Reference-tracking in Hinuq

Diana Forker, MPI EVA Leipzig, forker@eva.mpg.de

Conference on the Languages of the Caucasus,
7–9 December 2007

1. Introduction

Hinuq is the smallest of the five Tsezic languages spoken in western Daghestan in the Caucasus, by about 500 speakers. It belongs to the Nakh-Daghestanian language family. Typologically it is suffixing and Ergative case-marking with the basic word order SOV.
As is typical for these languages it has a noun class system with five classes that are used to mark agreement between nouns in the Absolutive case and verbs, and coreferential adjectives or adverbials (table 1). Only verbs that begin with a vowel can have a class prefix.[1]

number \ class	I	II	III	IV	V
singular	*∅-*	*y-*	*b-*	*y-*	*r-*
plural	*b-*	*b-/r-*	*r-*	*r-*	*r-*

table 1

(1) [očordiyu rek'u-y **geni** b-ut'-o ažey-ƛ'os]$_i$
old person.OBL-ERG pear(III) III- collect-PRS tree-SRABL

[łono **karzina** b-ič'er-ho]$_{ii}$ [...] hibayłu zaman-a-ł
three basket(III) III-fill-PRS this.OBL time-OBL-INTERESS

∅-aq'-o uži welosiped-ƛ'o [...] hayło uži-y
I-come-PRS boy bicycle-SRESS this.OBL boy-ERG

∅-eze-n hayło-ƛ'or-no **b**-ik'ek'-o
I-look-CVB he.OBL-SRLAT-and **III**-steal-PRS

'An old man is collecting pears from a tree, he is filling three baskets. […] At that time a boy on a bike is coming. […] This boy, looking at him, is stealing it.'

As can be seen in example (1), the agreement marking can be ambiguous.
This talk will be devoted to the reference-tracking system of Hinuq. First, the referential status of the class prefixes will be explored. Then the various referential devices will be investigated and compared to each other with respect to the grammatical role of their referents and to the syntactical status of their clauses.

[1] But indeed, not all of these verbs have class prefixes.

2. Do the class prefixes function as reference-tracking devices?

The class prefixes code gender and number of Absolutive arguments (almost like pronouns). The first expectation may be that clauses with class prefixed verbs contain fewer other reference-tracking devices (e.g. NPs or pronouns) than clauses with other verbs. This prediction is not born out as a comparison of the REFERENTIAL DENSITY (as defined in Bickel 2003) shows (table 2). The Referential density is the ratio of overt to possible argument NPs which may be nouns or pronouns. For clause (i) in (1) it is 1 because all possible arguments are overtly expressed. But for clause (ii) it is 0,5 (1/2) because there is no overt agent.

For my corpus of 453 clauses[2] the average Referential density of the 81 clauses containing verbs without class prefixes is 0.516. That is, about the half of the arguments are overtly expressed. For the other 372 clauses the average Referential density is 0.636 which means that almost two third of the arguments are overt.

(2) [haw-no xece-n xižina-ma] [hoboži hibayrutow q'ono rek'e
she-and let-PFT hut-INESS then same two person

∅-aq'e-n hayłu-de-n]
I-come-PFT she.OBL-APUDESS2-and
'And they let her in the hut. Then the same two men came to her.'

Referential density	clauses	density	variance
no class prefix	81	0.516	0.016
class prefix	372	0.636	0.0175

table 2

Explanations:
(1) The difference in density is not big enough to be significant. But the sample should be enlarged, including only Pear stories.

(2) If a new, enlarged study reveals nevertheless a significante difference in density, there is a suggestion made by B.Bickel (p.c.): Maybe the class prefixes demand an activation of the case-frame for Absolutive case which leads in turn to more overt Absolutive NPs in clauses with verbs containing class prefixes.[3]

(3) If there is still no significant influence of the class prefix vs. non class prefix difference on the Referential density, this may be due to the fact that about the half of the verbs in my dictionary has a class prefix, the other doesn't. Thus although it seems that class prefixed verbs are more frequent in texts, it is due to chance whether the verbs has it or not. This is not a reliable way for indicating reference. It is not the function of agreement morphology to identify referents in discourse.

[2] Four 'pear stories' collected in the usual way by showing the 'Pear story film' to informants and recording their narratives, four other spontaneously uttered narratives and one story from a grammar, checked with an informant

[3] See also Bickel (2003:717): "Actual use of NPs is governed by […] the degree to which the syntax of one's language requires case-frame activation. Since case frames are morphologically hosted by and structurally associated with NPs, it is plausible that frequent case-frame activation leads speakers to actually use overt NPs in discourse."

3. Topic continuity

If we exclude class prefixes, referential devices in Hinuq are: proper names, other lexical noun phrases[4], pronouns, numerals and zero anaphora.

