Mettouchi, Amina, Graziano Savà & Mauro Tosco (2015). Cross-linguistic comparability in CorpAfroAs. In: Mettouchi, Amina, Martine Vanhove and Dominique Caubet (eds.), Corpus-based Studies of Lesser-described Languages: The CorpAfroAs corpus of spoken AfroAsiatic languages. Studies in Corpus Linguistics 68. John Benjamins: Amsterdam-Philadelphia. vi, 332 pp. + index (pp. 221-255).

Cross-linguistic comparability in CorpAfroAs

Amina Mettouchi, Graziano Savà and Mauro Tosco
LLCAN, Paris / LLCAN CNRS / University of Turin

One of the aims of CorpAfroAs is to allow queries within and across the language samples composing the corpus. Through the study of phenomena represented in several languages of the corpus (directional morphemes, case, and gender) we show that CorpAfroAs indeed allows the retrieval of a body of data amenable to cross-linguistic comparison, within the Afroasiatic phylum and beyond. However, given the annotation scheme of the corpus, the retrieval of relevant data has to rely on information given in the accompanying grammatical sketches.

Introduction

When the CorpAfroAs project was submitted in 2006, one of the aims underlying the creation of a corpus composed of several single-language corpora within AfroAsiatic, was to provide a basis for cross-linguistic comparison. In order to provide such comparable annotations, homogenization was necessary because descriptive traditions diverged a lot in their terminology and their perspective (see Barontini et al. this volume), not to mention the variation linked to the language in which the analysis was previously conducted by members of the project (in our case French, Italian, Spanish, English, and Hebrew).

The annotations chosen in CorpAfroAs are based on form, and they are language-internal in the sense that categories are defined within each language and are not comparative in essence (for the distinction between the two types of categories see Lazard (1975), Comrie (1979), Bybee (1985), Haspelmath (2010), among others). Only morphosyntactic information is provided in the first annotation line, \ge, while other types of information (semantic or morphological verb class, syncretism, etc.) are given on the second annotation line, *rx*. The basis of the morphosyntactic annotation is a form/function pairing, where a form coding a function, regardless of its many contextual interpretations, is always annotated in the same way. For instance, the *s-* derivation in Berber is consistently annotated as

Causative (CAUS) despite the fact that it often has a transitivizing function when applied to an intransitive verb, and is sometimes used to derive a verb of sound from onomatopoeia. The same is true for lexical items: the same verb, whatever its contextual interpretations, is annotated in the same way. For instance, in Kabyle, verb *xdəm* is always annotated as 'make', even if in some contexts it can be translated as 'work' (physical activity or employment). This allows the verification of hypotheses that may emerge in the study of corpora: for instance, is the interpretation of the lexical item as 'work' limited to intransitive uses of the verb *xdəm*? An automatic search involving the retrieval of the structures containing this verb shows that this is indeed the case.

One of the assumptions underlying the annotation process in CorpAfroAs was that there is some degree of resemblance between a language-internal category and a comparative one (cf. Haspelmath 2010 among others). Thus, Perfective in language A is basically comparable with Perfective in language B, regardless of the fact that Perfective in a language that only has a binary opposition with Imperfective does not have the same properties as Perfective in a ternary system also involving an aorist for instance. The effect of this assumption is that retrieval of bodies of data for the verification of hypotheses is conducted directly on the corpus, through a search interface on the website, that allows complex queries based on labels (available as a list of glosses and their abbreviations. See the paper by Chanard in the present volume). For instance, it is possible to retrieve all the negative clauses containing a Perfective, in all the languages of the corpus that have the category Perfective and Negation, by using the abbreviations NEG and PFV.

Indeed, homogenization was necessary, but not sufficient to conduct an informed cross-linguistic study. Relying only on labels may lead to ineffective searches in the corpus: for instance subject in Kabyle is a bound pronoun, whereas in Beja it is sometimes a noun, sometimes a nominal extension, sometimes a pronoun. Moreover, without indications about the criteria used for subjecthood assignment, it is difficult to consider *a priori* that we are dealing with the same category. Comparing Subject in the two languages cannot be done without the preliminary examination of the way this category has been used by the annotators of the various single-language corpora.

This is why we decided to provide an accompanying grammatical sketch for each language, in which the labels used by each linguist of the project are defined: in each sketch, a complete list of labels is provided, and information on the definition of most glosses[1] is given.

1. The complete list of all glosses used in the various languages composing the corpus is available on the project's website http://dx.doi.org/10.1075/scl.68.website.

The use of corpora for cross-linguistic comparison is thus mediated through a grammatical description (and possibly several, since the end-user can also use other sources before searching the corpus data).

This paper illustrates the potential for cross-linguistic comparison of the CorpAfroAs corpus, through examples of searches concerning three phenomena: directional morphemes, case and gender. Each study is based on automatic searches in the corpus, after prior analysis of information given in the corresponding grammatical sketch, and some grammars of the languages under consideration. Those searches can be replicated by accessing the online corpus at the following address: http://dx.doi.org/10.1075/scl.68.website.

1. Directional verbal extensions in Chadic, Berber and Cushitic

Some Afroasiatic languages have grammaticalized a system of bound morphemes that originally indicate directionality of the movement denoted by the verb. Often, those morphemes are used for all kinds of verbs, and their meaning is extended to such notions as benefit for the speaker, or resultativity (Mettouchi 1997 for Western Kabyle (Berber)) or to affected argument, non-controlling argument, or point of view of the predicate (Frajzyngier 2012b for Wandala (Chadic)). The following description aims to show how data from CorpAfroAs can be the basis for a cross-linguistic study of those directional elements.

1.1 Distribution

Six languages of the corpus have such directional morphemes: Hausa, Zaar, Tamasheq, Kabyle, Gawwada, and Ts'amakko.

The Hausa Ventive morpheme is glossed DIR (Directional) in the corpus, and corresponds to verb class 6 (glossed v6 in *ge*). This is the Grade 6 conjugation of Newman (2000). It "indicates action in the direction of, or for the benefit of the speaker" (Caron 2012).

(1) *an saːmoː tà neː*[2]
an *saːmoː* *tà* *nèː*
4.PFV.NFOC get.DIR 3SG.F COP1.NFOC[3]
PNG.TAM V6 PRO.OBJ PTCL.SYNT
"We got it" (HAU_BC_CONV_02_SP2_260)

Zaar has the suffix *-dî*, which attaches to pronouns or verb complexes, and is glossed as CTP (Centripetal) in *ge* and PTCL (particle) in *rx*.

(2) *wò sutəd̂i /*
wò *su* *=tə* *-d̂i* /
3SG.FUT return =3S.OBJ -CTP /
PNG.TAM V =PRO -PTCL /
"He will come back" (SAY_BC_CONV_01_SP2_171)

In Western Kabyle there are two clitics, Proximal *=dd* (glossed PROX in *ge* and PTCL in *rx*) and Distal *=n* (glossed DIST in *ge* and PTCL in *rx*), which attach to verbs of all kinds (not only motion verbs) and, like pronominal clitics, climb to Mood-Aspect-Negation particles, or relativizers.[4]

(3) *amidawiɣ θamaʃaḥuts͡ /*
ad *=am* *=dd* *awi* *-ɣ* *tamaʃaḥuʧ* /
POT =ABSV2SG.F =PROX bring\AOR -SBJ1SG tale\ABS.SG.F /
PTCL PRO PTCL V14 PRO N.OV /
"I will offer you a tale" (KAB_AM_NARR_01_0003)

(4) *jəddmətta͡səffaḥt /*
i- *ddəm* *=dd* *taʦəffaḥt* /
SBJ3SG.M- grasp\PFV =PROX apple\ABS.SG.F /

2. Examples have the following layout: the first line contains a phonetic transcription with prosodic words; the second line contains a morphophonological transcription involving grammatical words with morpheme breaks; the third line, named *ge*, is the morphosyntactic glossing tier; the fourth line, named *rx*, contains information about parts of speech, syntax, semantics, etc. The translation is followed by the identifier of the example within the corpus. This identifier always has the same syntax: ISO code of the language, initials of the author, genre (conversation or narration), number of the file, speaker (if more than one speaker is involved), number of the intonation unit in the file. Single or double slashes signal a prosodic boundary, non-terminal (/) or terminal (//). See the general introduction to the volume for more details.

3. The list of abbreviations is available at http://dx.doi.org/10.1075/scl.68.website. It is an expanded version of the Leipzig Glossing Rules, and its extension has been supervised by Bernard Comrie within the CorpAfroAs project (see the Introduction in this volume for more details).

4. Clitic climbing in Kabyle and Tamasheq is obligatory in front of Mood-Aspect-Negation particles, relativizers and some conjunctions.

PRO- V23 =PTCL N.OV /
"He took an apple" (KAB_AM_NARR_02_028)

(5) *antrˤuħ arʃʃixiw /*
ad *=n* *t-* *ɾuħ* *ɣr* *ʃʃix* *-iw* /
POT DIST SBJ3SG.F- go\AOR to teacher\ANN.SG.M -POSS1SG /
PTCL PTCL PRO V24 PREP N.COV PRO /
"She would go to my teacher" (KAB_AM_NARR_03_0478)

In Tamasheq (Berber) there are two clitics, Proximal *=du* (glossed PROX in *ge* and PTCL in *rx*) and Distal =in (glossed DIST in *ge* and PTCL in *rx*), which attach to verbs of all kinds (not only motion verbs), and like pronominal clitics, can climb to Mood-Aspect-Negation particles, or relativizers.

