL.SBJV	inspect	2SG.OBJ	inspect	NMLZ	LOC

'^Because ^of ^that <+ *what we will merely do now* <+ is to think about it really well. //' (Girls_B_186)

POST-NUCLEUS

The post-nucleus governed constituents are the nucleus appendixes (APX, e.g. afterthoughts) and the post-nucleus constituent of it-Clefts (PST-CL1).

- PST-CL1

Clefts constitute a single intonation unit. In these constructions, the illocutionary nucleus is not on the predicate, which follows in the post-nucleus situation but on the specifying copula. In the following examples, the nucleus is bolded, and the ">+" sign that follows the nucleus indicates that there is a dependency relation with what follows.

(31) *^dò:mín < sɔ́ŋwa:rí < **^sé: dà:ʃì yá: môr lŭ:y** >+ ^əndá ʧà:yi vər tə sɔ́ŋwa:rês //*

dò:mín	*sɔ́ŋwa:rí*	*sé:*	*da:*	*-ês*	*yá:*	*mor*
because	chief_priest	only_if	person	DEF	3SG.COND	do_a_little

lu:	*-í:*	*kəndá*	*ʧà:yí*	*vər*	*tə*	*sɔ́ŋwa:rí*	*-ês*
get_old	RES	then	3PL.ICPL.ITER	give	3SG.OBJ	chief_priest	DEF

'^Because < a chief < *(it's) only when a man is a bit old* >+ ^then they make him a chief. //' (Rel_Har_008)

(32)   *"tò:" < **gí:** >+ kə́ mân //*

        tô: gí: kə́   mán

        well   DIST 2PL.AOR come

        '"Well" < *that* >+ you have come (for)? //' (= "well", is THAT what you have come for?) (Girls_A_090)

(33)   *"â:" < dzàŋ là:dì má: <+ **kakáp** >+ má gè:wàyéy //*

        â: dzàŋ   lá:dì   má: kakáp má   ge:wayé   -í:

        ah   day   sunday even all   1PL.FUT walk_around RES

        '"Ah" < on Sunday too < *(it's) everywhere* >+ we'll walk around. //'(Girls_A_010)

Afterthoughts, which are elaborations or correction of the illocutionary act of the nucleus, are expressed in a different IlU. They are preceded by a final intonation break and a pitch reset, and they receive a falling contour.

(34)   *móʃì makaranta < ma dyǎ:ŋo: //+ ^**sé: tə̀ ŋál kə́lá:sò:** //*

        mókʃì   makaranta ma   dyá: hń   -o:

        courting   school   1SG.FUT be_able NEG2 FCT

        sé: tə̀   ŋal   kə́lâ:s   -o:

        unless   3SG.SBJV look_for class   FCT

        'Dating in school < I couldn't do it //+ ^*unless he changed class. //*'
        (Girls_A_076)

Likewise, *kápwâ:sə̀ŋ*, 'all of them' in the long example (11d) above, is an appendix added to the nucleus as an afterthought after a final break.

3   Left-dislocation and clefting in Zaar

In this section I propose to run a survey of the Zaar corpus in order to do a comparative study of clefts, topics, and left-dislocated circumstantial adjuncts, tagged resp. NCL-CL1, TOP and

PR-Adv/Cls/Cnd in the Groups table. After defining them, I will contrast their prosodic and syntactic properties, their functions and finally give a representation of their syntactic structure.

When studying a Zaar corpus, the linguist is struck by the overwhelming presence of compound utterances (Cresti & Moneglia 2010:15) comprising a left-dislocated IlC sharing the same prosody consisting in a high tone, a flat contour, usually followed by a prosodic break marked by a pitch reset and fall and usually (but not necessarily) by a pause. In our corpus, out of a total of about 1,400 utterances, 586 have been tagged as compound utterances, while 571 have been tagged as simple (thetic, all-new) and 108 have been tagged as cleft.[6] These left-dislocated IlCs characterised by the same intonation pattern have been analysed as belonging to two different classes: Topics and left-dislocated circumstantial adjuncts. As for clefts, they constitute a single intonation unit with no break, and are characterised by a fall from a main stress falling on the cleft phrase.[7]

3.1   Definition and characterisation

3.1.1   Topics

They introduce a referent, selected out of the on-going conversation, or of the common knowledge of the speakers. These referents provides the necessary pragmatic information for the illocutionary act carried by the following nucleus. Example (35) shows two topics : *laː*, 'work' ; *məːríwôpm*, 'our children'.

---

[6]   Other categories have been tagged off and on, which explains why these figures must be taken as, at best, rough estimates.

[7]   See below for a definition of this term.