The corpus on which the analysis presented in this section is based is composed of ten 'pear stories' with a total amount of 469 clauses. There are no proper names in it, and numerals have been included to the pronouns. Thus, only three coding devices remain, namely lexical NPs, pronouns and zero anaphora. The distribution of them according to their argument position is given in the following table:

	lexical NPs	pronouns	zero	total
S	97	52	90	239
A	21	51	151	223
P	122	28	49	199
total	240	131	290	661

table 3

These data may be interpreted in two different ways. Either one groups together S and A following the widely accepted claim that there is a strong correlation between topic and subject (table 4). In the absence of any contrary indications the subject of a sentence will be interpreted as its topic and the predicate as a comment about this topic.[5]

	lexical NPs	pronouns	zero	total
S + A	118	103	241	462
P	122	28	49	199
total	240	131	290	661

table 4

Most of the pronouns and zero anaphors used in the narratives occur in the S and in the A position. This is in accordance with Givón's (1983) remarks about continuous topics. And looking at all referents in S and A position we find that only a relatively small amount of them (118 out of 462) are coded as lexical NPs whereas more than the half (241) occurs as zero anaphora.

For P the situation is reversed, almost two third of the P-positions are occupied by lexical NPs (122 out of 199). This may be explained by remembering that in topic-comment sentences[6] where the topic is the subject, the predicate is focused. Lambrecht (1994: 207) defines FOCUS as "the unpredictable or pragmatically non-recoverable element in an utterance." Therefore, objects which are part of the predicate are more often coded with full NPs than with pronouns or zero anaphora.

---

[4] It is difficult to differentiate between definite and indefinite NPs because bare nouns may be interpreted as either indefinite or definite according to the context. Sometimes the third person pronouns are used as definite articles. The result of comparing the RD values of these clearly definite NPs with all other NPs is for S, A and P similar: the RD value for unambiguously definite NPs is lower than for all other NPs. This is not surprising because definite NPs are usually not employed for introducing new referents, but for referring to already mentioned referents.

[5] A prerequisite of this interpretation is of course that the Ergative argument of a transitive sentence counts as the subject.

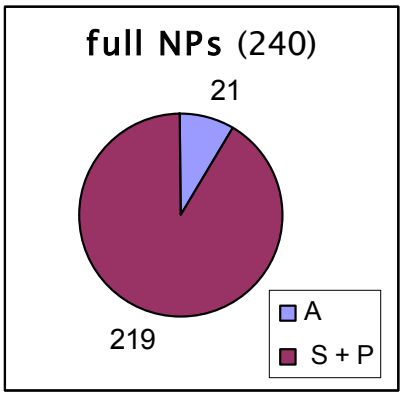
[6] which constitute the majority of the sentences in the corpus

Another way of interpreting table 3 is by grouping together S and P (table 5). This interpretation is supported by Du Bois' (1987) claims about the 'preferred argument structure'.

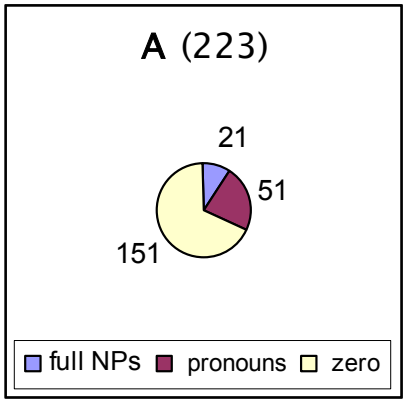
	lexical NPs	pronouns	zero	total
A	21	51	151	223
S + P	219	80	139	438
total	240	131	290	661

table 5

One of Du Bois' constraints is "Avoid lexical A's" (1987:823) which indeed matches well with the data in the above table 5 (see also graph 1), less than a tenth of the lexical NPs are found in the A position.[7] From another point of view just looking at the distribution of the three referential devices among all occurrences of referents in the A-position we find that the majority of them are coded by zero anaphora and only a relatively small proportion by lexical NPs (graph 2).



graph 1



graph 2

The other constrain formulated by Du Bois (1987:827) concerns the relation between the A-position and newly introduced referents; it says "Avoid new A's". The following table gives the number of new and given referents for each position.

	new	given	total
S	41	198	239
A	6	217	223
P	38	161	199
total	85	576	661

table 6

Indeed, we find that the majority of the referents occurring in the A-position have been already mentioned in the discourse.