(6) *iḍgeẓtid ɐhɐḍ /*
i- *əḍgeẓ* *=tu* *=du* *ɐhɐḍ* /
3SG.M- squeeze\PFV =ACC.3SG.M PROX night\ANN.SG.M /
PNG V.IA1/TAM PRO PTCL N.OV /
"The night surprised him" (TAQ_CL_NARR_01_026)

(7) *uhun oṣadin /*
uhun oṣa *=in* /
then arrive\PFV[3SG.M] DIST /
CONJ V.IA10/TAM.PNG PTCL /
"Then he went there' (TAQ_CL_NARR_02_71)"

The situation in Ts'amakko and Gawwada is more complex due to the number of verbal extensions.

In Ts'amakko, *=na* is an assertive element marking the actual existence of an entity, or reality of a fact, which appears after nouns and verbs. After verbs, it is glossed ASS in *ge* and V.CL in *rx*; *=nu* is a Dative or Ablative after noun phrases, and a complementizer marking a conditional clause after verbs, where it is glossed DAT in *ge* and CONJ.V in *rx*.

In Gawwada, *-na* and *-nu* are decomposed into MOV (mover), i.e. the element to which *-a*, *-u* (and marginally *-i*) need to be affixed in order to act as adpositions (and different from their use with nouns) for *n-* and either CFG (Centrifugal) for *-a*, and CTP (Centripetal) for *-u*. They can both attach to nouns and verbs. With locative nouns, *-a* and *-u* attach directly to the noun stem with no intervening *=n* (cf. Tosco 2012a).

1.2 Distribution and functions

The distribution of those bound morphemes is variable across the corpus. First of all, Zaar and Hausa only have one extension, the Ventive. Among the languages that have at least two extensions, there is no necessary balance between the two in terms of frequency of use. Whereas in Tamasheq the proportion between distal and proximal is roughly 40% / 60%, in Western Kabyle it is 0.1% / 99.9%.[5] The difference within Berber is especially striking since the Proximal and Distal extensions are of the same diachronic origin (>*d; >*n) throughout the language family. Ts'amakko and Gawwada also share historically identical morphemes *-na* and *-nu*. In Ts'amakko the proportion between complementizer and assertive is roughly 3% / 97%, in Gawwada the proportion between centripetal and centrifugal is roughly 21% / 79%. Complementizer and Centripetal are of the same origin, as are Assertive and Centrifugal.

Chadic and Berber languages tend to use the Proximal more extensively than the Distal. The latter for instance has disappeared in Eastern Kabyle dialects. The table in Frajzyngier (1987) shows that for a sample of thirty Chadic languages, all of them have Centripetal extensions, but only fourteen also have Centrifugal extensions.

In Gawwada and Ts'amakko on the contrary, the Distal/Centrifugal is used more extensively than the Proximal/Centripetal.

In Western Kabyle, the Centrifugal extension is used in a limited number of contexts:

(8) *innajas llitsin ðinəβgirˁəppwi /*
i- nna =jas lli =ṭ =in
SBJ3SG.M- say\PFV =DAT3SG open\AOR(IMP2SG) =ABSV3SG.F =DIST
d inəbgi n ṛbbi /
COP guest\ABS.SG.M GEN god /
"He said open it (the door), I'm (lit. it is) a beggar'. (KAB_AM_NARR_01_0677)"

(9) *a:::ʂmar sərsijin /*
a aʂmar sərs =iji =n /
VOC aʂmar be_placed\CAUS.AOR.IMP2SG =ABSV1SG =DIST /
"Amar please put me down!" (The ogress was put on a donkey by Amar) (KAB_AM_NARR_02_760)

5. Those counts are indicative, since they are based on different amounts of data, but they correspond to the overall distribution of the two extensions in the languages under consideration.

Mettouchi (2011) proposes that the function of the Distal clitic is to indicate that the process is construed relative to the deictic center of the addressee. Distance is not at play, since in (9), Aʂmar is holding the donkey, and in (8) the door is in front of the speaker. Viewpoint is more important: the speaker could have used a proximal clitic in examples (8) and (9), thus making the command more peremptory. In both examples, the use of the Distal clitic subordinates the speaker's viewpoint to the addressee's, with politeness side-effects. This shows that the distinction here is not motivated by direction of a movement, but by modal viewpoint/stance. The same holds for (5), where the verb could have been used without a directional clitic. Movement towards the addressee is a possible interpretation, but politeness is also at stake in (5). Spatial directionality cannot therefore be considered as a core function since most examples involve no movement, and no spatial distance from the addressee.

In Tamasheq, the distal extension is used mostly with motion verbs ('come', 'arrive', 'go', 'be on the point of arriving') as well as verbs of saying.

(10) *ikkain hartin oṣa /*
i- əkka =in har=tu =in oṣa /
3SG.M- go\PFV =DIST until=ACC.3SG.M =DIST arrive\PFV[3SG.M] /
PNG- V.IA9/TAM =PTCL CONJ=PRO =PTCL V.IA10/TAM.PNG /
"He went in this direction (to see it)" (TAQ_CL_NARR_01_088)

(11) *ənnɐɣasin ɐɾɐt /*
ənna -ɐ =as =in ɐɾɐt /
say\PFV -1SG =DAT.3SG =DIST thing\ABS.SG.M /
V.IA9/TAM -PNG =PRO =PTCL N.COV /
"I said something to him" (TAQ_CL_NARR_05_21)

The proximal extension is also used with motion verbs and verbs of saying, as well as other types of verbs.

(12) *ʒɐˈrɐkkɐnted'du muˈdɐrɐn /*
ʒɐrɐkkɐt -ɐn =tɐt =du muɖɐr -ɐn ʒɐrɐkkɐt /
dig_up\PFV -3PL.M =ACC.3SG.F =PROX animal\ANN. -PL.M dig_up\PFV /
V.XA2/TAM -PNG =PRO =PTCL N.OV -PNG V.XA2/TAM /
"Wild animals had dug her up" (TAQ_CL_NARR_01_095)

(13) *iɖgɐẓtid ɐhɐḍ /*
i- əɖgɐẓ =tu =du ɐhɐḍ /
3SG.M- squeeze\PFV =ACC.3SG.M =PROX night\ANN.SG.M /
PNG- V.IA1/TAM =PRO =PTCL N.OV /
"The night surprised him" (TAQ_CL_NARR_01_026)

Unlike Western Kabyle, in Tamasheq the use of directionals with motion verbs is widespread, as well as the interpretation in terms of location of a situation close to the speaker or far from him or her. The proximal and distal meanings are still central, even though the general function of each marker is larger as is shown by their use with verbs of saying, where they involve stance, and with other types of verbs, where we find some of the dimensions noticed in Kabyle: completion, present relevance.

Languages that use extensions very frequently, such as Western Kabyle, are likely to use them with a large variety of verbs, not only motion verbs. Indeed, the distribution of verbs in the Western Kabyle corpus of CorpAfroAs is consistent with findings in Mettouchi (1997), where beside motion verbs, the proximal clitic was also encountered with change of state verbs, and with verbs of saying, handling ('take', 'hold', etc.), finding, among others. Almost any verb is possible, since the proximal clitic has lost its original directional value, and more generally organizes the utterance around the deictic center of the (direct or reported) speaker or protagonist (Mettouchi 2011), with modal or aspectual dimensions as well as purely spatial ones.

In example 14, we can see the use of Proximal in two contexts. One is a verb of handling with motion ('take away') where the Proximal clitic is motivated by the focus on completion of the action, underlined by the conjunction *alamma* 'until': it is only when the bread is taken off the shelf that the father will know that his youngest daughter is old enough to feed herself if her stepmother neglects her.

The other context is negative and involves a verb that is not usually associated with a Proximal clitic. The motivation for the use of the Proximal clitic here is modal: the utterance is organized around the speaker's viewpoint and underlines stance: it is a categorical statement, almost an oath. This is reinforced by the use of the Negative Perfective with future time reference, usually in contexts of solemn oaths.

(14) *urdəzwiʤəɣ / alamma θəkksədd / fatˁima θuħrˁifθ / aɣrum əgðəkkwan //*
ur =dd zwiǧ -ɣ / alamma t- kks
NEG =PROX marry\NEGPFV -SBJ1SG / until SBJ3SG.F- take_away\PFV
PTCL =PTCL V23 -PRO / CONJ PRO- V23
=dd / Fatima tuḥrift / aɣrum g udəkkan //
=PROX / Fatima clever / bread:ABS LOC shelf:ANN //
=PTCL / NP ADJ / N.OV PREP N.OV //
"I won't marry until clever Fatima manages to take the bread from the shelf" (KAB_AM_NARR_01_0142)

The Hausa corpus also shows that motion verbs are not the only ones to be associated with Ventive extensions: beside 'return', 'leave', 'enter' or 'go', we find 'carry', 'get', 'do', 'sell', 'take', 'catch' (verbs of handling). In Zaar, the Centripetal extension

is associated with motion verbs ('return', 'go', 'arrive', 'leave', 'enter', 'thrust', 'pass by'), as well as verbs of handling ('take', 'hold', 'bring', 'weave', 'tie', 'rub', 'dig', 'gather', 'fetch'). It is remarkable that the same semantic subsets are associated with proximal/ventive/centripetal extensions in the three languages (Hausa, Zaar, and Kabyle).

(15) *mə́ ngâː ʒàdî //*
mə́ ngaː ʒà -dî //
1PL.AOR fetch water -CTP //
PNG.TAM V N -PTCL //
"We fetch water" (SAY_BC_CONV_02_SP1_007)

Gawwada and Ts'amakko use the Centripetal affix with various types of verbs, not necessarily motion verbs. As for the numerous Centrifugal affixes, they are mostly used with verbs of saying and telling: the proportion of affixation to verbs of saying compared to the two next most frequent verbs ('go' in Gawwada and Tsamakko, 'be there' in Gawwada, 'run' in Ts'amakko) is 8 to 1 in Ts'amakko, and 3 to 1 in Gawwada.[6] Other verbs used with the Centrifugal are 'return', 'arrive', 'run', 'jump' in Gawwada, 'arrive', 'tend cattle', 'eat' and 'come' in Ts'amakko.