(35)   **la: < mə:ríwôpm** < *ʧì gwàːsə̀ŋ tə́ laː hń́ //*

laː	mə̀ːri	kə́	=wopm	ʧì	gwàːsə̀n	tə́	laː	hń́
work	child.PL	POSL	1PL.POS	3PL.be	3PL.POS	with	work	NEG2

'As for work < our children < they themselves don't have any work. //'

In this example, the first topic is repeated in the nucleus; while the second one is represented by the 3rd person index of the person and TAM complex *ʧì* which assumes the function of subject of the verb. This hints at the fact that the topics, separated from the nucleus by a prosodic break, are not syntactically dependant on the verb or the predicate of the nucleus, and their relation to the nucleus is not syntactic. This relation is pragmatic, and is best befined by the notion of "aboutness".

The notion of "aboutness" (Sperber & Wilson 1986; Lambrecht 1994) has always been central to the definition of topics. The information function of the topic is to identify, through linguistic means, the domain of relevance for the illocutionary force carried by the nucleus, its pragmatic domain of identification. This is conveyed by the name "aboutness topic" commonly used to refer to this construction. (Krifka & Musan 2012; Schultze-Berndt 2013; Simard 2014). To paraphrase (Cresti & Moneglia 2010:18), the topic specifies the *pragmatic* aboutness of the nucleus.

Topics cannot enter in a *syntactic* relationship with the verbs of the nucleus. When a topic is in a pragmatic relation with the verb, and this relation corresponds to a dependency relation (the element could be an argument of the verb), the syntactic relation must be realised as a clitic so that the valency of the verb is saturated. In (36) above, the word *laː*, 'work', repeats the topic inside the nucleus to saturate the locative predicate *yi tə́*, 'have', lit. 'be with'. In (36), the topic *gə̀ːrí raːs,* 'old locust-bean tree' is co-referential with the adverb *ɗanì,* 'there' which is an adjunct of the verb *gìː,* 'bury'. In (37), a clitic (the direct object pronoun *ʃì,* 'them') saturates the valency of the verb *gìː,* 'bury'.

(36) *gyá: gə̀~ ra:sə́n tsə́n < tá gì: ʃí **dân** //*

gyá:	gə̀:rí	ra:s	-ín	tsəní	tá	gi:	ʃí	dáni
PL	old	locust_bean	PROX	like_this	3PL.FUT	bury	3PL.OBJ	there

'These old locust-bean trees like this < they would bury them *there. //*'

(Bury_Har_109)

(37) *ŋa: gə̀t < tá gì: **ʃí** ɓə́ɮə̀ŋ //+ kápwâ:sə̀ŋ //*

ŋa:	gə̀t	tá	gi:	ʃí	ɓə́ɮə̀ŋ	káp	wá:sə̀ŋ
young	woman	3PL.FUT	bury	3PL.OBJ	outside	all	3PL.POS

'Girls < they would bury *them* outside //+ all of them. //' (Bury_Har_103)

3.1.2 Left-dislocated circumstantial adjuncts

If left-dislocated circumstantial adjuncts share the same intonation pattern as topics in Zaar, their function and properties set them apart. It is agreed that adverbials and other circumstantial adjuncts are frame-setters that limit the applicability of the main predication to a certain restricted domain (Chafe 1976). Using the concept of common ground, Krifka & Musan (2012) establish a difference between contrastive topics and frame-setters which can be extended to aboutness topics:

> With contrastive topics, the current common ground management contains the expectation that information about a more comprehensive, or distinct, entity is given; contrastive topics indicate that the topic of the sentence diverges from this expectation. With frame setters, the current common ground management contains the expectation that information of a different, e.g., more comprehensive, type is given, and the frame setter indicates that the information actually provided is restricted to the particular dimension specified. (Krifka & Musan 2012:32)

In Zaar, this difference in the management of information is paralleled by a syntactic difference which confirms that topics and frame-setters belong to different functional levels:

topics are pragmatic, belong to Information Structure, whereas frame-setters belong to the (micro-)syntactic structure. If we compare examples (36) and (38), we observe that no adverb (such as *ɗáni*, 'there') need modify the verbs *tu*, 'meet'. This is being already done by the left-dislocated adjunct *ɗa gìp kìmsə́y*, 'in Kimsə'.

(38) **ɗa gìp kìmsə́y** <+ *kə ɫə́ tutə //= kə fu:tə //*

ɗa	*gìp*	*kímsə*	*-í:*	*kə*	*ɫə*	*tu*	*=tə*
at	inside	Kimsə	-DIST	2SG.AOR	go	meet	3S.OBJ

kə	*fu:*	*=tə*
2SG.AOR	tell	3S.OBJ

'*In Kimsə* <+ you go meet him //= (and) tell him.' (Boys_B_188)

The same is true for example (39) where the frame-setter **ɗaŋgənín tsə́n**, 'right now' is dependent on the verb *ɲom*, 'wrestle', and this has no temporal adjunct as a dependant modifier inside the nucleus.

