

To appear in *Faits de Langues* n°52/1 (2021)

From a corpus-based to a corpus-driven definition of clefts in Kabyle (Berber): Morphosyntax and prosody

Amina Mettouchi*

1. Introduction¹

Clefts in Kabyle have a syntactic form that appears to be distinctive: an NP, preceded by a copula, and followed by a relativizer and a clause. However, if we strictly follow that morphosyntactic characterization, and run the corresponding query on an electronic corpus, a number of structures are retrieved that have various functions and meanings.

Working with the assumption that differences in function and meaning are correlated to differences in form, I will examine the formal correlates of those differences, and show that they are of a prosodic nature. I will ultimately propose a definition of clefts based on the interaction of formal means belonging to morphology, syntax and prosody.

In doing so, I will not simply claim that prosody plays a role in the interpretation of clefts, which is a widely held view. The perspective adopted in this paper is more radical, in the sense that I consider that there is no such structure as a cleft in Kabyle, without precise, defining prosodic features that I will investigate here. Those prosodic features are as fundamental to the definition of clefts as the morphosyntactic features usually discussed in the literature.

Moreover, the corpus-oriented aspect of the paper is not an end in itself, but rather a way of systematically testing linguistics claims about forms or constructions: a good part of a linguist's work consists in precisely describing and analyzing constructions and their functions in a testable and falsifiable way. Even if coding properties are not sufficient to characterize a construction, and one needs behavioral properties as well, the latter can only be tested once the coding properties of the form or construction have been defined. This is what this paper aims to do for Kabyle clefts.

Following a presentation of Kabyle, and of the corpus, I will discuss corpus findings based on a strictly morphosyntactic definition of clefts and show that the search retrieves more constructions than it should, and leaves aside some structures that have the same function as that of canonical clefts. In a third part I will show that the introduction of prosodic components to a language-internal definition of clefts in Kabyle allows the delimitation of a

* LLACAN, EPHE –PSL. Mail: amina.mettouchi@ephe.psl.eu

¹ My gratitude goes to all the Kabyle speakers I have recorded over the years, and in particular Tounsia Rabia, Cherifa, Yamina-Zahra, and Zouina Mettouchi. I thank the anonymous reviewers for constructive comments on this paper, as well as audiences in Paris and Cologne for their feedback. In particular, I would like to thank Nikolaus Himmelmann for his insightful comments on an earlier version of this paper. I am extremely grateful to citizens of France and Europe who, through their taxes, allow the maintenance of public funding for research, thus facilitating the conduct of sustainable scientific work.

homogeneous group of structures sharing a function, that of narrow focus. I will label that construction the cleft, considering that other structures are of a different nature, even if they involve similar morphosyntactic materials: they are compositional structures involving a non-verbal predicate, and a relative clause, and display functional variations on more or less loosely associated assertions and presuppositions.

2. General Background

2.1. Kabyle

Berber languages are spoken in northern Africa, in a zone delimited by the Atlantic Ocean to the West, the Mediterranean to the North, the oasis of Siwa (Egypt) to the East, and the Sahel to the South. Those languages constitute a family within the Afroasiatic phylum. Well-known members of the family are, among others, Kabyle (spoken in northern Algeria), Tachelhuyt (Shilha) (spoken in southern Morocco), and Tamashek and Tahaggart (also called Tuareg), spoken in southern Sahara.

Kabyle has about four million speakers in the north of Algeria. The variety investigated in this paper is a Western-Central one, spoken in the village of Ait Ikhlef, close to the town of Bouzeguene. I collected all the data on fieldwork between 1992 and 2019.

In Kabyle, as in all Berber languages, a minimal predication consists either of a verb and its bound personal pronoun, or of a non-verbal predicate (Chaker 1983, Mettouchi 2017). In addition to this core, the clause may contain noun phrases, and prepositional phrases, as well as adverbs. Within noun phrases, modifiers follow the modified constituent. The language has two genders (masculine and feminine) and two numbers (singular and plural), marked on adjectives, on nouns, and on pronominal affixes and clitics hosted by verbs, nouns and prepositions. It also has two states (absolute and annexed), marked on nouns.

2.2. Corpus

The data on which the study is based have been collected over the course of almost three decades of fieldwork in Kabylie (Algeria). The annotated online corpus on which the queries have been tested and refined is a subpart of that body of data, it contains one hour (23 336 morphemes) of transcribed, segmented, annotated and translated narratives and conversations, collected in the field between 2007 & 2011. Examples in this paper are taken from the one-hour corpus.

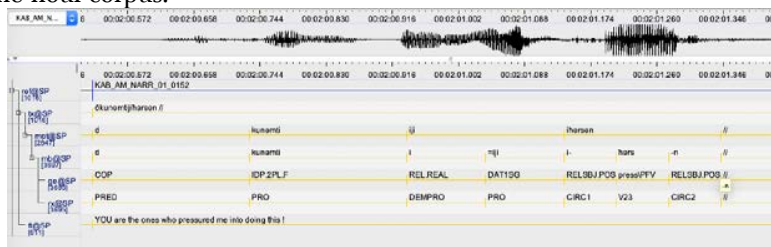


Figure 1: Layout of the Kabyle Corpus (illustrating example [2] in 3.1.)

As shown in Figure 1, morphosyntactic annotation is displayed on two tiers, “ge” and “rx”, allowing the automatic retrieval of complex queries based on forms.² Queries are not just tools for the retrieval of statistical tendencies or collocations. They are, crucially, a way of testing hypotheses: a formal definition for clefts, translated into an automatic query, should enable the retrieval of all and only the clefts in the corpus. Retrieving less than all the clefts, or retrieving structures that look like clefts but have a different function altogether, means that the formal definition is inaccurate: either it is erroneous, or it lacks a feature that is crucial but has not been identified yet. Studying all the departures from the expected list of hits helps refine the original formal definition, in order to reach an optimally accurate characterization of the phenomenon under study.

The data were segmented into intonation units³ on the basis of native speaker perception, and acoustic control with Praat.⁴ An intonation unit is a segment of speech that has a coherent intonation contour, and is delimited by its boundaries, which bear a ‘boundary tone’. In Kabyle, Intonation Units are marked by one or more of the following cues:⁵

Main external cues:

- (1) final lengthening;
- (2) initial rush;
- (3) pitch reset;
- (4) pause;
- (5) creaky voice.

Main internal cues:

- (1) declination;
- (2) tonal parallelism, or isotony.

The total number of non-pausal intonation units for the online corpus is 2,671. With breath intakes and silent pauses, the total number of units is 3,974.⁶ All examples in the paper and their larger contexts, as well as their corresponding sound files can be checked in the online corpus.⁷

3. Clefts and cleft-like structures: Morphosyntactic characterization

The heuristic concept of ‘cleft’ with which I started my investigations is loaded with as little assumptions as possible: I start from the corpus search of

² For the syntax of queries using regular expressions, see https://llacan.cnrs.fr/fichiers/manuels/fELAN/ELAN-CorpA_search.pdf

³ Annotated as /: non-terminal boundary; //: terminal boundary.