In Western Kabyle, contrary to Hausa where the Ventive remains attached to the verb, Proximal particles are subject to clitic climbing with Mood-Aspect-Negation preverbal particles, or relativizers, and must attach to those preverbal morphemes (this is also the case for Absolutive or Dative pronouns) (see ex.16). The list of hits for a search involving the Proximal or Distal clitics cannot directly provide a list of associated verbs. Partial searches are necessary to recover all examples, after which the visualization of those examples makes it possible to retrieve the contextual elements at play in the interpretation of meaning: types of verbs, but also types of pronouns, presence of modal markers, types of aspect-mood used, etc. The precise study of those contexts highlights the frequent use of this Proximal clitic with Dative pronouns (19% of clauses (85 out of 451) containing a Proximal particle also contain a Dative pronoun, the proportion of clauses with Dative pronouns in the whole corpus being 13%).

(16) *aɣiddəħku səʦʦi θimuʃuħa //*
ad =aɣ =dd t-ħku səʧi timuʃuħa
POT =DAT1PL =PROX SBJ3SG.F-tell\AOR grandmother\SG tale\ABS.PL.F
PTCL =PRO =PTCL PRO-V13% N.KIN N.OV
"My grandma would tell us folktales" (KAB_AM_NARR_03_0245)

6. Counts are based on 100 verbs for each language.

This finding is consistent with the tendency of Zaar to have the centripetal extension attached to Benefactive markers (7% of clauses containing a centripetal extension also contain a Benefactive marker, the proportion of clauses with Benefactive markers in the whole corpus being 1%).

(17) *àː láːrmí mə́ndî ʧíɣə́j /*
àː láːrmí mə́ndî ʧíkə́j
àː láːr =mí mə́n -dî ʧík -íː
3SG.PFV bring =1PL.OBJ BEN -CTP thus -RES
PNG.TAM V =PRO PTCL -PTCL ADV -ASP
"as he brought [him] to us like this" (SAY_BC_CONV_02_SP2_044)

Finally, in Zaar, we notice the regular association of Resultative (glossed RES in *ge*) and the Centripetal extension:[7] 14% of clauses containing a centripetal extension also contain a resultative marker, the proportion of clauses with Resultative markers in the whole corpus being 10%. This may suggest that, as in Western Kabyle, movement towards the deictic center of the speaker can be associated with Completed or Perfect aspects, or the attainment of a goal.

(18) *ŋgwôyŋ tùlíːdî /*
ngôkn tùlíːdî
ngôkn tul -íː -dî
he_goat arrive -RES -CTP
N V -ASP -PTCL.EXT
"He-goat arrived" (SAY_BC_NARR_03_SP1_653)

This is interesting, since Western Kabyle, which did not grammaticalize the function 'resultative', regularly uses the Proximal particle to convey this meaning (Mettouchi 1997), as shown in example (14). On the other hand, Tamasheq, which has a Resultative aspect (glossed RES in *ge*), does not show any correlation between that aspect and the Proximal or Distal clitics.

Those qualitative findings serve as a basis for a larger cross-linguistic comparison of directional morphemes, should other Berber, Chadic and Cushitic languages be added to CorpAfroAs. They help formulate heuristic hypotheses on centripetal/proximal extensions in Chadic and Berber: once the markers start to be used outside a strictly spatial domain, it seems that the notion of direction towards a deictic center is extended to impact on the situation (with resultative meaning), or on the participants (with beneficial/detrimental meaning). It can even, as in Western Kabyle, take on modal values, such as viewpoint (especially with verbs

7. Ader Hausa, not represented in CorpAfroAs, has a combination of centripetal and resultative in the form of Grade 4 suffix *-ikkee* (Caron 1989:147).

of saying, or in *irrealis* or negative contexts). On the other hand, the findings concerning the Cushitic languages Gawwada and Ts'amakko show that extension of grammaticalization can also concern the Centrifugal extension. The very strong co-occurrence pattern with verbs of saying indicates that what is probably at stake, apart from direction of motion, or localization, is that the function of the particle is modal. And indeed, the centrifugal is glossed Assertive in Ts'amakko.

2. Case in AfroAsiatic

The second study is about cross-linguistic comparison of case in some languages of the CorpAfroAs corpus. It presents a typology of case values and a discussion on marking of syntactic roles in general.

The languages taken into consideration are Kabyle and Tamasheq for Berber, Hausa for Chadic, Hebrew and Moroccan Arabic for Semitic, Wolaytta for Omotic and Afar, Gawwada and Ts'amakko for Cushitic.

2.1 Defining case in CorpAfroAs

According to a common definition:

> Case is a system of marking dependent nouns for the type of relationship they bear to their heads. Traditionally the term refers to inflectional marking, and, typically, case marks the relationship of a noun to a verb at the clause level or of a noun to a preposition, postposition or another noun at the phrase level (Blake 2001:1).

Case as defined above is one of the possible coding means of syntactic roles. However, any marking of syntactic property of nouns and pronouns is often defined as case. While case is sometimes reduced to syntactic role marking, some theories expand its functional properties and use it to indicate more abstract semantic roles (Fillmore 1968). This is because often, but not always, case and syntactic-role correspond to some semantic characteristics. For example, a Nominative noun encoding the Subject in a sentence often acts as the agent.

Case is a form associated to a syntactic-marking function and typically a case system is ordered in case declensions with suffixes as case markers. Latin, Greek and Turkish are languages with such a system. However, other approaches to case allow case markers to be marked by clitics to the nouns or the phrase and pre-/post-positions. This is because sometimes pre-/post-positions, nouns and phrase clitics and inflectional case markers are connected on a grammaticalization line and in fact may express the same function. The degree of boundedness of the case marker can also go in the other direction, so that case is marked by word

suppletion and the whole form changes according to the expressed case. This is typical of case-determined pronominal paradigms.

Typical labels such as Nominative, Accusative and Dative are used in CorpAfroAs to indicate case. Syntactic role marking labels are Subject, Direct Object, Indirect Object etc. and semantic roles are agent, patient, recipient etc. The use of these labels in the corpus reflects the kind of analysis applied and has loose correlation with segmental properties. The preference goes to general syntactic role marking if the element is not considered 'case-like'. Unbound elements such as pre-/post-positions tend to receive lexical glosses such as 'to', 'on', 'with'. However, they can be interpreted as grammatical role markers and glossed accordingly: one preposition in Kabyle is glossed DAT because the function of this element is considered similar to the one typically coded by Dative case. Semantic roles can be coded by any bound or unbound form. It should be added that syntactic roles can also be inferred, among other coding means, from agreement and the position of the word in the clause.

In the languages of the CorpAfroAs corpus analyzed in this paper, case systems are rather poor. One language has a suffixal case system. In others case is coded by different forms, which are integrated in one system. Other languages have cases only in pronouns.

2.2 A description of case marking in AfroAsiatic

2.2.1 *Case suffixes and apophony*

The only language of the CorpAfroAs corpus with an exclusive series of case affixes creating a declension is Wolaytta (Omotic). The declension applies to both nouns and pronouns. Eight nominal case suffixes operate in this language: Nominative (NOM), Accusative (ACC), Genitive (GEN), Dative (DAT), Locative (LOC), Directive (DIR), Instrumental (INS), Comitative (COM).

The Nominative in Wolaytta, and in several Ethiopian languages, is typologically interesting because it is not part of a system that can be defined as Nominative-Accusative or Ergative. It marks the Subject in an intransitive clause and the agent in a transitive clause and indicates the Subject of a copula clause. In the last case the predicative element, i.e. noun, pronoun or adjective, is marked by the Accusative case.

Nominative and Accusative case affixes are gender-sensitive. Therefore, M or F precede, separated by a dot, the case glosses (see 3.6. below for more details).

In example (19) the masculine noun *gaammo* 'lion', is marked by the Nominative case *-i*:

(19) gaammóy ʔissí ʔíndé míizza laaggíis //
gaammóy ʔissí ʔíndé míizza
gaammó -í ʔissó -í ʔindé miízza
lion -M.NOM one -LINK female.old cow
N -CASE NUM -CONNECT ADJ N
laaggíis //
laagg -iis //
drive -3MSG.PAST.AFF.DECL
v1 -TAM
"the lion drove one old cow" (WAL_AA_NARR_05_lion_15)

Gawwada and Ts'amakko, both Cushitic languages of the Dullay cluster, have only one suffixal case: Associative (ASSOC).

The case is actually expressed by three case-sensitive forms according to the three-gender distinction of these languages. As for the meaning, the three suffixes indicate both a location in a sentence and a possessor in a noun phrase.

See example (20) from Gawwada, where the noun *kolle* 'river' is marked as locative by the Associative feminine case *-atte* after deletion of the final Feminine gender marker *-e*.

(20) *ʃeːtte ʂagaba gollaj / gollatte /*
ʃeːtte ʂakapa kollaj#
ʃeːt -t -e ʂak -a =pa kollaj# /
girl -SING -F be_there -IPFV.3SG.M[8] =LINK kollaj# /
N -PNG -PNG V -TAM.PNG =CONJ FS /
kollatte
koll -atte /
river -ASSOC.F /
N -PNG /
"There was a girl at the…at the river" (GWD_MT_NARR_07_012–013)

In Afar, Nominative (NOM), which has similar characteristics as the one described above for Wolaytta, and Genitive (GEN) indicate case marking by apophony and movement of the accent to the word-final syllable. In fact, only masculine nouns in the unmarked Absolutive (ABS) case that end in a vowel are marked for NOM and GEN. For both cases the apophony is *a* > *i* and the accent moves to the last syllable of the word. If the word-final syllable is underlyingly accented, the case is marked by apophony only.

8. The M agreement of an F noun is caused by the loss of agreement between the subject and the verb. This is due to Gawwada's Subject focusing strategy in the formation of thetic sentences (Tosco 2010: 325).