(39) *á wû tu ["tô:" <* **ɗaŋgənín tsə́n** *<+ tá ɲôm tə́ kádì //] //*

á	*wul*	*tu*	*tô:*	*ɗangəní*	*tsəní*	*tá*	*ɲom*	*tə́*	*kádì*
3SG.AOR	say	comp	well	now	like_this	3PL.FUT	take	with	dog

'He said ["well" < *right now* <+ he will wrestle with Dog. //] //'
(Hyena_S1_282)' (Mbrt_S1_410)

The adjunct is dependent on the verb it modifies. It contributes to the semantic component of the nucleus by restricting the circumstantial scope of its referential value. When a circumstantial adjunct appears in pre-nucleus position (left-dislocation, PR-Adv), it keeps a direct dependency relation with the verb, and no clitic or resumptive element is needed. This is marked in our macrosyntactic punctuation by the plus sign added to the chevron (<+). By contrast, it confirms the non-compositional nature of topics, working as a syntactic island (Cresti & Moneglia 2010:34–38), which is indicated in our annotation by a simple chevron

(<). Clausal adjuncts, whether circumstantial (PR-Cls, cf. (40)) or conditional (PR-Cnd, cf.

(41) share these properties with adverbial adjuncts.

(40) **ɗam mə ɬə́ tulí:** <+ ^sé: mə ɬə́ tu::: è: gàri gón //=

ɗan	mə	ɬə	tul	-í:	sé:	mə	ɬə	tu	è:	gàri	gón
as	1SG.AOR	go	arrive	RES	then	1SG.AOR	go	reach	FILL	town	QLT

'As I arrived <+ *then I reached er... a village. //='* (Boys_A_151)

(41) **yá: mǎni** <+ wò ɬyan wàhála > áy //

yá:	man	-i	wò	ɬya	-ni	wàhála	áy
3SG.COND	come	SPCF	3SG.FUT	drink	-INCH	suffering	eh

'If she comes <+ *she will suffer* > *indeed. //'* (Boys_B_289)

3.1.3 Clefting

Inspired by Higgins' seminal work on English (Higgins 1979), a large literature has been

devoted to the study of clefts and their equivalent in languages of the world (e.g. Geluykens

1984; Declerck 1988; Hedberg 2000; Lambrecht 2001; den Dikken 2006; Hedberg & Fadden

2007; Gundel 2008), including an early work on Hausa, the largest and best studied language

of the Chadic family (McConvell 1973). Like Hausa and many Chadic languages, Zaar

exhibits constructions that are related to the English cleft structures. The basic cleft structure,

also called "it-Cleft", e.g. *'It was* CHICKEN WINGS *that Peter ordered for lunch.'* is defined by

Hartmann & Veenstra (2013:1) as follows:

> *The term cleft describes a specific syntactic pattern which serves to separate a*
>
> *discourse prominent constituent structurally from the rest of the clause. [...] In*
>
> *its classical form, a cleft is a bi-clausal copulative construction consisting of*
>
> *an impersonal pronoun (the cleft pronoun), a copular verb, the informationally*
>
> *prominent phrase (the cleft phrase) and an embedded relative clause (the cleft*
>
> *clause).*

The sentence 'It was CHICKEN WINGS that Peter ordered for lunch.' can thus be analysed as :

(42) 'It   was   CHICKEN WINGS   (that) Peter ordered for lunch'
       (Cleft Pronoun)   (COP)   Clefted Constituent   Cleft Clause

Over the years, this definition, directly inspired by generative syntax studies of the English

language has to take into account variations due to languages that don't have a copula or a

cleft pronoun, as some languages lack expletive subjects or a copula, or both. (Gundel

2008:70) If the cleft pronoun is absent, one ends up with an "it-Cleft" structure with no 'it' in

it. This is the case in Zaar which, in its "classical cleft" uses copulas without an expletive

pronoun (e.g. (43), (44)), and can even omit the copula altogether (45). Two copulas are used

in Zaar for clefting, with the meaning '(it) is X': X *kən* (COP2, the most frequent); and *nə* X

(COP1).