⁴ <http://www.fon.hum.uva.nl/praat/>.

⁵ See (Izre’el & Mettouchi 2015) for more information on the segmentation of the CorpAfroAs corpus of spoken Afroasiatic languages, of which the Kabyle corpus is a part.

⁶ Each unit has been numbered following a precise methodology: ISO code of the language, initials of the collector/annotator, genre (NARR is for narrative, CONV for conversation), number of the file/recording in the corpus, number of the unit in the file/recording. Thus, all examples are easily found in the corpus.

⁷ <http://dx.doi.org/10.1075/scl.68.website>.

a construction that involves a relative-like component, as well as a non-verbal predicate composed of a copula and a noun phrase, and has an interpretation in terms of focus. This ‘canonical’ cleft is typically the one speakers offer in elicitation tasks when prompted for contrastive focus (Mettouchi 2003a, 2003b).

The difference between Mettouchi (2003a, 2003b) and the approach I have undertaken in this paper, is that in the present study, this “canonical cleft” is not the stable object of the study, but only its starting point. My question now is how I can formally delimit a construction that has a specific function, starting from that heuristic concept. My aim is to be driven by what I actually encounter in the corpus, rather than by prior assumptions on constructions and their functions: the initial construction, which I refer to as ‘canonical cleft’ for want of a better term, is bound to be redefined by the corpus study. The corpus study itself is not a quantitative search of occurrences of the ‘canonical cleft’, nor is it a search on the pragmatic values or the contexts of use of that ‘canonical cleft’. The aim is the retrieval, in my spontaneous data, of all and only the instances of a construction defined by the association of a function and form, which I will call “cleft”. The ultimate goal of the study is to provide a more accurate and complete definition of that construction in Central-Western Kabyle.

In the literature on the Berber language family, focus constructions have mostly been studied in a syntactic perspective, in various frameworks, sometimes functional and historical (Basset 1952; Galand 1957, 1984, 2002; Leguil 1984), sometimes formal (Ennaji and Saddiqi 1986). Most of those studies are based on Tashelhit data (Southern Morocco). Pragmatic perspectives are often included in the studies, but although prosody is mentioned as an important element in the description of clefts, or more generally focus constructions, it is not formally investigated, except in Louali & Mettouchi (2002), Mettouchi (2003a, 2003b), Mettouchi *et al.* (2004). Several papers about the information structure of various Berber languages in general have been published, as well as several papers on the intonation of some illocutionary types in various Berber languages, however, none of them to our knowledge have specifically broached the subject of the intonation of clefts, or contrastive focus, apart from those quoted above.

3.1. *First type of hits: the canonical cleft*⁸

Running a query aimed at retrieving the sequence {Copula (*d*) – Noun Phrase – Relativizer (*i* or *ara*) – verbal clause}, that represents the canonical structure for clefts, as in (1), yields a number of hits, such as examples (2) to (4):

⁸ Examples are preceded by general information about their context in the recording, and by a translation of the paragraph in which they occur. The first line is in prosodic words, and IPA, it closely follows the sound files. To allow separation into morphemes, given extensive assimilation processes in the realization of underlying segments, a second line is provided, in grammatical words (further breakable into morphemes), and broad phonetics transcription, with emphatics represented orthographically with a subscribed dot. The reference of each example (see footnote 6) allows the reader to search the corpus online.

Context: Mother Nuja is an ogress, she wanted to eat Amar but he tricked her into falling asleep and burnt her daughters to death.

Paragraph from which the example is taken: Mother Nuja woke up and realized her daughters had been burnt to death. She broke into a run, reached a river, she was running and the horses were with her, she said “May God punish you Amar! You miserable boy! *You killed my seven daughters!* You to whom I offered hospitality, wait for me!”⁹

- (1) $\partial s\partial\beta\text{f}\alpha j\partial s\text{si } i\partial\eta n y i \partial^{\text{c}} //$
 d s\partial b f \alpha j \partial s s i i t n y i d //
 COP seven daughter\PL REL.REAL kill\PFV:SBJ2SG //
 “You killed MY SEVEN DAUGHTERS!” (KAB_AM_NARR_02_MIDGET_460)

Context: The seven little orphaned girls are tricked by a neighbour into asking their father to remarry. The father refuses again and again, afraid that his new wife might abuse the girls. But they insist.

Paragraph from which the example is taken: He said “All right my daughters, we’ll do as you wish, the choice is yours. If she is good to you, you’ll have been right, if not, (remember) *you (are the ones who) put pressure on me* until I agreed to remarry!”

- (2) $\partial k u n \partial m t i j i h \partial r s \partial n //$
 d k u n \partial m t i i = j j i i h \partial r s n //
 COP IDP2PLF REL.REAL=DAT1SG press\PFV:RELSBJ //
 “YOU put pressure on me.” (KAB_AM_NARR_01_0152)

Context: The three speakers are reconstructing the genealogy of the family. They do not agree on the number of wives their ancestors had.

Paragraph from which the example is taken: Speaker 1: they said he married seven women; Speaker 2: Hajj Tahar didn't marry seven (women); Speaker 1: he did!; Speaker 2: certainly not! *It's Hajj Rabah who married seven (women)*

- (3) $\partial \partial l h \partial d \zeta r \partial b \partial h i g u y \partial n s \partial \beta f \alpha //$
 d l h \partial d \zeta r \partial b \partial h i j u y \partial n s \partial b f \alpha //
 COP pilgrim Rabah REL.REAL take\PFV:RELSBJ seven //
 “It's Hajj Rabah who married seven (women)!”
 (KAB_AM_CONV_01_SP2_047)

Context: The speaker talks about how hard life was for the Kabyle woman in the old days, and how power relationships were distributed.

Paragraph from which the example is taken: “The mother-in-law, she would control everything, the daughter-in-law couldn't express herself, the mother-in-law would decide how much flour you could have, she would fill

⁹ Abbreviations: ABS absolute state; ABSV absolutive pronominal paradigm; ANN annexed state; AOR aorist; ASSOC associative; SHREF shared reference demonstrative; COL collective; COP copula; DAT dative; F feminine; GEN genitive; HESIT hesitation; IDP independent pronoun; IPFV imperfective; M masculine; PFV perfective; PL plural; POS positive; POSS possessive pronominal paradigm; PROX proximal; SBJ subject pronominal paradigm; REAL realis; REL relator; RELSBJ subject relativization circumfix; SG singular.

A list of glosses with definitions, explanations and references can be found on https://corpafroas.huma-num.fr/Archives/KAB/PDF/KAB_AM_ALISTOFGLOSSES.PDF

you a minimal measure, only then could you go and prepare bread, only when she had prepared a little sauce could you cook, you could not go against her commands. Now... in the past we had attic shelves (where food was kept), and the mother-in-law, she was the head of the household, it was her who would gather provisions on the attic shelves. If she wanted to give you things, she would be the one to decide. If she wanted to give you dried figs, *she would (be the one to) give them to you*. But now those mothers-in-law, poor things, their status has gone, they have no power anymore.”