2.2.2 *Case clitics*

Other syntactic roles of nouns and pronouns in Gawwada, Ts'amakko and Afar are indicated by a series of clitics. In Gawwada and Ts'amakko the domain of case marking by clitics is not the noun but the noun phrase. They attach after the last element of a noun phrase and do not replace the last vowel of a modified noun, in contrast to what happens with the Associative case. The clitics in Ts'amakko are Dative (DAT), Diffusive (DIFF), Comitative-Instrumental (COM) and Locative (LOC). It is to be noted that DAT in Ts'amakko marks both a recipient-receiver and a source-provenance. Gawwada glosses differ in that there is no LOC clitic and the Ts'amakko Dative *=nu* corresponds to a combination of the Mover (MOV) morpheme *=n* followed by the Centripetal (MOV-IN) affix *-u*. The Gawwada *=n-u* is opposed to *=n-a*, Mover-Centrifugal (MOV-OUT), and *=n-í*, Mover-Specific (MOV-SPEC). This means that Gawwada has two additional case clitics *=n-a* (MOV-OUT) and *=n-í* (MOV-SPEC). The description is summarized in the following table:

	IN *-u*	OUT *-a*	SPEC *-í*
MOV *=n*	*=n-u*	*=n-a*	*=n-í*

In the following example from the Ts'amakko corpus, the Diffusive *=ma* follows the modifier *linq'e* 'clean' since it marks the whole Noun Phrase rather than the Head Noun *do:llo* 'skin mat'.

(21) *bagannaŋki qawko do:llo / li:::nq'e agí:ppi / garmitto //*
bagaɗnanki q'awko do:llo
bagaɗ -n -anki q'awk -o do:ll -o /
run.P -FUT -IPFV.1PL man -M skin_mat -M /
V -TAM -TAM.PNG N -PNG N -PNG /
linq'ema gí:ppi / garmitto //
linq'e=ma gí:f ~p -i / garm -itt -o //
clean=DIFF go_to_sleep ~SEMELF -PFV.3SG.M / lion -SING -M //
ADJ=CASE.CL V ~V.DER -TAM.PNG N ~N.DER -PNG
"We'll run. The one who sleeps on the clean mat is a lion".
(TSB_GS_NARR_001_SP1_248–250)

Gawwada and Ts'amakko case clitics also attach to pronouns. They attach to the Object pronouns, labeled OBJ in CorpAfroAs, following directly the pronominal morpheme. The other main pronominal paradigm is Subject (SBJ). Gawwada glosses differ here in preferring Oblique (labelled OBL) for the Ts'amakko Object and in using the Subject paradigm only for the participants, while non-participants use the aforementioned Specific (SPEC) *-í* or a Generic (GEN) *-a*. Therefore, only non-core case is marked by case clitics on pronouns.

In Ts'amakko, when a pronoun is marked for locative, the case clitic *=ta* rather than the Locative case is used. This is shown in the following example, where the 1SG.OBJ pronoun *ʔe:ta* is followed by the Locative *=ta:*

(22) *eta sabbete ita magĝi //*
ʔe:ta sabbete
ʔe: =ta sabb -ete ʔita magĝi //
1SG.OBJ =LOC top -LOC.P away go_away.IMP.SG //
PRO.IDP CASE.CL N.LOC CASE ADV.LOC V //
"Get away from me" (TSB_GS_NARR_006_SP1_35)

Afar also makes use of case clitics for nouns and pronouns. These are *=h* Centripetal (CPT), *=k* Centrifugal (CFG), *=l* Instrumental (INS) and *=t* Locative (LOC).

2.3 Syntactic roles marking in pronouns

In the rest of the languages under analysis, i.e., Hausa and Zaar (both Chadic), Kabyle and Tamasheq (both Berber), Hebrew and Moroccan Arabic (both Semitic), case and syntactic roles are only indicated in pronouns. Case marking in the Berber languages is Accusative (ACC) and Dative (DAT) in Tamasheq and Absolutive (ABS) and Accusative (ACC) in Kabyle. The following glosses are also used in Berber: SBJ for pronominal Subject and ABSL (Absolute) and ANN (Annexed). The latter two do not indicate case but state of the nouns in the context of the clause and the phrase. How the two states are selected according to the syntactic context in which they appear is one of the big questions of Berber linguistics (see Mettouchi and Frajzyngier (2013), for the most recent hypothesis that has an impact on general typology).

Other pronominal series that indicate syntactic roles are the Object (OBJ) pronominal clitics and Subject (SBJ) independent pronouns in Hebrew and the Possessive (POSS) and Object (OBJ) pronominal clitics of Moroccan Arabic. Case syncretism between POSS and OBJ in Moroccan Arabic is analyzed as Oblique case and labeled OBL. Finally, Hausa has Object (OBJ), Benefactive (BEN) and Possessive (POSS) pronominal paradigms, while what in the other languages is presented as a subject pronominal paradigm here is labeled IDP, i.e. "Independent".

2.5 Cross-linguistic queries on case in CorpAfroAs

The description presented above shows that in the CorpAfroAs corpus case is poorly expressed and case systems largely integrate morphological marking of syntactic role. The only exception is Wolaytta with its full-fledged case declension. When conducting queries in the CorpAfroAs corpus, therefore, one should be

aware of the fact that syntactic roles may or may not be indicated by case glosses. For example it is noteworthy that the core case glosses NOM and ACC are used for case suffixes and less so in pronominal paradigms. The syntactic role labels SBJ and OBJ are preferred for pronominals.

The corpus also shows that case marking is not necessarily a modification of a word. Ts'amakko and Gawwada show a case concord system where the domain of case marking is the noun for case suffixes, but the noun phrase for case clitics. The noun is marked by the clitic if it is the only element of a noun phrase. The structure of those languages being Head-Modifier, if any modifier, including a relative clause, follows the Head Noun, the clitic attaches to the modifier. If there is more than one modifier, the case marker will still follow that rightmost modifier. This is not valid in the case of pronouns, which are directly followed by the case-clitics.

According to one of the principles of the CorpAfroAs methodology, a single gloss is associated to each grammatical form and each gloss reflects the meaning and the function of the form. The choice of the gloss is therefore an outcome of the language-internal analysis suggested by the grammatical system of each language. This is visible also in the glossing of case.

3. Gender in AfroAsiatic

3.1 Overview

Both gender and number are robust categories in AfroAsiatic languages in general, and in this respect the languages of our corpus are good representatives of the language family as a whole: gender is marked in all of them with the exception of Zaar, and Juba Arabic (a creole/pidgin). Number (which will be tackled here only insofar as it interacts with gender) seems to be marked in all languages of the project. Moreover, gender and number interact in many interesting and different ways, as will be shown below.

The robustness of gender in AfroAsiatic is shown in agreement with a gendered nominal head on modifiers, as well as on the verb, where the gender of the subject (be it overtly expressed as a noun, pronominalized, or contextually given) governs agreement on the form of the verb.

The correlation of grammatical gender with sex in animates may be weak, and sometimes it is non-existent.

The Afroasiatic gender system is based upon a binary Masculine (M) vs. Feminine (F) distinction, with the latter generally being the marked member of the opposition.

Number is based minimally upon a Singular (SG) vs. Plural (PL) opposition, with the latter being again the marked member. Against these family-wide generalities, a number of deviations are observed. Within the gender system, F is, occasionally, the unmarked member: such a situation has been described for Zayse and Zargulla (Omotic; Hayward 1989) but is not represented in the corpus. Variation within the number system is more widespread and diversified and involves both the number of elements in opposition and their markedness value. A common departure from the basic SG vs. PL opposition involves a Collective from which a *nomen unitatis*, or Singulative (SING) is derived: in this case, the markedness values are reversed, with SING often being marked. Other variation may involve the presence of a separate Dual (not represented in the corpus). More restricted variations may yield a Plurative alongside a Singulative, and the reanalysis of Plural as a third gender (in the sense of a partially lexically-specified classification of nouns; see below 3.6.).

Gender and number may interact in agreement as well as in the actual shape of the exponents.

3.2 Categories affected by gender

Among the languages in our corpus, nouns, personal pronouns and verbs favor the expression of gender. Adjectives too are often, but to a lesser degree, gender-marked. Moreover, a few languages (represented in our corpus by Afar) may lack the category of adjectives altogether. Other categories mark gender in at least a subset of their members. The conditions affecting the marking of gender may be lexical or morphosyntactic; e.g., demonstratives in Kabyle do not show gender-variation when they occur as affixed nominal modifiers, but they do as pronouns. Cf (23) vs. the pronominal use in (24).

(23) *a-rgaz-agi* "this man" — *t-a-qʃiʃ-t-agi* "this girl"
ABSL.SG-man-PROX — F-ABSL.SG-child-SG.F-PROX

(24) *wagi* "this one (M)" — *tagi* "this one (F)"
PROX.SG.M — PROX.SG.F
wigi "these ones (M.PL)" — *tigi* "these ones (F.PL)"
PROX.PL.M — PROX.PL.F

Berber languages have gendered numerals; when the native numerals have been superseded by (Arabic) loans, as in Kabyle, gender is marked on the inherited numbers 'one' and 'two,' and absent in the Arabic-derived numerals from 'three' onwards:

(25) *jiwən* "one (M)" *jiwət* "one (F)"
sin "two (M)" *snat* "two (F)"

Where original numerals have been retained (as in some other Berber varieties) gender-agreement applies to the whole category of numerals.