(43)   *"tô:" < tə́ yîsə́ŋə́y tu [**kyâ:ŋ** >+ mbwá:tə //] //*

       tô: tə́   yisə́ŋ   -í: tu   kyá:ni   kən   mbwa: =tə
       DM   3PL.AOR   know   RES   COMP   2S.IDP   COP2   shoot   3S.OBJ
       '"Well" < they know that [*(it) is YOU* >+ (who) shot it. //]//' (Hunt_Har_047a)

(44)   ***nə ɓərtín** >+ ka ɓəl > fá: //*

       nə   ɓərti -ín   ka   ɓəl   fá:
       COP1   root   PROX   2SG.FUT   dig   indeed
       '(It) is THIS ROOT >+ (that) you will dig > indeed. //' (Moral_Har_069)

In (45), no copula is used for the cleft structure

(45)   *"â:" < dzàŋ là:dì má: <+ **kakáp** >+ má gè:wàyéy //*

       â: dzàŋ   lá:dì   má: kakáp   má   ge:wayé   -í:
       ah   day   sunday   even   everywhere   1PL.FUT   walk_around   RES
       '"Ah" < on Sunday indeed <+ *(it is) EVERYWHERE* >+ (that) we will stroll. //'
       (Girls_A_010)

Zaar also possesses wh-Clefts, also called pseudo-Clefts, where the cleft clause is a free

relative clause, which appears in sentence initial position: '*What Peter ordered for lunch was*

CHICKEN WINGS.' Example (47) below illustrates the structure in Zaar with the *nə* (COP1)

copula:

(46)   ^àmá: mə́n yó:ɗan ʧǎ: fî <+ **nə mə́n mársəŋ** //

   àmá: mə́n   yó:ɗan   ʧǎ: fî   nə   mə́n   mársəŋ

   but   people   which   3PL.ICPL   do   COP1   people   Lusa

   'But the people who did it <+ were THE PEOPLE OF LUSA. //' (Cal_Har_010)

NB: The it-Cleft equivalent of (46) would be 'But it was THE PEOPLE OF LUSA who did it.'
Cleft structures in Zaar correspond a single intonation constituent with no internal prosodic
break. This is paralleled by a close monosentential syntactic integration of cleft structures.
The dependency relationship of the clefted constituent is preserved and no clitic or lexical
duplication is needed. In (43), the cleft clause *mbwá: tə*, 'shoot it' has no subject clitic
standing for the clefted element *kyâ:n*, 'you' nor does any adverb or lexical equivalent stand
for *kakáp*, 'all' in (45). In (47) below, no COD clitic stands for the clefted element *gí:*, 'this'.

(47)   *"tô:"* < **gí:** >+ *tə̀tàyá: fû:mí ʧ́ĩ: //*

   tô: gí: tə̀tàyá: fu: =mí   ʧík   -í:

   DM   DIST   3PL.REM.ICPL   tell   1PL.OBJ   thus   DIST

   '"Well" <it is THIS >+ (that) they used to tell us like that. // (lit. THIS >+ they
   told us like that. //)' (Moral_Har_088)

3.1.4   Syntactic representation

The differences in the properties explored in the previous section can be neatly represented
using dependency graphs, as developed in the Raphsodie *Protocol for micro-syntactic coding*
(Kahane et al. 2013) and the annotating tool *Arborator* (Gerdes 2013). The tagging of
peripheries, discourse markers, etc. has been adapted to account for the properties described
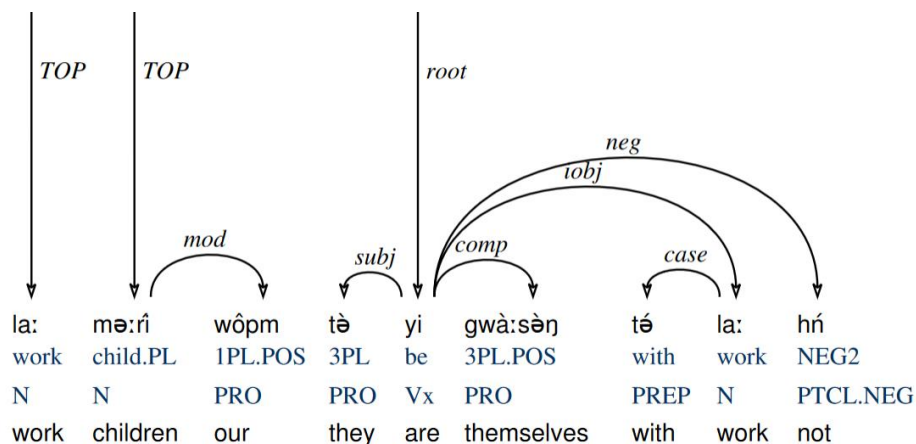in Zaar. The meaning of the tags will be given when they first appear.

TOPICS

Topics are represented as independent of the root, e.g. the defective verb *yi (tə́)*, 'have' (lit.
'be (with)') in (35). The two topics have been labelled as TOP in the graph.

(35) *la: < mə:ríwôpm < ʧi gwà:sə̀ŋ tə́ la: hń //*

‘As for work < our children < they themselves don't have any work.’