- (4) ðnətstsaθ aramddifkan //
 d nəttat ara=am=dd ifkan //
 COP IDP.3SG.F REL.IRR=DAT2.SG.F=PROX give\PFV:RELSBJ //
 “She would (be the one to) give them to you.” (KAB_AM_NARR_03_0710)

The structure is composed of a copula, followed by a noun phrase, and by a clausal construction, whose relative-like nature is shown by the fact that in (2) and (3) it is the subject-relativization form (an invariable form based on the third masculine singular form of the verb and suffixed with *-n*) that is used. The structure also contains the same relativizer as the one used in relative clauses (5), *i* or *ara* (depending on the realis/*de re* or irrealis/*de dicto* status of the clause). However, whereas in relative clauses the relativizer can be omitted (6), this is never the case in clefts.

Context: The speaker deplors the fact that nowadays women leave their houses and gardens in a pitiful state, they don't work like in the old days.

Paragraph from which the example is taken: “if you could see their fields, they leave them in a pitiful state, those olive-trees. Because the women of the past, they are dead. *Those women of the past who worked hard like that*, they are not here anymore. Today, women refuse to work like that. They give orders, they trim their nails, that's all.”

- (5) lxalaθnni nziç igxəðmən akken / 227 / ulaʃiθənt //
 lxalat-nni n zik i
 woman\ABS.PL.F=SHREF GEN long_ago REL.REAL
 ixdəmn akkən / ulaʃ=itnt //
 make\PFV:RELSBJ thus / NEG.EXS=ABSV3.PL.F //
 “Those women of the past who worked hard like this, they are not here anymore.” (KAB_AM_NARR_03_1107-1111)

Context: The speaker compares food in the old days, which was healthy, and processed food today, which she criticizes.

Paragraph from which the example is taken: “People had herds, they would take it to pasture, and the father would slay a kid or lamb for his children. He would hook it in the cold, there were no fridges, and take parts little by little to feed his children. The meat was cooked little by little, it was sweet. That's how delicious it was. *Nowadays we have frozen meat*, we don't know if it's beef or what, that they slew.”

- (6) θura ðaçsum ikunzilin //
 tura d aksum ikunzilin //
 now COP meat\ABS.SG.M freeze\PFV:RELSBJ //
 “Nowadays we have frozen meat.” (KAB_AM_NARR_03_0938)

The structures in examples (1) to (4) also share a function, that of expressing focus, in the sense of Lambrecht, for whom focus is “the semantic component of a pragmatically structured proposition whereby the assertion differs from the presupposition” (Lambrecht 1994: 213).

In (1) the presupposition is that it is the young hero and his seven brothers who should have been killed - but he was cleverer than the ogress and poured the boiling water on her seven daughters instead. In (2) the presupposition is that the pressure to remarry is due to the father, whereas actually it is the little girls, manipulated by their neighbour, who insist. In (3) the presupposition is asserted by another speaker in the preceding turn in the conversation: Hajj Tahar married seven women (whereas actually it is Hajj Rabah who did). In (4) the presupposition, triggered by an assessment of the situation nowadays, in which young women are the heads of their own homes, is that young brides would be allowed to decide which tasks to do, and whether they wanted to eat something, in their households. Actually, in the old days, everything was kept under lock and key by the mother-in-law, who decided about everything in the house.

More precisely, those examples are instances of narrow focus: if we take focus structure to be “the conventional association of a focus meaning [distribution of information] with a sentence form” (Lambrecht 1994:222), the focus domain in all of them is a single constituent, and not a whole predicate or sentence: in (1) it is “my seven daughters”, in (2) “you”, in (3) “Hajj Rabah”, in (4) “she”.

3.2. Second type of hits: equative clause followed by a relative clause

The abovementioned formal query also retrieves structures that contain the same formal features, but whose meaning and function are different, as in (7):

Context: The speaker talks about the simplicity of the life people had in the countryside in the old days.

Paragraph from which the example is taken: “I remember, I used to go with my grandma, what can I say, we had cows, we had sheep, we had goats, *the milk was milk from the sheep, that their children consumed*, we would go, grandma and I, I would take the village herd to pasture, we would arrive in the afternoon, we would milk the she-goat, we would drink milk from our livestock, *that life was lived on the ground, it was a life that they made with their hands, it was a life that we gave back to the earth.*”

- (7) ajfki / BI_414 / ðajfki lmal / 192 / ɪgɪsətstɛn warraw ə:: / ənsənt //
 ajfki / d ajfki n lmal /
 milk\ABS.SG.M / COP milk\ABS.SG.M GEN herd\ANN.COL /
 i təttn warraw ə:: /
 REL-REAL eat\IPFV:SBJ.3.PL.M offspring\ANN.PL.M HESIT /
 -nsnt //
 -POSS.3PL.F //
 “The milk, it was milk from their livestock - that their children consumed.”
 (KAB_AM_NARR_03_0160-0163)

Example (7) is not centred on a divergence between a presupposition (that children drank milk from other animals than sheep (e.g. cows), or that they

drank processed, as opposed to unprocessed milk), and an assertion (children drank sheep's milk, children drank unprocessed milk). Rather, it simply asserts the fact that milk came from the sheep owned by the family, just like everything in those days was done locally, rustically, in close relationship with nature and the simplicity of rural life. The mention that it is drunk by children is an afterthought, flagging a previous subtopic about the healthy food that people consumed in the old days. It is not to be interpreted in the scope of this construction. We do not have a focus construction here, and this is made very clear when we examine prosodic evidence.

Indeed, the first striking formal difference in comparison to examples (1) to (4) is that whereas the former are all contained within one intonation unit, example (7) spans two intonation units, with a 192ms-long pause in-between. Does that mean that there is a correlation between the narrow-focus interpretation of the morphosyntactic sequence in (1) to (4), and the fact that it is enclosed within one prosodic unit, whereas when it spans two intonation units, as in (7), the interpretation of the structure is necessarily different? Should the definition of clefts include that parameter?

Things are not so straightforward, and the data show that this correlation between interpretation in terms of canonical cleft, and inclusion within a single intonation unit needs to be checked further. Indeed, clefts may span two intonation units, although extremely rarely (there is only one example in my data, discussed in 4.4.). More importantly, in several cases, the morphosyntactic structure under investigation is indeed encapsulated within a single intonation unit, but does not have an interpretation in term of narrow focus: inclusion of a cleft-like morphosyntactic structure within a single intonation unit is not a sufficient condition for the structure to be an actual cleft. Example (8) is an instance of this:

Context: The speaker talks about the simplicity of the life people had in the countryside in the old days.