Similar restrictions operate in other languages of the AfroAsiatic phylum and in the corpus. In Table 1, a language will be considered as marking gender on the relevant category if it marks it in a subset (minimally, one element) of the members of that category:

Table 1. Gendered categories in the CorpAfroAs languages

language	family	Noun	Pers. Pro.	Adj.	Dem.	Num.	Poss.	Def.	Verb
Afar	Cushitic	+	+	missing[9]	–	–	+	missing	+
Arabic: Moroccan	Semitic	+	+	+	+	–	+	–	+
Arabic: Tripoli	Semitic	+	+	+	+	–	+	–	+
Arabic: Juba	Semitic	–	–	–	–	–	–	–	–
Beja	Cushitic	+	+	+	+	+	+	+	+
Gawwada	Cushitic	+	+	+	–	+	+	missing	+
Hausa	Chadic	+	+	+	+	–	+	+	+
Hebrew	Semitic	+	+	+	+	+	+	–	+
Kabyle	Berber	+	+	+	+	+	+	missing	+
Tamasheq	Berber	+	+	+	+	+	+	missing	+
Ts'amakko	Cushitic	+	+	+	–	+	+	missing	+
Wolaytta	Omotic	+	+	–	+	–	+	+	+
Zaar	Chadic	–	–	–	–	–	–	–	–

The defining characteristic of gender is agreement, and evidence for gender must be found outside nouns: a language may be said to have a gender system only if different agreement patterns are found on various target categories, and these ultimately depend on controllers (typically, nouns) of different types (cf. Corbett 1991, 2006).

The following sections will provide evidence of the morphological expression of gender on nouns (3.3.) and pronouns (3.4.) before discussing gender agreement (3.5.) and the interaction of gender with number (3.6.).

9. "Missing" implies that the corresponding word-class does not exist in the language in question. In the case of Afar (and other East Cushitic languages not represented in CorpAfroAs), the semantic class of "adjectives" is represented by different categories of verbs.

3.3 Gender in nouns

As anticipated in 3.1. and as is common in gender systems, little if any relationship is found between grammatical gender and natural sex. The following are two Cushitic examples among many. As will be further expounded in 3.6 below, Gawwada and Ts'amakko often overtly mark number — in (26), the Singulative — before gender on nouns:

(26) *hisk-att-o / hesk-att-o* "woman" (Gawwada/Ts'amakko)
woman-SING-M

(27) *loʔ-o* "cow" (Gawwada/Ts'amakko)
cow-M

Conflict between morphological (gender-assigned) and semantic (sex-determined) agreement are not uncommon; e.g., Gawwada *hisk-att-o* 'woman,' morphologically M, governs agreement with the verb in the 3F form when subject, although, e.g., morphological agreement is always followed by an agreeing possessive or adjective, which occur in the M.

Languages without gender marker, such as Juba Arabic, may express the sex of animate entities lexically, for example with the word *mára* 'woman;' e.g.. *ásed* 'lion,' *ásed ábu mára* 'lioness' (where *ábu*, literally 'father,' is used, as in Arabic, as a relative marker).

Languages where one gender only is marked on the head are very common; in such a case, the unmarked member is the M, with F being marked by a suffix, a prefix, or both. In Moroccan Arabic (Semitic), only F is in general overtly marked. The marker is suffixal:

(28) *əl=ħbəq* "the basil" (ARY_AB_narr_01_004)
ART=basil[-M]
DET=N.M

(29) *əl=qbiːl-a* "the tribe" (ARY_AB_narr_01_020)
ART=tribe-F
DET=N-PN

(28) further shows that whenever a category (in this case, and most typically in the domain of gender, M) is not formally marked in the language, it is not *per se* retrievable from the glosses (in (28) M is added in brackets for comparative purposes).

Often, both genders are overtly marked, for example, in languages of the Cushitic group. In Gawwada, affixal *-o* and *-e* mark, respectively, the M and F gender (as well, for *-e*, the PL, as detailed in 3.6. below):

(30) *paʃ-o* "field" (GWD_MT_NARR_011_019)
field-M
N-PNG

(31) *pij-e* "land" (GWD_MT_NARR_011_017)
land-F
N-PNG

Also in Wolaytta, M and F nouns have different endings, generally followed by gender-sensitive determiners and case markers:

(32) *gaammó-a* "the lion" (WAL_AA_NARR_05_lion_05)
lion-DEF.M.ACC
N-PNG-CASE

(33) *ʔindé-ó* "the old one" (WAL_AA_NARR_05_lion_26)
female_old-F.ACC
ADJ-PGN

Covert (zero) gender marking is by no means rare. E.g., Moroccan Arabic *daːr* 'house' is unmarked as F; the agreeing adjective that follows is duly marked as F by *-a*:

(34) *f=əl=daːr waːḥəd-a*
in=DEF=house a_single-F
PREP=DET=N.F ADJ-PNG
"In one house" (ARY_AV_NARR_02_398)

Or, in the following example, by the verbal form, which is again marked as F:

(35) *əl=daːr ʕaːdi t-ṭiːḥ ʕla=na*
DEF=house FUT 3F-fall\IPFV along=OBL.1PL
DET=N.F PTCL PNG-V PREP=PRO.PNG
"The house will fall on us" (ARY_AV_NARR_02_044)

In Beja (Cushitic), gender is recovered *inter alia* from gender-sensitive (in)definite markers, as shown below:

(36) *i=taktʔi* "the scarecrow" (BEJ_MV_NARR_09_jewel_48)
DEF.M=scarecrow
DET=CN.M

(37) *tiː=koːba* "the container" (BEJ_MV_NARR_09_jewel_43)
DEF.F=container
DET=N.F

Gender marking may be affected by a following modifier, as in the Semitic *status constructus*, represented in the corpus by Arabic varieties. In this construction, the head precedes a (nominal or pronominal) modifer in genitival constructions; a F head is in this case followed by the affixal F gender *-t* which is dropped in isolation and in other syntactic configurations:

(38) *ħkaːj-t haːjna*
story-F\CS Hayna
N-PNG NP
"The story of Hayna" (ARY_AB_NARR_01_014)

Although gender tends to be marked suffixally, it can also be expressed by a prefix or by both a prefix and a suffix (a circumfix), as in one of the Kabyle examples of (23) Kabyle *t-aqʃiʃ-t-agi* ('F-ABSL.SG-child-SG.F-PROX') 'this girl.'

One and the same language can use both prefixes and suffixes in different word classes or subclasses. E.g., in Gawwada, while gender is marked on nouns by a final vowel, it is marked by a prefixal consonant on, *inter alia*, the possessives, where it marks the gender of the head noun:

(39) *kaf-k-o h-aːju*
family-SING-M M-POSS.1SG
N-PNG-PNG PNG-PRO.POSS
"My family"(GWD_MT_NARR_002_009)

(40) *pij-e t-aːni* "our land" (GWD_MT_NARR_002_209)
land-F F-POSS.1PL
N-PNG PNG-PRO.POSS

This is further coupled for a few persons (in Gawwada, 2SG and 3SG) with gender-agreement with the possessor:

(41) *harʤ-ú=sa h-iːsi*
hand-M\DEM=DIST M-POSS.3SG.F
M-PRO.DEM=PTCL.DEM PNG-PRO.POSS
"That hand of hers" (GWD_MT_NARR_009_101)

3.4 Gender in personal and other pronouns

Gender agreement in the personal pronouns is very widespread among the languages in the corpus. The most common situation is the presence of three forms for the non-participants, a M.SG and a F.SG one, and a gender-indifferent PL one. Other languages have much richer systems, where gender is present also in the forms for the addressee, both sometimes Singular and Plural:

The following table shows the independent (emphatic or nominative according to the language) pronominal forms in a subset of the languages of the corpus:

Table 2. Independent personal pronouns in selected CorpAfroAs languages

	Beja	Gawwada	Hausa	Kabyle	Zaar
1SG	*(un)'ani*	*ʔano*	*niː*	*nəkk (i(ni))*	*mi*
2SG.M	*(um)ba'ruːk*		*kai*	*kəʃʃ(i(ni))*	
2SG		*ʔato*			*ki*
2SG.F	*(um)ba'tuːk*		*keː*	*kəmm(i(ni))*	
3SG.M	*(um)ba'ruː*	*ʔiso*	*ʃiː*	*nətstsa*	
3SG					*ʧi*
3SG.F	*(um)ba'tuː*	*ʔise*	*ita*	*nətstsat*	
1PL.M				*nəkkni*	
1PL	*(an)hi'nin*	*ʔine*	*muː*		*mì*
1PL.F				*nəkknti*	
2PL.M	*(am)ba'raːk(na)*			*kunwi*	
2PL		*hune*	*kuː*		*kì*
2PL.F	*(am)ba'taːk(na)*			*kunnəmti*	
3PL.M	*(am)ba'raː*			*nutni*	
3PL		*ʔusunde*	*suː*		*ʧì*
3PL.F	*(am)ba'taː*			*nutənti*	

The simplest system in Table 2 is represented by Zaar, where gender does not play a role at all in the personal pronouns (the same happens in Juba Arabic).

A very widespread pattern is exemplified by Gawwada, which has gender-specific forms for the 3SG only (the same obtains, among the languages of the corpus, for Afar, Ts'amakko and Wolaytta).

Other languages show different stages of complexity: Hausa opposes M and F forms in 3SG and 2SG, but not in the PL, while Beja has separate M and F forms both in the SG and PL and for both the 2nd and 3rd persons. The same is true in Hebrew. Arabic dialects vary between these possibilities, while, among the languages in our corpus, Kabyle represents the farthest development in gender marking, with separate M and F forms for all the persons except 1SG.