(Wom_B_221)

TOP	TOP			root				neg

la:	mə:rî	wôpm	tə̀	yi	gwà:sə̀ŋ	tə́	la:	hń
work	child.PL	1PL.POS	3PL	be	3PL.POS	with	work	NEG2
N	N	PRO	PRO	Vx	PRO	PREP	N	PTCL.NEG
work	children	our	they	are	themselves	with	work	not

[*TOP*: topic; *mod*: modifier; *subj*: subject; *comp*: complement; *iobj*: indirect object; *case*: case marking (the *case* relation is used for any case-marking element which is treated as a separate syntactic word, including prepositions, postpositions and clitic case markers); *neg*: negation]
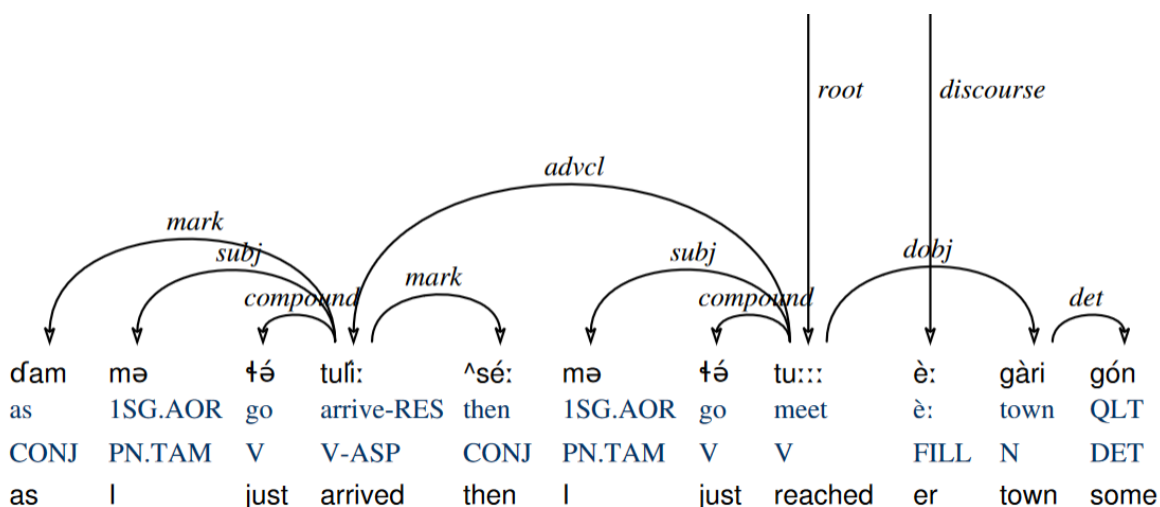
FRAME-SETTERS

Frame-setters are represented as dependants on the root of the graph, e.g. the compound verb *ɬə́ tu,* ‘arrive’ in the case of (48) below.

(48) *ɗam mə ɬə́ tulí: <+ ^sé: mə ɬə́ tu::: è: gàri gón //=*

‘When I arrived <+ then I reached er... a village. //’ (Boys_A_151)

| ɗam	mə	ɬə́	tulí:	^sé:	mə	ɬə́	tu:::	è:	gàri	gón
as	1SG.AOR	go	arrive-RES	then	1SG.AOR	go	meet	è:	town	QLT
CONJ	PN.TAM	V	V-ASP	CONJ	PN.TAM	V	V	FILL	N	DET
as	I	just	arrived	then	I	just	reached	er	town	some

[*mark*: marker (a word linking a finite clause subordinate to another clause); *subj*: subject (accounts for the function of the Person and Number index in the PN.TAM morphological complex); *compound*: verb compounding (in Zaar, verb compounding accounts for a Serial Verb Construction); *advcl*: adverbial clause modifier; *dobj*: direct object; *det*: noun determiner]

CLEFTS

As we have seen in the presentation of clefts in Zaar, copulas can be omitted in nominal predications, as in example (48). This justifies the analysis of nominal predication in the Universal Dependency grammar where the copula is not the head of the clause but rather the dependent of a lexical predicate. Such an analysis is motivated by the fact that many languages often or always lack an overt copula in such constructions. (de Marneffe et al. 2014). In this analysis of cleft structures, the cleft clause is a dependent of the cleft phrase, which is the root of the graph.