Paragraph from which the example is taken: "I remember, I used to go with my grandma, what can I say, we had cows, we had sheep, we had goats, the milk was milk from the sheep, that their children consumed, we would go, grandma and I, I would take the village herd to pasture, we would arrive in the afternoon, we would milk the she-goat, we would drink milk from our livestock, *that life was lived on the ground, it was a life that they made with their hands, it was a life that we gave back to the earth.*"

- (8) θamɕijθ tsamɕijθ əlqafa / tsamɕiθt ixəðmənt sifassənsənt / tsamɕijθ iddəntsarra
 ðəgwaçalnni /
 tamɕift d tamɕift n lqafa /
 life\ABS.SG.F COP life\ABS.SG.F GEN ground\ANN.SG.F /
 d tamɕift i xdəmnt s
 COP life\ABS.SG.F REL.REAL make\PFV:SBJ.3PL.F INSTR
 ifassn-nsənt / d tamɕift i=dd
 hand\ANN.PL.M-POSS.3PL.F / COP life\ABS.SG.F REL.REAL=PROX
 nttarra dg wakał-nni /
 give_back\IPFV:SBJ.1PL in land\ANN.SG.M-SHREF /
 "That life was lived on the ground, it was a life that they made with their
 hands, it was a life that we gave back to the earth." (KAB_AM_NARR_03_0173-
 0176)

It would have been easy to interpret the second and the third intonation units, in isolation, as narrow/contrastive foci: “it’s LIFE that they made with their own hands, it’s LIFE that we gave back to the earth”.

However, the actual interpretation is in terms of modification: we have a referent, *tamɣift*, which is first specified, through an equational structure, by genitive modification of the referent, ‘life of the ground’ → a life close to the ground, that of the house where the hearth used to be, where people used to sit in order to eat, and the soil outside, that people used to weed and plough and gather their subsistence from.

From there follow two further specifications of the type of life the speaker is thinking about: one which insists on the fact that there was no industrial processing, the other that everything they took from the soil they gave back to the soil: organic produce, as well as their own bodies after death.

The function of the copula which starts the second and third intonation units is to mark equation with the discourse topic *tamɣift* at the beginning of the first intonation unit. The function of the relative clauses inside those two units is to restrict the reference of the antecedent *tamɣift*. They do not express a presupposition whose confrontation with the assertion “it is (a) life” would allow the emergence of a differential element of information that would be the focus of that structure. What example (8) features is equational clauses followed by restrictive relative clauses.

Does that mean that only “context”, in the sense of a cluster of inferences based on previous utterances, general knowledge, and situational factors, can distinguish actual clefts from other structures that look like clefts but are actually composed of a main and a subordinate clause, which marks a resurfacing subtopic, as in (7), or provides referential information about the NP, as in (8)?

While I agree that some inferences and some elements of extra-linguistic knowledge are indeed necessary to interpret correctly what is produced by speakers, I consider that falling back on ‘context’ must be restricted to cases where no formal means can be found to trigger the interpretation under investigation. This is not the case for our structures, for which taking into account prosodic features is crucial, and allows the distinction between clefts and cleft-like structures to be made.

4. Prosody as a Core Coding Means for Clefts

Indeed, what clefts have in common in Kabyle, is not only their morphosyntactic structure, and the fact that they are normally inside a single intonation unit. It is also, and crucially, their prosodic pattern.

4.1. Prosodic prominence: location of the peak and relative pitch

One feature that all cleft-like structures which are instances of narrow focus share, is the F_0 peak on the relativizer, as well as a more or less pronounced compression of the pitch range on the clefted constituent (see also Mettouchi 2003a et 2003b).

This is visible in the prosodic contours of examples (1) to (3), reproduced here as (9) to (11):

- (9) ḏsəβfajəssi iθənyid̪ʰ //
 d səβfə jəssi i tnyid //
 COP seven daughter\PL REL.REAL kill\PFV:SBJ2SG //
 “You killed MY SEVEN DAUGHTERS!” (KAB_AM_NARR_02_MIDGET_460)

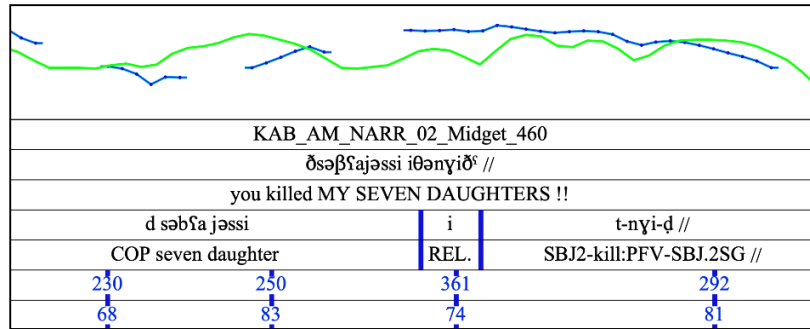


Figure 2: Example (9) [(1)]¹⁰

The values in the bottom tier before last, in this example, show that the pitch prominence of the structure (361 Hz) is on the relativizer *i*, accompanied in this case by an intensity peak, but this is not systematic (see example (11) for instance).

The same profile can be found for example (10), with a peak on the relativizer at 472 Hz.

- (10) ḏkunəmtijihərsən //
 d kunəmti i=iiji ihərsn //
 COP IDP.2.PL.F REL.REAL=DAT.1.SG press\PFV:RELSBJ.POS //
 “YOU put pressure on me.” (KAB_AM_NARR_01_0152)

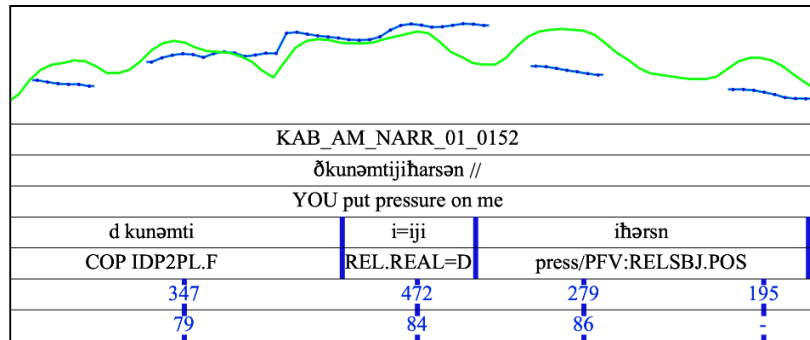


Figure 3: Example (10) [(2)]

- (11) ḏəlhadʒ rabaḥ iguyən səβfə //
 d lhədz Rabaḥ i juʁn səβfə
 COP pilgrim Rabaḥ REL.REAL take\PFV:RELSBJ seven
 “It’s Hajj Rabaḥ who married seven (women)!” (KAB_AM_CONV_01_SP2_047)

¹⁰ Pitch, also called *F*₀ (in Hertz) is represented by the blue line, whereas intensity (in decibels) is represented by the green one. On each graph, the values for *F*₀ (e.g. 230; 250) and Intensity (e.g. 68, 83) are in the bottom two tiers.