Gender-marking in pronouns therefore proceeds along the following cline:

Ø → 3SG → {2SG, 3SG} → {2, 3} → {1PL, 2, 3}

Gender in other pronominal categories is subject to heavy language-specific constraints: sometimes object (or oblique) pronouns follow the distribution patterns

of independent pronouns, often with a few reductions; e.g., the Kabyle pronominal clitics do not oppose 1PL.M and 1PL.F; the dative clitic of 3SG is likewise gender-neutral:

Table 3. Gender in Kabyle pronominal clitics

	Dative clitic	Absolutive clitic
1SG	*(i)yi*	*(i)yi*
2SG.M	*(j)ak*	*(i)k*
2SG.F	*(j)am*	*(i)kəm*
3SG.M		*(i)t*
3SG	*(j)as*	
3SG.F		*(i)tst*
1PL	*(j)ay*	*(j)ay*
2PL.M	*(j)awən*	*(i)kwən*
2PL.F	*(j)asənt*	*(i)tənt*
3PL.M	*(j)asən*	*(i)tən*
3PL.F	*(j)asənt*	*(i)tənt*

Sometimes different patterns emerge: in Gawwada, 2SG Oblique (used as direct objects and with postpositions) and Associative pronouns have separate M and F forms (in this as in other Cushitic languages there are no 3rd person object pronouns):

(42) *ho* *he*
2OBL.SG.M 2OBL.SG.F
hola *hela*
2ASSOC.SG.M 2ASSOC.SG.F

Gender may further affect other pronominal categories, such as the Interrogative pronouns of Gawwada and other Cushitic languages (both M and F forms contrasting with a single PL form):

(43) *h-ú-nka* "which one (M)?" *t-í-nka* "which one (F)?
M-M\DEM-which F-F\DEM-which
h-í-nka "which ones?"
PL-PL\DEM-which

3.5 Gender agreement

Given the huge typological differences between and within each Afroasiatic language group (cf. Frajzyngier 2012a), it is no wonder that agreement patterns are very diversified, too. A selection of the main features is exemplified below.

3.5.1 *Gender and gender agreement in Adjectives*

One of the simplest and more widespread agreement patterns involves the presence of the same (or a similar) allomorph of the head noun on the modifier, as in the following examples from Moroccan Arabic: in (44) a Ø-marked M.SG noun is followed by a Ø-marked adjective, while in (45) a F.SG noun is followed by an agreeing F.SG adjective. The same pattern is used in plural nouns: in Hebrew (Semitic; 46) a M.PL head is similarly followed by an agreeing adjective. In (44) M.SG, being the unmarked value for gender and number, is not overtly marked on either the head or the modifier:

(44) *təqliːd* *ʃaːdi*
tradition(-M.SG) common(-M.SG)
N.M ADJ.SG.M
"A common tradition" (ARY_AB_NARR_01_275)

(45) *əl=məʀʀ-a* *əl=taanj-a*
DEF=time-F DEF=second-F
DET=N.F-PNG DET=ADJ-PNG
"The second time" (ARY_DC_NARR_01_SFCC_068)

(46) *anaʃ-im* *umlal-im*
man-M.PL unfortunate-M.PL
N-PNG ADJ-PNG
"Miserable people" (HEB_IM_CONV_2_SP1_065)

Agreement with the Head operates across an intervening noun modifier in a genitival construction. In the following example from Hebrew, the Adjective (*meuʁgan-et*) agrees with its F Head noun (*xevʁa-t*), which is further modified by the noun (*jelad-im*) immediately following it:

(47) *xevʁa-t* *jelad-im* *meuʁgan-et*
society-F.SG child-M.PL organize\ACT.PTCP.F.SG
N.F-CS N-PNG V-PNG
"An organized children's company" (lit.: "a society of children, an organized one"); (HEB_IM_NARR_4_SP1_076)

Verb-final languages (such as the Cushitic and Omotic languages of the Horn of Africa) may have either Head-Modifier clause order (as represented in the corpus by Gawwada and Ts'amakko) or Modifier-Head (as in Afar and Wolaytta). In Wolaytta the adjectives do not agree in gender, number and case with the head noun:

(48) *woggá góda-í* "the big chief" (WAL_AA_NARR_05 lion_43)
big chief-M.NOM
ADJ N-CASE

(49) *sagi mhiːn eː-stʔeː*
big place 3SG.M-sit_down\REFL.IPFV
ADJ N.M PNG-DER.V1
"He stays in a remote place" (BEJ_MV_NARR_09_jewel_54)

Identification of a separate category of Adjectives is more problematic for other languages (such as, among the languages of the corpus, Gawwada, Ts'amakko, and especially Afar), where adjectival concepts may be conceived of in verbal terms. Gender agreement is nevertheless found, as in the following example from Gawwada:

(50) *ʃiːn-am-k-o* *piʃ-a=tta=kka* *ʔan=woʔ-i*
smear-PASS-SING-M white-M=INS=CONTR SBJ.1=want-PFV.1SG
V–V.DER-PNG-PNG ADJ-PNG=CASE=PTCL PRO.SBJ=V-TAM.PNG
"I want the white butter" (GWD_MT_NARR_006_033)

(51) *haːr-itt-e=si* *daʃamm-aj*
fish-SING-F=PROX big\INT-F
N-PNG-PNG=DEICT ADJ-PNG-PRO
"This very big fish" (GWD_MT_NARR_004_071)

3.5.2 *Gender and gender agreement in definite markers, demonstratives and other nominal modifiers*

A few languages possess definite markers. In Arabic and Hebrew they are invariable for gender and number, but in other languages (in)definite markers are gender-sensitive, as in Beja:

(52) *i=tarab=eː* *ti=balami=t=eː* *firʔa-tiːt*
DEF.M=half=3PL.ACC DEF.F=supply=INDF.F=POSS.3PL.ACC go_out-CVB.ANT
DET=N.M=PRO DET=N.F=DET=PRO V1.PNG
"They shared their food supply and" (BEJ_MV_NARR_07_cold_14)

(53) *naːt=* *ka=soː-j-a*
thing=INDF.F NEG.IPFV=CAUS-say\PFV-3SG.M
N.F=DET PTCL=DER.V1-V1.IRG
"He did not tell him anything (else)" (BEJ_MV_NARR_07_cold_75)

Demonstratives are likewise gender-marked in a few languages, such as in Moroccan Arabic, where *haːdaːk* (DIST.M) and *haːdiːk* (DIST.F) contrast with a gender-neutral form *diːk*.

Also in Wolaytta both the Distal and Proximal demonstratives have different gender-sensitive forms:

(54) *he-ge-á* *ʔússa*
DIST.DEM-M.NMLZ-DEF.M.ACC heifer
N.M
"That (group of) heifers" (WAL_AA_NARR_05_lion_20)

(55) *ha-nn-ó-kka* *ʔeh-éetí*
PROX.DEM-F-F.ACC-INCL bring-2PL.PRES.AFF.Q
DEICT-PGN-PGN-[ABSENT] V1-TAM
"Just this one (F) you bring?" (WAL_AA_NARR_05_lion_34)

Gawwada and Ts'amakko have no Definite markers and their Demonstratives are invariable; they have instead a special class of pronominal heads ('the one which...'). They are formed by a prefix gender marker (*h-*/*k-* for M and PL, *t-* for F) followed by the suffix gender markers of nouns, yielding semantically empty words. The combination of prefixes and suffixes unambiguously differentiates M, F, and PL, as exemplified in (56):

Table 4. The gendered pronominal heads of Gawwada and Ts'amakko

	Gawwada	Ts'amakko
M-M	*h-o*	*k-o*
F-F	*t-e*	*t-e*
PL-PL	*h-e*	*k-e*

(56) *h-o* *daːmm-a* *maːtt-a*
M-M big-M Maatta-M
PNG-PNG ADJ-PNG NP-PNG
"The big one is (called) Maatta" (GWD_MT_NARR_002_021)

In a few languages (e.g., Beja) numerals agree in gender with the head they modify; in others, gender agreement is restricted to lower numerals, and minimally to 'one,' as in the following example from Gawwada:

(57) *haːr-itt-e* *toʔ-ott-e=si*
fish-SING-F one-SING-F=PROX
N-PNG-PNG N.NUM-PNG-PNG=DEICT
"This one fish" (GWD_MT_NARR_004_056)

In other cases, gender agreement is limited to 'one' and 'two,' as in Kabyle (cf. (25) above), or in the following example from Hebrew:

(58) *ʃn-ej* *jelad-im* "two children" (HEB_IM_NARR_4_SP1_017)
two-M.PL child-M.PL
N-CS N-PNG

Plurality is often not marked on the noun:

(59) *qaw-h-o* *lakki* "two men" (GWD_MT_NARR_005_090)
man-SING-M two
N-PNG-PNG NUM

The number 'one' has separate M, F, and PL forms (the latter meaning 'some, a few') in Gawwada and Ts'amakko:

Table 5. Gendered 'one' in Gawwada and Ts'amakko

	Gawwada	Ts'amakko	
M	*toʔ-okk-o*	*do-okk-o*	("one-SING-M")
F	*toʔ-ott-e*	*do-ott-e*	("one-SING-F")
PL	*toʔ-okk-e*	*do-okk-e*	("one-SING-PL")

3.5.3 *Gender and agreement in verbs*

As anticipated, in AfroAsiatic subject nouns command gender-agreement on the form of the verb, although this is rare in the PL (and *a fortiori*, where existent, in the Dual). Gender-agreement for the addressee (the 2nd person) in the verbal form is found only in Kabyle and Tamasheq among the languages represented in the corpus; much rarer, and not found in our corpus, is gender-agreement for the speaker (the 1st person). In contrast, gender agreement for a non-participant (the 3rd person) in the SG is almost universal, with different M.SG and F.SG forms:

(60) *jhaːm* *dhaːj* *jʔ-i=t*
leopard(-M.SG) DIR come-AOR.3SG.M=COORD
SBJ.N.M POSTP V2.IRG-TAM.PNG=CONJ
"A leopard came towards them and" (BEJ_MV_NARR_15_leopard_016)

(61) *hiːja* *ma=lqːa-t* *ma=t-diːr*
3SG.F NEG1=find\PFV-3SG.F NEG1=3F-do\IPFV
PRO.IDP PTCL.NEG=V-PNG PTCL.NEG=PNG-V
"She did not know what to do" (ARY_AB_NARR_01_120)