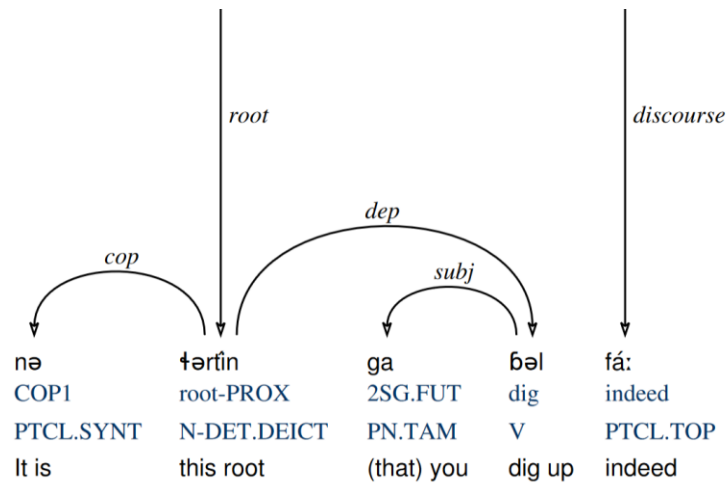(47)　　*"tô:" < **gí:** >+ tə̀tàyá: fû:mí ʧí: //*

　　　　'"Well" < *it is THIS* >+ that they used to tell us like that. //' (lit. THIS, they told us like this.) Moral_Har_088
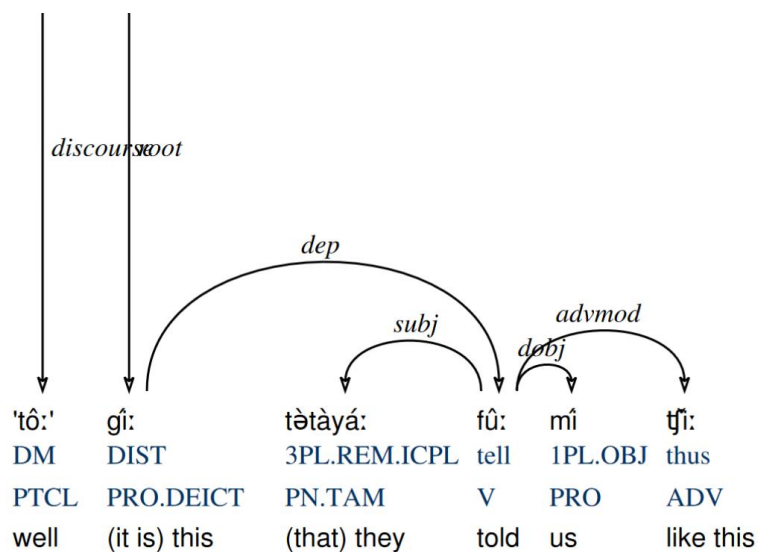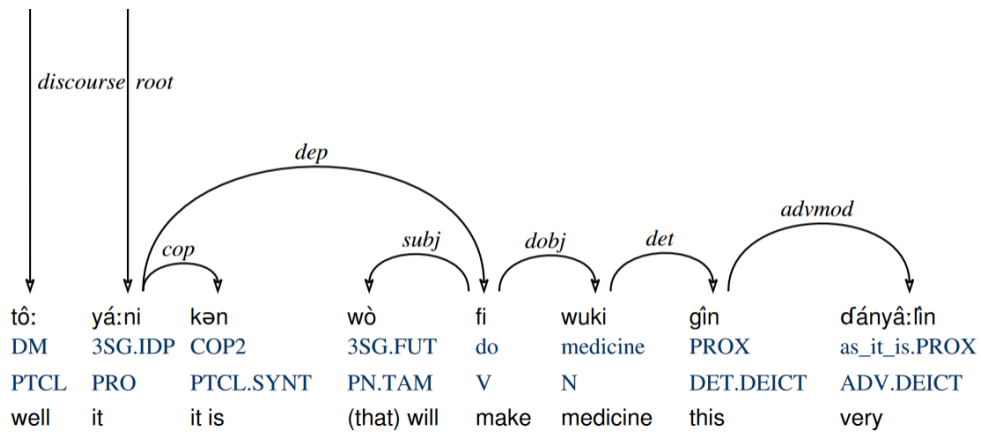

[*discourse*: interjections and other dialogic peripheral elements, e.g. phatic, allocutive, expressive, etc; *dep*: dependent (dependency of the cleft clause on the

clefted phrase); *subj*: subject (cf. ex. 49); *dobj*: direct object; *advmod*: adverbial modifier (a non-clausal adverb or adverbial phrase that serves to modify the meaning of the word)]

When the copulas *nə* or *kən* are used, they are represented as dependents of the lexical predicate, i.e. the clefted phrase, as in examples (45) and (50) below.