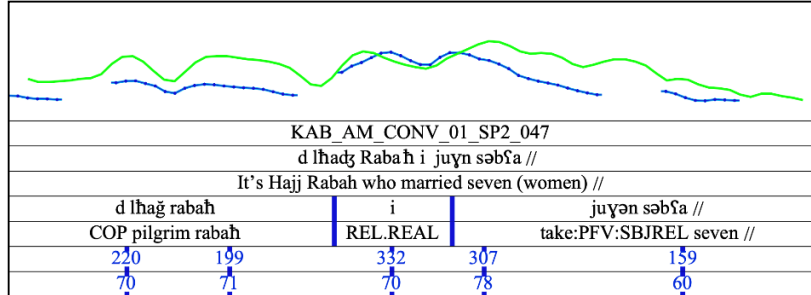


Figure 4: Example (11) [(3)]

In example (11), the main prominence is also on the relativizer, with a peak at 332 Hz. The mean pitch in the first part of the structure is 190 Hz, whereas it is 248Hz in the second. The strong compression observed for the clefted constituent in this example is proportional to the degree of polemical debate at this point in the conversation: the two women are strongly contradicting each other about who had married seven women.

Mettouchi (2003a: 146) claimed that “if the speaker who utters the cleft anticipates divergence from Speaker B, then the F₀ peak will be very high. If not, it will be medium high”. Actually, it is not a question of absolute height, but rather of relative prominence, of contrast between the F₀ value on the relativizer, and that observed on the clefted constituent. The part of conversation analyzed in Mettouchi (2003a) showed several marked contrasts in prominence, whereas for instance example (12) below, although clearly exemplifying the cleft prosodic pattern, is less markedly polemical, with a prominence on the relativizer at 238Hz, and on the clefted constituent at 176Hz.

- (12) ðnətsaθ aramddifkan //
 d nəttat ara=m=dd ifkan //
 COP IDP.3SG.F REL.IRR=DAT.2SG.F=PROX give\PFV:RELSBJ //
 “SHE would give them to you.” (KAB_AM_NARR_03_0710)

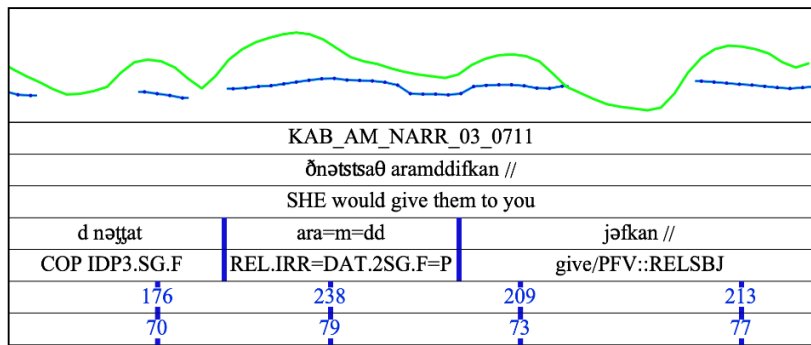


Figure 5: Example (12) [(4)]

The prosodic profile is still distinct from cleft-like (i.e. non-cleft) structures (because the F₀ peak on the relativizer is still higher than on any syllable in

the preceding constituent), but the contrast is less marked. The various types exemplified in figures 2 to 5 point to a continuum within narrow focus between pragmatically neutral narrow focus, and more polemical instances of divergence between presupposition and assertion, culminating in contrastive focus (exemplified in [11] [3]).

There is no continuum however between those cleft structures and the non-cleft structures in Figures (6) and (7). Indeed, structures that look like clefts but are actually different in function, and in form if we take prosody into account, show clearly different profiles.

- (13) ajfki / BI_414 / ðajfki lmal / 192 / igisætstæn warraw ə::: /ənsənt //
 ajfki / d ajfki n lmal /
 milk\ABS.SG.M / COP milk\ABS.SG.M GEN herd\ANN.COL /
 i tættn warraw ə::: /
 REL.REAL eat\IPFV:SBJ.3.PL.M offspring\ANN.PL.M HESIT /
 -nsnt //
 -POSS.3PL.F //
 “The milk, it was milk from their livestock - that their children consumed.”
 (KAB_AM_NARR_03_0160-0163)

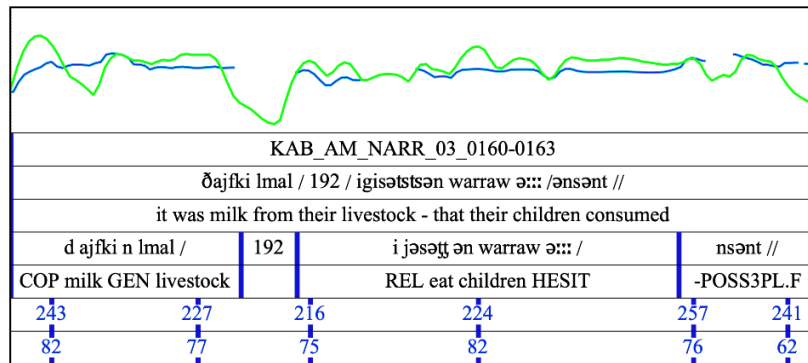


Figure 6: Example (13) [(7)]

- (14) θamfiθ tsamfiθ əlqafa / tsamfiθt ixəðmənt sifassənnənt / tsamfiθ iddəntsarra
 ðəgwaçalnni /
 tamfiθt d tamfiθt n lqafa / d
 life\ABS.SG.F COP life\ABS.SG.F GEN ground\ANN.SG.F / COP
 tamfiθt i xdəmnt s
 life\ABS.SG.F REL.REAL make\PFV:SBJ.3PL.F INSTR
 ifassn-nsənt / d tamfiθt i=dd
 hand\ANN.PL.M-POSS.3PL.F / COP life\ABS.SG.F REL.REAL=PROX
 nttarra dg wakał-nni /
 give_back\IPFV:SBJ.1PL in land\ANN.SG.M-SHREF /
 “That life was lived on the ground, it was a life that they made with their
 hands, it was a life that we gave back to the earth.” (KAB_AM_NARR_03_0173-
 0176)

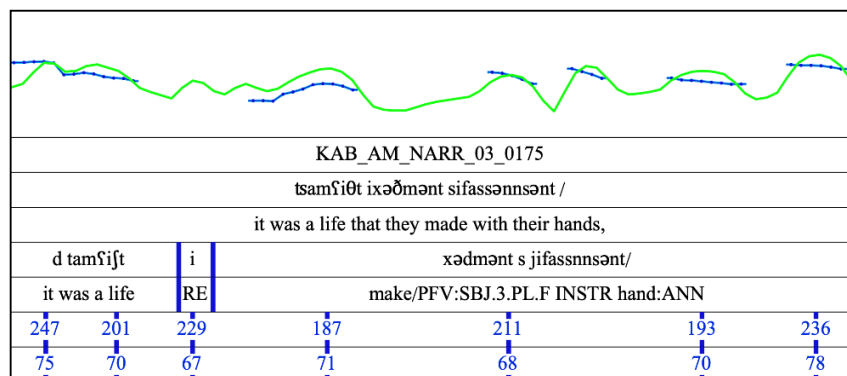


Figure 7: example (14) ([8])

In examples (13) (Figure 6) and (14) (Figure 7), the main F₀ peak is on the first part of the construction (copula followed by noun phrase, 243 Hz in [13], 247 Hz in [14]), considerably higher than the relativizer (216 Hz in [13] and 229 Hz in [14]).