(62) *ka=i-sərʱ-u=h*
REAL=3-take_to_graze\IPFV-PL=OBL.3SG.M
TAM=PNG-V-PNG=PRO.PNG
"They take them to graze" (ARY_AB_NARR_01_273)

(63) *ʒid-it* *idammən-iw* *ad=tn*
be_sweet\PFV-QLT.PL blood\ABSL.PL.M-POSS1.SG POT=ABSV3SPL.M
V.QLT-PRO N.COV-AFFX PTCL=PRO
t-sw-mt
SBJ2-drink\AOR-SBJ2PL.F
CIRC1-V13%-CIRC2
"My blood attracts you and you will drink it?" (KAB_AM_NARR_01_M_340)

Gender-agreement is also found in participial forms, as in Hebrew:

(64) *Hi* *haj-ta* *koʁ-et*
3F.SG be\PFV-3F.SG name\ACT.PTCP-F.SG
PRO.IDP V-AFFX.PNG V-PNG
"She used to call" (HEB_IM_NARR_4_SP1_097)

Subject gender marking may appear on phonologically separate morphemes, as in Hausa:

(65) *tà* *ʤe:* *gida:* *à* *gàʈi:-n-sù*
3SG.F.AOR go home at town-GEN-3PL.GEN
PNG.TAM V0 N PREP N-SYNT-PNG
"She went home to their town" (HAU_BC_NARR_02_SP1_021)

On the other hand, different paradigms in the same language may show various syncretism patterns, whereby gender and/or number oppositions are lost. For example, negative paradigms in Cushitic usually have a single form in the Past or Perfect; in Gawwada, a single form is used for all Singular subjects in the Negative Past.

3.6 The interaction of gender, number and case

In a few case-rich languages core cases may have different case forms for gender and number. In Wolaytta this happens for the Nominative, the Accusative, and the Definite and Indefinite Genitive. An affix *-í* marks the Nominative case in SG.M nouns and in PL nouns irrespective of gender, while *-á* signals SG.F nouns. The same syncretism of the M gender with the PL number is found in the Accusative,

with affixal *-á* marking both SG.M nouns and all PL nouns, and affixal *-ó* being reserved for SG.F nouns.

In Afar, gender plays a role together with the phonological shape of the word in conditioning the expression of the Subject and Genitive case: only vowel-final M nouns change their final vowel of the Basic (or Absolute) case form into *-í*. F nouns, as well as consonant-final M nouns (and a few exceptions of the vowel-ending ones) do not have overt case-marking. The accented nature of the M case affix causes a change in the accent pattern, which becomes the sole marker of case for M *i*-final nouns:

Table 6. Subject/Genitive case-marking on V-final M nouns in Afar

Absolute case	Subject and Genitive case	
áwka	*awkí*	"boy"
abbatiːnu	*abbatiːní*	"authority"
absiːse	*absiːsé*	"supervisor"
gínni	*ginní*	"demon"

Also, in Beja the Definite markers are case-sensitive for the Nominative (and the Accusative) cases:

(66) *uː=mha* "the morning" (BEJ_MV_NARR_09_jewel_59)
DEF.SG.M.NOM=morning
DET=SBJ.N.M

(67) *tuː=tiji* "the monster" (BEJ_MV_NARR_09_jewel_49)
DEF.SG.F.NOM=snake
DET=SBJ.N.F

The Associative (or Locative) case, which is the only morphological case of Gawwada and Ts'amakko (Cushitic), also has different gendered case forms:

Table 7. The gendered associative case in Gawwada and Ts'amakko

	Gawwada	Ts'amakko
ASSOC.M	*-ito*	*-ilo*
ASSOC.F	*-atte*	
ASSOC.PL	*-ete*	

The association between gender and number marking is pervasive; a good example of a typical relation in gender and number marking is shown in Tamasheq verb conjugational pattern. Basically, SG is marked by a prefix, generally *j-/i-*but

Ø in certain verb classes for M and *t-* for F, but no suffix. PL is instead marked by different gendered suffixes but no prefix:

Table 8. The interplay of gender and number marking in Tamasheq verbs

	prefix	suffix
3SG.M	*i-, j-*, Ø	Ø
3SG.F	*t-*	Ø
3PL.M	Ø	*-ɐn*, -V*n*
3PL.F	Ø	*-nɐt*

Gender switch coupled with number is common in many Cushitic languages, such as Somali (not represented in the corpus). In such a system, usually called 'gender polarity,' the gender of a noun of specific noun classes is reversed in the PL. The latter is usually marked by a suffix, but certain noun classes may be marked by gender switch alone.

Again in Cushitic, while gender is an inherent property of nouns, number is often not an obligatory category and may be seen as a matter of derivation (cf. Mous 2012: 361–363).

A special situation is provided by two closely-related languages of the Dullay branch of the Cushitic group (Gawwada and Ts'amakko; cf. Savà 2005), which are analyzed in the corpus as having a three-fold gender system, with PL alongside M and F, and a three-fold number system: preternumeral (or basic), SING, and Plurative (PLUR). Like M and F nouns, PL nouns are marked by a final vowel (typically, *-o* for M, and *-e* for both F and PL). Number marking may or may not be present in the shape of a noun.

The internal morphological composition of nouns may be captured by the following template

STEM ± NUMBER MARKING + GENDER MARKING

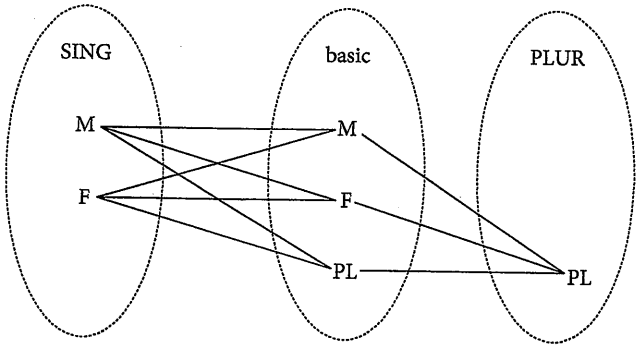
In short, number marking always precedes gender marking, and while overt expression of number may be absent, the marking of gender is always part and parcel of a noun form.

While the vast majority of count nouns are M or F in their basic form, a few are PL. Many mass nouns are PL. As anticipated, the gender of nouns denoting inanimate countable entities is not semantically motivated: they may either be Masculine, Feminine, or (in a minority of cases) Plural.

Number derivation operates from a basic noun, with the addition of either a Singulative or a Plurative affix before the gender marker. Against the free gender-association of basic nouns, Singulative nouns may only be either M or F in gender, and Plurative nouns are always PL in gender.

The interplay of gender and number in Gawwada is graphically illustrated in Table 9:

Table 9. The interaction of gender and number in Gawwada



The simplest case involves probably a number-unmarked (or basic) noun, semantically both a singular and a generic and either M or F in gender, and a number-derived Plurative expressing a plural:

(68) *paʃ-o* "field" *paʃ~ʃ-e* "fields"
field-M field~PLUR-PL

In (69) the referent is a sex-differentiated animate, and a Singulative Feminine form is further derived:

(69) *har-o* "dog" *har-itt-e* "bitch"
dog-M dog-SING-F
har~r-e "dogs (bitches)"
dog~PLUR-PL

For many nouns, having either animate or inanimate referents, no number-unmarked form is found: a Singulative acts both as a singular and a generic, against which a Plurative form acts as a plural:

(70) *ʔasp-itt-e* "storm" *ʔasp-idd-e* "storms"
storm-SING-F storm-PLUR.PL

Even a morphological Singulative may act as a semantic generic or collective, from which a further, or second, Singulative (with a singulative meaning) can be derived:

(71) *ʔinn-akk-o* "fly; flies" *ʔinn-att-akk-o* "a single fly"
fly-SING-M fly-SING-SING-M

Not infrequently, the morphologically simplest (i.e, not gender-marked) form is a semantic plural, from which a Singulative is derived:

(72) *ʔilk-e* "teeth" *ʔilk-akk-o* "tooth"
tooth-PL tooth-SING-M

As expected, semantics plays a role in the selection of gender, but not a decisive one; while (72) above may give the — partially correct — impression that the Plural gender is mostly selected for collective entities or mass nouns (from which a Singulative acts as a *nomen unitatis*), exceptions are by no means uncommon:

(73) *ker-e* "headrest" *ker-add-e* "headrests"
headrest-PL headrest-PLUR-PL

(74) *minn-e* "house" *minn-add-e* "houses"
house-PL house-PLUR-PL

Finally, (75) shows a Plural (and semantic collective) noun for an animate entity against which both a pair of gendered Singulatives (reflecting natural gender opposition) and a Plurative are derived:

(75) *ʔorr-e* "potters" *ʔorr-itt-o* "a potter (man)"
potter-PL potter-SING-M
ʔorr-itt-e "a potter woman"
potter-SING-F
ʔorr-add-e "(many) potters"
potter-PLUR-PL

3.7 AfroAsiatic languages as gendered languages par excellence?

Apart from Chadic, where many languages have no gender at all, Afroasiatic languages are 'gendered' languages *par excellence*: 'a few gender morphemes, foremost among them the F marker *-t*, show an extraordinary persistence across time and space, and may be seen as a shibboleth for the whole phylum. Also the gender system as a whole, with its binary distinction between a Masculine and a Feminine, is very persistent — no additions to the system of genders is observed (except for the possible use of Plural as a gender; cf. 6. above). Conversely, absence of gender is found only in Chadic and in typologically 'deviant' languages, such as the Arabic-derived Juba Arabic and Ki-Nubi creoles.