(44)　　**nə ɬərtín** *>+ ka ɓəl > fá: //*

　　　　'(It) is THIS ROOT >+ (that) you will dig > indeed. //' (Moral_Har_069)

nə	ɬərtɨn	ga	ɓəl	fá:
COP1	root-PROX	2SG.FUT	dig	indeed
PTCL.SYNT	N-DET.DEICT	PN.TAM	V	PTCL.TOP
It is	this root	(that) you	dig up	indeed

(48)　　*"tô:" < **yá:ni kən** >+ wò fi wuki gín dányâ:lín //*

　　　　'"Well" < it is THIS >+ (that) will make this very medicine. //'

　　　　(Moral_Har_076)

tô:	yá:ni	kən	wò	fi	wuki	gîn	dányâ:lin
DM	3SG.IDP	COP2	3SG.FUT	do	medicine	PROX	as_it_is.PROX
PTCL	PRO	PTCL.SYNT	PN.TAM	V	N	DET.DEICT	ADV.DEICT
well	it	it is	(that) will	make	medicine	this	very

4   Conclusion

In this paper analysing peripheries in relation with syntax and information structure in Zaar, a Chadic language spoken in Nigeria, we have argued a minimal annotation representing in a simple and concise way the interface between information structure and syntax was essential to retrieve meaningful data. The article uses the concept of macrosyntax, based on illocutionary units, for this new level of annotation using existing morphosyntactic tiers in Elan. With the corresponding annotation script, a pilot 90 min (15,000 words) corpus has been annotated and a preliminary study of peripheries in this language has been done on this annotated corpus. We have argued that, although topics and frame-setters share the same intonation pattern, their syntactic properties call for a specific syntactic representation for which we have used a system adapted from the Universal Dependency Grammar. Some concluding comments can be done concerning the system introduced in this paper to annotate the information structure of Zaar, and how this structure is patterned in the language. I have chosen this punctuation, and developed a corresponding set of tags bearing in mind that it should be as theory-neutral as possible in order to implement a genuine bottom-up methodology, with a heuristic aim in mind, and the hope the results can be used for typological comparisons. Another quality of this system of annotation is related to the fact that the notion of stacks accounts easily and intuitively for disfluencies, discontinuities and ellipses, and is perfectly adapted to the restitution of the oral flow. Despite the apparent accidents, interruptions and ellipses, the restitution of the stacks proves that meaning, syntax and information progress and develop like the fugues and counterpoints of a musical score, which a description limited to the boundaries of a canonical grammatical sentence has been unable to account for.

Finally, in the way Zaar shapes sound into meaning with the help of intonation, syntax and semantics, it appears that the left periphery is dominant and clefts are a device that is all the

more meaningful as it is sparsely used. The three components of Zaar Illocutionary Units come forth with a clear specialisation: the pre-nucleus establishes the frame/ground/site around the speaker's point of view; the nucleus carries the action/opinion, etc. in relation to the site; the post-nucleus seeks the hearer's approval, reactions or comments.

REFERENCES

Andersen, Hanne Leth & Henning Nølke (eds.). 2002. *Macro-syntaxe et macro-sémantique: actes du colloque international d'Århus, 17-19 mai 2001*. Berne: Peter Lang.

Berrendonner, Alain. 1990. Pour une macro-syntaxe. *Travaux Linguistiques de Gand*(21). 25–36.

Berrendonner, Alain. 2011. Unités syntaxiques & unités prosodiques. *Langue française* n°170(2). 81–93.

Blanche-Benveniste, Claire, Mireille Bilger, Christine Rouget, Karel van den Eynde & Piet Mertens. 1990. *Le français parlé: études grammaticales*. Paris: CNRS.

Caron, Bernard. 2015. Tone and Intonation. In Amina Mettouchi, Martine Vanhove & Dominique Caubet (eds.), *Corpus-based Studies of Lesser-described Languages. The CorpAfroAs corpus of spoken AfroAsiatic languages*, 43–60. Amsterdam-Philadelphia: John Benjamins.

Chafe, Wallace L. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View. In C. Li & Sandra A. Thompson (eds.), *Subject and Topic*, 25–56. New York/San Francisco/London: Academic Press.

Chanard, Christian. 2014. *ELAN-CorpA-V4.7.3*. http://llacan.vjf.cnrs.fr/res_ELAN-CorpA.php.

Cresti, Emanuela. 1999. Force illocutoire, articulation topic/comment et contour prosodique en italien parlé. *Faits de langues* 7(13). 168–181. doi:10.3406/flang.1999.1250.

Cresti, Emanuela. 2000. *Corpus di italiano parlato*. . 2 vols. Firenze, Italie: Accademia della Crusca.

Cresti, Emanuela & Massimo Moneglia (eds.). 2005. *C-ORAL-ROM: integrated reference corpora for spoken Romance languages*. (Studies in Corpus Linguistics v. 15). Amsterdam ; Philadelphia, PA: J. Benjamins.

Cresti, Emanuela & Massimo Moneglia. 2010. Informational patterning theory and the corpus-based description of spoken language: The compositionality issue in the topic-comment pattern. In Massimo Moneglia & Alessandro Panunzi (eds.), *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*, 13–45. Firenze, Italie: Firenze University Press.