4.2. Other structures with slightly different morphosyntax, same function, and same prosodic pattern

A closer look at the data shows that some structures, which don't have the same morphosyntactic form exactly, but are functionally similar and display the same prosodic pattern, had been left out by the original query.

Among them we find structures involving prepositional phrases (15), quantifiers or adverbs (16), in which there is no copula before the clefted constituent, as well as non-verbal clefted clauses (17).

Context: The conversation is about the genealogy of the family, and has reached Mahmud's generation. The speakers discuss the number of children his father had, and their names. The first speaker mentions three daughters, the second speaker remembers only two. The first speaker reminds her that the eldest daughter died in childbirth, and the second speakers remembers about all that.

Paragraph from which the example is taken: SP₁= he had five sons and three daughters

SP₂= How come, three daughters?

SP₁= three indeed! Louisa died, God bless her

SP₂= That's right I remember! There's Nadia, Fadila...

SP₁= Fadila and Nadia are the youngest

SP₂= that's right, those are Mahmud's sisters

SP₁= yes, he had three sisters. Her (his mother's) eldest daughter died, she died giving birth to a boy

SP₂=she got married and got a baby, this girl, *she was in Algiers, she lived in Algiers*, she had a house. The boy poor thing, he died.

SP₁= Yes, he died. God bless him.

- (15) ildzajr iθəlla / ildzajr iθətsɪf //
 i lzajr i tlla // i lzajr
 LOC Algiers REL.REAL exist\IPFV:SBJ.3SG.F // LOC Algiers
 i tətɪf //
 REL.REAL live\IPFV:SBJ.3SG.F //
 “She lived in Algiers.” (KAB_AM_CONV_01_SP2_279)

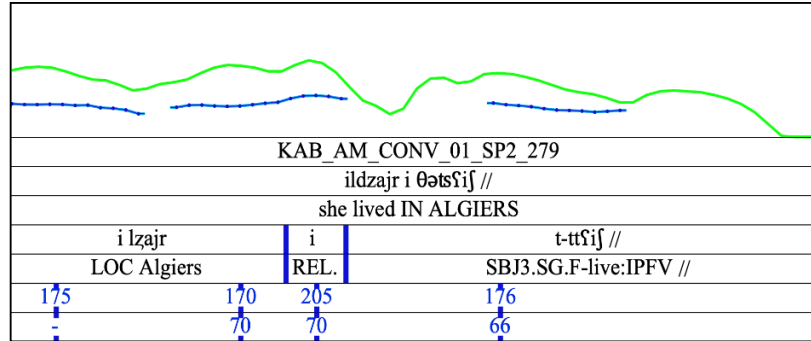


Figure 8: Example (15)

In (15), the presupposition triggered by the preceding context is that the woman under discussion lived in Ait Ikhlef, as the rest of the conversation was about people living in that village. The structure goes against that presupposition and provides a differential element in the assertion, which is that she lived in Algiers. We have a typical narrow (though not contrastive) focus.

In (16) the context is more polemical, the two main speakers have been arguing about the number of wives Hajj Tahar had, Speaker 2's initial assumption being that he had married either three or seven women. Example (16) goes against that presupposition, and provides five as the right number, in a typical contrastive focus.

Context: The two speakers are reconstructing the genealogy of the family. They do not agree on the number of wives their ancestors had. Speaker 1 goes through the names of all the wives Hajj Tahar had (her initial assumption was uncertain, she hesitated between three or seven wives).

Paragraph from which the example is taken: SP 1: (...) he married Tatahart (...)

SP2: I know none of those!

SP3= but that one, isn't she the mother of Arezki At Mhend?

SP1 (to speaker 3): Oh no, that one's name is Taatment, it's uncle Ali (who married her), may God bless him, and his father Arezki with him - no it's not father Ali who married her...

SP2 (to speaker 1): that's four...

SP1= five!

SP2= Don't include Ali's house (in the survey), so that's five

SP1= so *Hajj Tahar married five* (women)

SP2= yes

SP1= isn't that so?

- (16) xəmsidd juɣ ʒəddi lħadʒ tʃafiarʃ //
 xəmsa i=dd juɣ ʒəddi
 five REL.REAL=PROX take\PFV:SBJ.3SG.M grandfather\ANN.SG.M
 lħağ tʃafiar //
 pilgrim tʃafiar //
 "My grandfather Hajj Tahar married FIVE (women)."
 (KAB_AM_CONV_01_SP1_133)

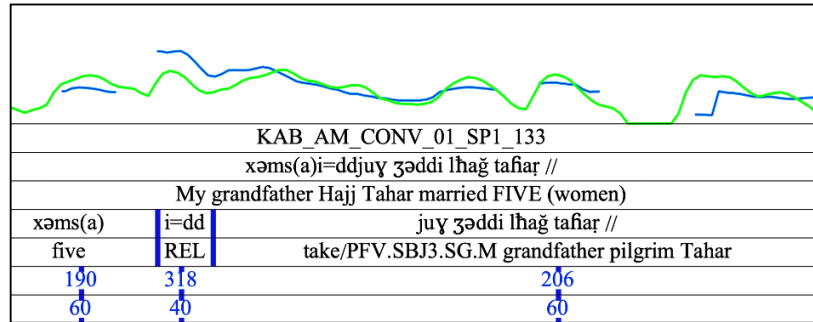


Figure 9: Example (16)

Example (17) is taken from a personal recount developing comparisons between the condition of women in the old days, and their present situation. The structure under study comes after a stretch of the discourse where the speaker says that nowadays, Kabyle women can and do work outside, with husbands sometimes staying at home and minding the children. Those women are the heads of their own houses, unlike in the old days, where the traditional situation would be that young women would come and live in the patriarch's house, and be under the strict rule of their mother-in-law, which is what example (17) states. There is therefore a contrast here between a possible previous assumption, based on the current situation in Kabylie, where brides are the heads of their domestic space, and the assertion, which is that actually, in the old days, it was the mother-in-law who held that role.

Context: The speaker talks about how hard life was for the Kabyle woman in the old days, and how power relationships were distributed.