Gender is marked in a number of lexical categories and subcategories and plays a central role in agreement. On the other hand, gender in AfroAsiatic is not only a means of reference, but has acquired semantic functions such as diminutive, sometimes pejorative (Frajzyngier 2012a: 522). As mentioned above, gender

— alone or in combination with an affix — may come to mark number in the so-called 'gender polarity' of certain Cushitic languages.

As we have tried to show in this paper, all or most of these properties and values are evidenced and can be neatly investigated in CorpAfroAs.

Conclusion

The three studies conducted in this paper show that a corpus-based analysis can lead to interesting discoveries concerning features of Afroasiatic languages, provided some information is given in the grammatical sketch of the corresponding language.

Automatic retrieval of directional particles in the corpus allows a quick assessment of the distribution of those morphemes, as well as the semantic types of associated verbs. Contexts facilitate the analysis of discourse factors and modal dimensions. It appears that for the six languages under consideration, the directional morphemes have grammaticalized outside the domain of space and motion, and have acquired aspectual, modal or interactional dimensions. A thorough comparative study of those morphemes within AfroAsiatic is yet to be conducted, on the basis of this preliminary exploration.

The analysis of labels pertaining to the domain of Case shows that case systems largely integrate morphological marking of syntactic role. Various morphological means are used to mark Case, depending on the languages, and the corpus allows the end-user to retrieve the relevant forms, within their context. Thus, it is also possible, as was done in this paper, to investigate one case label (Nominative) across the corpus, and thanks to the associated grammatical sketches, conduct an informed comparison. However, the limits of a comparison based on labels and grammatical sketches is apparent in the fact that each case label has to be considered within a system. The paper by Frajzyngier and Mettouchi in this volume proposes an alternative solution for cross-linguistic comparison, to be implemented in a project funded by the Agence Nationale de la Recherche for 2013–2016, CorTypo.[10]

Finally, Gender is shown to be a pervasive category within AfroAsiatic, and CorpAfroAs provides rich and varied examples illustrating not only the morphological marking of Gender, but also its uses in agreement, for reference-tracking, and for semantic distinctions. Further, more fine-grained comparisons, for instance the cross-linguistic comparison of the use of gender for diminutive mark-

10. DOI: http://dx.doi.org/10.1075/scl.68.website.

ing, are yet to be conducted, on a larger corpus for which CorpAfroAs provides a pilot version.

References

Azeb Amha. 2012. 'Wolaytta Corpus'. Corpus recorded, transcribed and annotated by Azeb Amha. In Amina Mettouchi & Christian Chanard (eds). The CorpAfroAs Corpus of Spoken AfroAsiatic Languages. DOI: http://dx.doi.org/10.1075/scl.68.website. Accessed on 12/07/2013. (=WAL_AA)

Barontini, Alexandrine. 2012. 'Moroccan Arabic Corpus (Meknes)'. Corpus recorded, transcribed and annotated by Alexandrine Barontini. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: http://dx.doi.org/10.1075/scl.68.website. Accessed on 12/07/2013. (=ARY_AB)

Blake, Barry. 2001. *Case*. Cambridge: CUP. DOI: 10.1017/CBO9781139164894

Bybee, Joan L. 1985. *Morphology: A Study of the Relation between Meaning and Form* [Typological Studies in Language 9]. Amsterdam: John Benjamins DOI: 10.1075/tsl.9

Caron, Bernard. 1989. The verbal system of Ader Hausa. In Current Progress in Chadic Linguistics [Current Issues in Linguistic Theory 62], Zygmunt Frajzyngier (ed.), 131–169. Amsterdam: John Benjamins. DOI: 10.1075/cilt.62.08car

Caron, Bernard. 2012a. Zaar grammatical sketch. *ANR CorpAfroAs: A Corpus for Afro-Asiatic Languages*. http://dx.doi.org/10.1075/scl.68.website (18 April 2013).

Caron, Bernard. 2012b. 'Zaar Corpus'. Corpus recorded, transcribed and annotated by Bernard Caron. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: http://dx.doi.org/10.1075/scl.68.website. Accessed on 12/07/2013. (=SAY_BC)

Caron, Bernard. 2012c. Hausa grammatical sketch. *ANR CorpAfroAs: A Corpus for Afro-Asiatic Languages*. http://dx.doi.org/10.1075/scl.68.website (18 April 2013).

Caron, Bernard. 2012d. 'Hausa Corpus'. Corpus recorded, transcribed and annotated by Bernard Caron. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: http://dx.doi.org/10.1075/scl.68.website. Accessed on 12/07/2013. (=HAU_BC)

Comrie, Bernard. 1975. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge: CUP.

Corbett, Greville G. 1991. *Gender*. Cambridge: CUP. DOI: 10.1017/CBO9781139166119

Corbett, Greville G. 2006. *Agreement*. Cambridge: CUP.

Caubet, Dominique. 2012. 'Moroccan Arabic Corpus'. Corpus recorded, transcribed and annotated by Dominique Caubet. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: http://dx.doi.org/10.1075/scl.68.website. Accessed on 12/07/2013. (=ARY_DC)

Fillmore, Charles J. 1968. The case for case. In *Universals in Linguistic Theory*, Emmon Bach & Robert T. Harms (eds), 1–88. New York NY: Holt, Rinehart, and Winston.

Frajzyngier, Zygmunt. 1987. Ventive and centrifugal in Chadic. *Afrika und Übersee* 70(1): 31–47.

Frajzyngier, Zygmunt. 2012a. Typological outline of the Afroasiatic phylum. In *The Afroasiatic Languages*, Zygmunt Frajzyngier & Erin Shay (eds.), 505–624. Cambridge: CUP.

Frajzyngier, Zygmunt. 2012b. *A Grammar of Wandala*. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110218411

Hayward, Richard J. 1989. The notion of 'default gender': A key to interpreting the evolution of certain verb paradigms in East Ometo, and its implication for Omotic. *Afrika und Übersee* 72: 17–32.

Lazard, Gilbert. 1975. *La catégorie de l'éventuel. In Mélanges E. Benveniste*, 347–358. Paris: Société de linguistique.

Lux, Cécile. 2012. 'Kabyle Corpus'. Corpus recorded, transcribed and annotated by Cécile Lux. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: http://dx.doi.org/10.1075/scl.68.website. Accessed on 12/07/2013. (=TAQ_CL)

Malibert, Il-Il. 2012. 'Modern Hebrew Corpus'. Corpus recorded, transcribed and annotated by Il-Il Malibert. In Amina Mettouchi & Christian Chanard (eds). The CorpAfroAs Corpus of Spoken AfroAsiatic Languages. DOI: http://dx.doi.org/10.1075/scl.68.website. Accessed on 12/07/2013. (=HEB_IM)

Mettouchi, Amina. 1997. La particule D en berbère (kabyle): Transcatégorialité des marqueurs énonciatifs. In *Proceedings of the 16*th International Congress of Linguists, Paper No 0270. Oxford: Pergamon.

Mettouchi, Amina. 2011. The grammaticalization of directional clitics in Berber. Paper presented at the workshop 'Come and Go off the grammaticalization path', convened by Jenneke van der Wal and Maud Devos at the 44th Annual meeting of the Societas Linguistica Europaea.

Mettouchi, Amina. 2012. 'Kabyle Corpus'. Corpus recorded, transcribed and annotated by Amina Mettouchi. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: http://dx.doi.org/10.1075/scl.68.website. Accessed on 12/07/2013. (=KAB_AM)

Mettouchi, Amina & Frajzyngier, Zygmunt. 2013. A previously unrecognized typological category: The state distinction in Kabyle. *Linguistic Typology* 17(1): 30–59. DOI: 10.1515/lity-2013-0001

Mous, Maarten. 2012. Cushitic. In *The Afroasiatic Languages*, Zygmunt Frajzyngier & Erin Shay (eds), 342–422. Cambridge: CUP.

Newman, Paul. 1983. The efferential (alias causative) in Hausa. In *Studies in Chadic and Afroasiatic Linguistics*, Ekkehard Wolff & Hilke Meyer-Bahlburg, 397–418. Hamburg: Buske.

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

Savà, Graziano. 2005. *A Grammar of Ts'amakko*. Cologne: Rüdiger Köppe.

Savà, Graziano. 2012. 'Ts'amakko Corpus'. Corpus recorded, transcribed and annotated by Graziano Savà. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: http://dx.doi.org/10.1075/scl.68.website. Accessed on 12/07/2013. (=TSB_GS)

Tosco, Mauro. 2010. Why contrast matters: Information structure in Gawwada (East Cushitic). In *The Expression of Information Structure. A Documentation of its Diversity across Africa* [Typological Studies in Language 91], Ines Fiedler & Anne Schwarz (eds), 315–348. Amsterdam: John Benjamins. DOI: 10.1075/tsl.91.12tos

Tosco, Mauro. 2012a. The grammar of space of Gawwada. In *Proceedings of the 6th World Congress of African Linguistics, Cologne, 17-21 August 2009*, Matthias Brenzinger & Anne-Maria Fehn (eds), 523–532. Cologne: Rüdiger Köppe.

Tosco, Mauro. 2012b. 'Gawwada Corpus'. Corpus recorded, transcribed and annotated by Mauro Tosco. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: http://dx.doi.org/10.1075/scl.68.website. Accessed on 12/07/2013. (=GWD_MT)

Vanhove, Martine. 2012. 'Beja Corpus'. Corpus recorded, transcribed and annotated by Martine Vanhove. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: http://dx.doi.org/10.1075/scl.68.website. Accessed on 12/07/2013. (=BEJ_MV)

Vicente, Ángeles. 2012. "Moroccan Arabic Corpus (Ceuta)'. Corpus recorded, transcribed and annotated by Ángeles Vicente. In Amina Mettouchi & Christian Chanard (eds). *The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. DOI: http://dx.doi.org/10.1075/scl.68.website. Accessed on 12/07/2013. (=ARY_AV)