Declerck, Renaat. 1988. *Studies on Copular Sentences, Clefts, and Pseudo-clefts*. Cornell University Press.

Dikken, Marcel den. 2006. Specificational copular sentences and pseudoclefts. A case study. In Martin Everaert & Henk C. van Riemsdijk (eds.), *The Blackwell companion to syntax*, vol. IV, 292–409. Malden, MA, USA: Blackwell Pub.

Geluykens, Ronald. 1984. *Focus phenomena in English: an empirical investigation into cleft and pseudo-cleft sentences.* Wilrijk, Belgium: Universiteit Antwerpen.

Gerdes, Kim. 2013. Collaborative Dependency Annotation. *Proceedings of the Second International Conference on Dependency Linguistics (DepLing 2013)*, 88–97. Prague: Matfyzpress.

Gundel, Jeanette K. 2008. Contrasting perspectives on cleft sentences. In María de los Ángeles Gómez-González, J. Lachlan Mackenzie & Elsa González Álvarez (eds.), *Languages and cultures in contrast and comparison*, 69–8. Amsterdam, Pays-Bas.

Hartmann, Katharina & Tonjes Veenstra. 2013. Introduction. In Katharina Hartmann & Tonjes Veenstra (eds.), *Cleft structures*, 1–32. Amsterdam: J. Benjamins.

Hedberg, Nancy. 2000. The Referential Status of Clefts. *Language* 76(4). 891–920.

Hedberg, Nancy & Lorna Fadden. 2007. The information structure of it-clefts, wh-clefts and reverse wh-clefts in English. In Nancy Hedberg & Ron Zacharski (eds.), *The Grammar-Pragmatics Interface: Essays in honor of Jeanette K. Gundel*, 49–76. Amsterdam: J. Benjamins.

Higgins, Francis Roger. 1979. *The pseudo-cleft construction in English*. New York: Garland.

Kahane, Sylvain, Kim Gerdes, Paola Pietrandrea & Christophe Benzitoun. 2013. *Protocol for micro -- syntactic coding*. http://www.projet-rhapsodie.fr/plus/tutoriels/doc_download/201-protocol-for-micro-syntactic-coding.html.

Kahane, Sylvain & Paola Pietrandrea. 2012. La typologie des entassements en français Kahane. vol. 1. SHS Web of Conferences. http://www.shs-conferences.org/.

Krifka, Manfred & Renate Musan. 2012. Information structure: Overview and linguistic
    issues. In Manfred Krifka & Renate Musan (eds.), *The Expression of Information
    Structure*, 1–43. Berlin, Boston: De Gruyter Mouton.
    http://www.degruyter.com/view/product/177467 (2 May, 2016).

Lacheret, Anne, Paola Pietrandrea & Atanas Tchobanov. 2014. Rhapsodie: a Prosodic-
    Syntactic Treebank for Spoken French. http://hal.upmc.fr/hal-00968959/document (23
    March, 2016).

Lambrecht, Knud. 1994. *Information structure and sentence form : topic, focus, and the
    mental representations of discourse referents*. Cambridge England ; New York:
    Cambridge University Press.

Lambrecht, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics*
    39(3). 463–516.

Marneffe, Marie-Catherine de, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D.
    Manning, Ryan McDonald, Joakim Nivre, et al. 2014. Universal Dependencies.
    Online documentation (Version 1). http://universaldependencies.org/ (16 June, 2016).

McConvell, Patrick. 1973. Cleft sentences in Hausa: a syntactic study of focus. London:
    University of London Doctoral dissertation.

Mettouchi, Amina, Martine Vanhove & Dominique Caubet (eds.). 2012. *The CorpAfroAs
    Corpus. ANR CorpAfroAs: a Corpus for Afro-Asiatic languages*.
    http://corpafroas.huma-num.fr/.

Mettouchi, Amina, Martine Vanhove & Dominique Caubet (eds.). 2015. *Corpus-based
    Studies of Lesser-described Languages. The CorpAfroAs corpus of spoken AfroAsiatic
    languages*. Amsterdam: Benjamins.

Schultze-Berndt, Eva. 2013. About the shifty notion of contrast. Identifying subtypes of topics
    in corpus data of two Australian languages. *Labex TCA-ISGR, Llacan (2013-11-12)*.
    Villejuif.

Simard, Candide. 2014. Another look at right-detached NPs. In Aicha Belkadi, Kakia
    Chatsiou & Kirsty Rowan (eds.), *Proceedings of Conference on Language
    Documentation and Linguistic Theory 4*. London: SOAS.
    www.hrelp.org/eprints/ldlt4_17.pdf.

Sperber, Dan & Deirdre Wilson. 1986. *Relevance: communication and cognition*. Oxford: B.
    Blackwell.