Paragraph from which the example is taken: "The mother-in-law, she would control everything, the daughter-in-law couldn't express herself, the mother-in-law would decide how much flour you could have, she would fill you a minimal measure, only then could you go and prepare bread, only when she had prepared a little sauce could you cook, you could not go against her commands. Now... in the past we had attic shelves (where food was kept), and the mother-in-law, *she was the head of the household*, it was her who would gather provisions on the attic shelves. If she wanted to give you things, she would be the one to decide."

- (17) ðnətʃtsat iðrʃrʃaj bb°əxxam /
 d nətʃtat i d rʃaj n
 COP IDP.3SG.F REL.REAL COP commander\ABS.SG.M GEN
 wəxxam /
 house\ANN.SG.M /
 "It was she who was the commander of the house." (KAB_AM_NARR_03_0703)

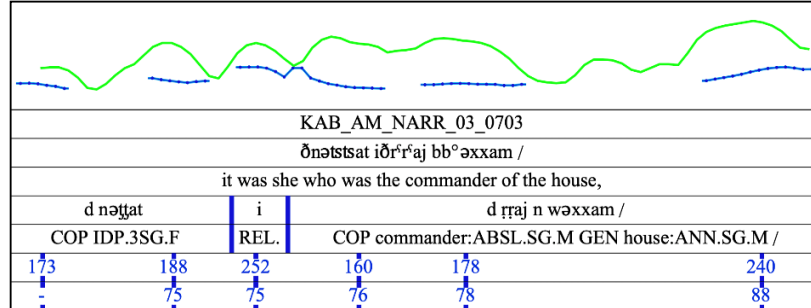


Figure 10: Example (17)

Despite the fact that the cleft clause is non-verbal, its prosodic profile is similar to those of verbal cleft clauses in examples (9) to (11). This shows how crucial prosody is as a component of the construction.

The common function between the structures in (15) to (17), and those under (9) to (12) suggests that they represent the same construction type, with presence of a copula at the onset of the clefted constituent being dependent on the nature (nominal vs non-nominal) of that constituent, and the nature of the clefted clause possibly non-verbal as in (17). A strong support for that characterization of clefts is the common prosodic pattern that those structures share:

- An F_0 peak on the relativizer
- A lowering of F_0 on the clefted constituent, proportional to the degree of prominence of the relativizer – that proportion expressing degree of contrastiveness
- A single intonational contour for both parts of the cleft.

4.3. A counterexample?

In the one-hour corpus, one example does not exactly fit the above characterization: it has the same morphosyntactic form, the same prosodic pattern with F_0 prominence on the relativizer, but it was segmented into two intonation units by one of two native speakers in the transcription process.

Context: The speaker fondly remembers her childhood, when she would help her grandmother, and take the sheep to pasture.

Paragraph from which the example is taken: She (the grandmother) would take the village herd to pasture. She would gather bundles of wood, she would carry one and I would carry one, although I was a little girl. We would arrive, we would light a fire, we would sit around, we would talk, she would tell us folktales, life in those days was sweet, *it was life in those days, which was the good life*. But now, it's different.

- (18) θamɸiɸnɸni nziç igəβninən // tsamɸiɸnɸni nziç / i ðəlɸali //
- | | | | |
|---------------------|---------------------|----------|------------|
| tamɸiɸt-nni | n | zik | i |
| life\ABS.SG.F-SHREF | GEN | long_ago | REL.REAL |
| d | tamɸiɸt-nni | n | zik / |
| COP | life\ABS.SG.F-SHREF | GEN | long_ago / |

ibninn // i d lʕali //
 be_tasty\PFV:RELSBJ // REL.REAL COP good //
 “Life of those days, it was sweet.... It was life in those days, which was the
 good life.” (KAB_AM_NARR_03_0516-0518)

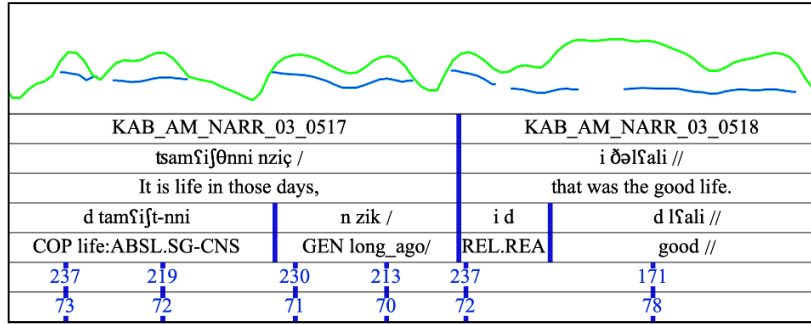


Figure 11: Example (18)

As far as function is concerned, the above example is not contrastive: the several preceding intonation units are all about the good old times, and even if the whole recording is built on an opposition between past and present, the passage under study does not insist on the comparison, and is more of a development on the happy memories the speaker has of her childhood. And example (18) is the conclusion of that series of happy memories going back to intonation unit # 0464, that is fifty-two intonation units (including pauses) before (a total duration of 38 seconds). In that context, interpretation of the structure under consideration (0517-18), especially as its prosodic contour parallels that of intonation unit 0516, is more in terms of a discourse antitopic, allowing the underlying semi-active idea to resurface, in a similar way as in example (7) the relative clause flagged a previous subtopic about the healthy food that people consumed in the old days.

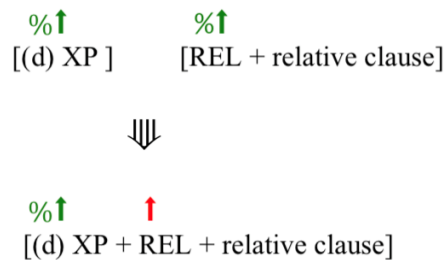
However, the prominence on the relativizer is clearly there (fig.11). On the other hand, the clefted constituent is not compressed, and starts at the same F₀ pitch as the relativizer (237 Hz).

Everything looks as if the speaker was hesitating between a simple characterization of life in the old days, triggered by the preceding context, and a contrast, triggered by the overall argumentative structure of the whole recording, which very clearly opposes life in the old days and life nowadays.¹¹ Besides, after a breath intake, she immediately resumes her comparison with *mi tura / mačči kif kif //*, ‘as for now, it’s different’ (KAB_AM_NARR03_0520-0521). This fluctuation might explain the inter-annotator disagreement on the presence of a boundary between intonation units 0517 and 0518.

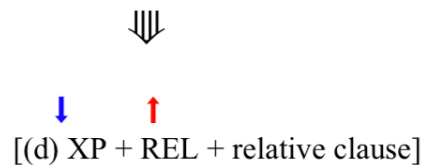
Indeed, the prominence on the relativizer could be interpreted in two ways: either as typical of a cleft structure (in that case the standard assumption would be that 0517 and 0518 form one unit), or as a pitch reset (0518, distinct from 0517), marking the left-boundary of a new intonation unit, with an interpretation of 0518 in terms of simple relativization, or afterthought.

¹¹ For a detailed summary of NARR03, see Mettouchi (2015).