Both constraints on the expressions filling out A-positions can be brought together under Givón's term of 'topic continuity'. According to Givón (1983:18), discourse is most naturally continuous. Continuity from one clause to the next is the most expected and unsurprising

[7] In fact, three speakers even did not employ any lexical NPs in the A-position.

situation and is therefore either not marked morphologically (zero anaphor), or is encoded with minimal morphological marking. Discontinuity, on the other hand, is surprising thus more morphologically and/or structurally more marked.
One of the measurements suggested by Givón in order to investigate topic continuity is the REFERENTIAL DISTANCE which assesses the number of clause boundaries between the previous occurrence in discourse of a referent and its current occurrence in a clause. Thus, in example (3) in clause (iii) the pronoun *hałoy* gets the referential index 1 because its referent is identical to the referent of the zero anaphora in the immediately preceding clause. The possessive pronoun *hayłos* in clause (v) gets also the referential index 1 in this clause, because its referent is identical to the class prefix (zero anaphora) of clause (iv). In turn, this class prefix (zero anaphora) gets 3, because the last mention of its referent was in clause (i) with the Genitive NP *užiš*.

(3) [hayło uži-š sede-z y-aš-iš šapka]$_{i}$
this.OBL boy-GEN1 one.OBL-DAT IV-find-PST hat

[hag y-aši-nos]$_{ii}$ [hezo hałoy šaruk'a-n y-ik'-no]$_{iii}$
this IV-find-SEQ then he.ERG whistle-and IV-beat-CVB

[hezor-no ∅-utir-iš]$_{iv}$ [toƛ-iš hayło-s šapka]$_{v}$
back-and I-turn.CAUS-PST give-PST he.OBL-GEN1 hat
(i) One found the hat of this boy. (ii) Having found it, (iii) he whistled (iv) and made (him) turn back (v) and gave his hat (to him).

The average Referential distance (RD) of the three selected coding devices for S, A and P is presented in the following table:

Referential distance	lexical NPs	pronouns	zero
S	11.001	1.48	1.861
A	4.717	1.669	1.689
P	10.806	1.595	2.143

table 7

There is a small difference between pronouns and zero anaphora, in any case the RD of zero anaphora is even higher than the RD of pronouns (which is contradicts the typological predictions that Givón (1983:30ff) makes on the basis of various studies of different languages). This shows that it is very common for Hinuq speakers to drop overt arguments so that in the appropriate context a single verb like *neƛiš* ('gave') on its own would be a complete utterance. On the other hand, the use of pronouns is not very common.
The RD value of lexical NPs is generally higher than that of pronouns and zero anaphors (they can 'look back' up to 11 clauses).
The low RD value for lexical As (which means high continuity) results from the interaction of the aforementioned constraints:
"Avoid new A's" "Avoid lexical As"
→ As are subjects, thus topical, continuous → less coding material

Then the question is: Why is the RD value of the expressions in the S-position which functions as subjects and should thus be topical (continuous) so high? The explanation is that intransitive sentences are used for the introduction of new referents. These presentational

sentences make use of indefinite NPs and a limited set of intransitive predicates like 'be', 'be at', 'live', 'arrive', 'have' and 'appear'. The sentences have typically VS word order and sometimes the numeral 'one' is used to mark thematically important referents (4).

(4) hes zoq'e-n rek'we
one be-PFT person
'There was a man.'

4. Finiteness

In order to form complex sentences, Hinuq makes extensive use of converbs. Many sentences are composed of one or more non-finite converbal clauses together with a finite clause. The converbal clauses often express temporal, causal or other circumstances of the main event described by the finite clause. Therefore, main clause and converb often share a number of arguments.

Referential density	clauses	density	variance
finite	207	0.736	0.0221
non-finite	150	0.515	0.0221

table 8

The difference in the Referential density between finite and non-finite clauses is quite clear. It may be explained by looking at the converbs which have been mainly used. By far the most common converb is the 'narrative converb' which coordinates two or more simultaneous or sequential events in a narrative sequence (5). These events typically share the actor with the event described in the finite clause. The actor is either completely left out or only once overtly expressed, usually in the finite clause.

(5) [k'onk'a-n b-ixer-no] [karžina-n k'onk'a-ƛ'o-n gor-no]
bike-and III-uplift-CVB basket-and bike-SRESS-and put-CVB

[∅-oxe-s]
I-leave-PST
'(He) uplifted the bike, put the basket on the bike and left.'

Other converbs (e.g. the converb meaning 'after') occur often with a subject that is different from the one of the main clause and therefore overt, but they are less used.

5. References

Bickel, Balthasar. (2003) 'Referential Density in Discourse and Syntactic Typology'. *Language* 79: 709-736.

Du Bois, John W. (1987) 'The discourse Basis of Ergativity'. *Language* 63: 805-855.

Givón, Talmy. (1983) *Topic Continuity in Discourse*. Amsterdam: John Benjamins.

Lambrecht, Knud. (1994) *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.