In turn, the interpretive ambiguity of that structure points to the possible origin of the Kabyle cleft prosodic profile. Prominence on the relativizer is a rare typological feature, even among Berber languages (Louali and Mettouchi 2002), and one explanation for this pattern could be that the structure was originally biclausal, this being realized by a distribution over two distinct intonation units. The high pitch on the relativizer would have been originally due to its position at the left-edge of the second intonation unit, a typical position for pitch reset. This high boundary-tone %↑ would then have been reinterpreted in terms of prominence ↑, as an Fo peak within a cleft structure, enclosed in a single intonation unit.



And instead of lowering the pitch on the relativizer in order to place the focal accent on the clefted constituent, which is the most widespread pattern cross-linguistically, Kabyle kept it on the relativizer, and resorted to more or less pronounced compression (↓) of the clefted constituent in order to express the crucial component of focus, which is the differentiation between assertion and presupposition.



This choice of having the prominence on the relativizer, and the clefted constituent compressed (Kabyle), rather than the prominence on the clefted constituent, and the cleft clause compressed (standard cross-linguistic profile), is available because, as already stated in Mettouchi 2003a&b, the new information in the focus is not the constituent itself but the establishment of a difference between assertion and presupposition. And this can be formally marked in various ways, cross-linguistically.

5. Conclusion

The systematic investigation of a morphosyntactic structure, in a language-internal perspective where any functional difference is thoroughly investigated in order to find its formal correlate, makes it possible to precisely

define the cleft in Central-Western Kabyle: it is a construction whose function is to express narrow focus, and which is characterized by:

- the juxtaposition of:
 - a phrasal constituent (NP, ADV, QNT+N, PREP+N,...) preceded by a copula when its head is nominal,
 - and a clausal constituent introduced by the relativizer *i* (*realis/de re*) or *ara* (*irrealis/de dicto*),
- where the relativizer bears the main prosodic prominence of the structure, and the relationship between the two parts of the cleft is marked by:
 - an F₀ peak on the relativizer,
 - a lowering of F₀ on the clefted constituent, proportional to the degree of prominence of the relativizer, that proportion expressing degree of contrastiveness,
 - a single intonational contour for both parts of the cleft.

Prosody plays a central role in the definition, and is in no way a superimposed optional pattern, as is still postulated in various approaches (e.g. Drubig & Schaffar 2001). In the perspective adopted in this paper, prosody is a formal domain no less central and crucial to grammar as morphology, syntax, phonology.

Finally, as far as corpus design is concerned, the annotation of prosodic peaks appears necessary, and should be implemented for automatic queries to result in the retrieval of all and only the clefts in the corpus, and also, to allow the discovery of new constructions involving F₀ peaks, beside the marking of narrow focus. However, this annotation, which relies on relative rather than absolute F₀ values, is complex to implement, and thorough testing procedures involving the comparison of acoustic and perceptual decisions should be conducted. One way of annotating those peaks would be to use ANALOR (Avanzi et al. 2008), which has already been used on data from the CorpAfroAs corpus.

References

- Avanzi M., Lacheret A. & Victorri B., 2008, Analor, un outil d'aide pour la modélisation de l'interface prosodie-grammaire, *CERLICO*, Poitiers, p. 27-46.
- Basset A., 1952, La langue berbère, *Handbook of African Languages*, 1. London, Oxford University Press for International African Institute,
- Chaker S., 1983, *Un parler berbère d'Algérie (Kabylie): Syntaxe*, Aix-en-Provence, Publications de l'Université de Provence.
- Drubig H. B. & Schaffar W., 2001, Focus constructions, in Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language Typology and Language Universals*, Berlin and New York, Walter de Gruyter, p. 1079-1105.
- Ennaji M. & Saddiqi F., 1986, The Syntax of cleft sentences in Berber, *Studies in Language* vol.10:1., p. 53-77.
- Galand L., 1957, Un cas particulier de phrase non-verbale: 'l'anticipation renforcée' et l'interrogation en berbère, in *Mémorial André Basset*, Paris, Maisonneuve, p. 27-37.

- Galand L., 1984, Typologie des propositions relatives, la place du berbère, *Lalies* 6, Paris, Presses de la Sorbonne Nouvelle, p. 81-101.
- Galand L., 2002, Propositions relatives, thématization et rhématisation: l'exemple du berbère, in *Etudes de Linguistique Berbère*, Louvain & Paris, Peeters, p. 331-356.
- Izre'el Sh. & Mettouchi A., 2015, Representation of speech in CorpAfroAs: Transcriptional strategies and prosodic units, in Amina Mettouchi, Martine Vanhove & Dominique Caubet (eds.), *Corpus-based Studies of Lesser-described Languages: The CorpAfroAs corpus of spoken Afroasiatic languages*, Amsterdam and Philadelphia, John Benjamins, p. 13-41.
- Lambrecht K., 1994. Information structure and sentence form, *CSL* 71, Cambridge, Cambridge University Press,
- Leguil A., 1984, Une mutation syntaxique en français et en kabyle, *BSLP* 79/1., p. 323-332.
- Louali N. & Mettouchi A., 2002, Structures intonatives en berbère: l'énoncé prédicatif à particule *d*, in *Proceedings of SP2002 International Symposium on Speech Prosody, Aix-en-Provence, 11-13 April 2002*, p. 463-466.
- Mettouchi A., 2003a, Contrastive Focalization on clefts in Taqbaylit Berber, in Amina Mettouchi & Gaëlle Ferré (eds.), *Proceedings of IP2003 Interfaces Prosodiques, Nantes 27-29 mars 2003*, p.143-148. <http://llacan.vjf.cnrs.fr/pers/mettouchi/IP2003.html>.
- Mettouchi A., 2003b, Focalisation contrastive et structure de l'information en kabyle (berbère), in Jacques François & Anne Lacheret (eds.), *Mémoires de la Société de Linguistique de Paris: Fonctions et moyens d'expression de la focalisation*, Paris and Louvain, Peeters, p. 81-97.
- Mettouchi A., 2015, Aspect-Mood and discourse in Kabyle (Berber) spoken narratives, in Doris L. Payne & Shahar Shirtz (eds.), *Beyond Aspect: the expression of discourse functions in African languages*, Amsterdam and Philadelphia, John Benjamins, p. 119-143.
- Mettouchi A., 2017, Predication in Kabyle (Berber), in Amina Mettouchi, Zygmunt Frajzyngier & Christian Chanard (eds), *Corpus-based cross-linguistic studies on Predication (CorTypo)*, <http://cortypo.huma-num.fr/publications.html>.
- Mettouchi A., Smail N. & Louali N., 2004, Intonational structures in Berber: the non-verbal predicate *d*+XP in Tarifit and Taqbaylit, in Dymitr Ibrizimov & Rainer Vossen (eds), *Proceedings of the 2. Bayreuth-Frankfurter Internationales Kolloquium zur Berberologie 2002 (11-13 July)*, Köln, Rüdiger Köppe, p. 111-117.