On the Placement and Morphology of Udi Subject Agreement

Berthold Crysmann
Computerlinguistik
Universität des Saarlandes
http://www.coli.uni-sb.de/~crysmann/papers/UdiAgr.html

DRAFT — Comments welcome

In this paper I shall address the form and distribution of subject agreement markers in Udi, a Caucasian language spoken in the Republics of Azerbaijan and Georgia (former USSR).[1] Subject-verb agreement in Udi, just like in most other languages, is a property strongly associated with the verb (or predicate): in Udi, the lexical class of the verb determines to a large extent the allomorphy observed in the agreement paradigm (cf. Table 1). Furthermore, the verb, and proper parts of it, can also serve as a host, to which the agreement marker can attach, sometimes exhibiting infixation. In the presence of negation, wh-expressions, and focussed constituents, however, realisation of subject agreement on the verb is illicit. Instead, the agreement marker attaches to one of the aforementioned syntactically independent elements. Even when separated from the verb, these markers exhibit more or less the same allomorphy as if they were attached to the verb.

I will argue that this morphosyntactic paradox can be resolved, if we assume that word-level signs can contribute more than one domain object on their DOM list (Kathol, 1995). Udi agreement markers are, then, introduced in the lexical entry of the verb, accounting for their morphological properties, while at the same time, they will enjoy a representation as an (independent) domain object, enabling us to

[1] The data cited in this paper are entirely taken from the work of Alice C. Harris, either from published articles (Harris, 1984, 1992, 1996) or from the handout of her 1997 ESCOL presentation (Harris, 1997). For those examples which Harris attributes to others, I retain the original reference alongside the source used for this paper.

As for the orthographic convention, I mark pharyngalised vowels with an underdot, and ejectives with an apostrophy.

capture their placement properties. In particular, I will build on the proposals for a linearisation-based analysis of morphosyntactic paradoxa in Fox and European Portuguese (Crysmann, 1999, to appear) and show how the ideas developed there can be fruitfully employed in the analysis of Udi as well. It is of note that the class of elements to which Udi agreement markers attach is actually quite similar to the class of syntactic elements triggering proclisis in European Portuguese.

1 Data

1.1 Syntactic distribution of Udi Agreement

In Udi, subject agreement is expressed by a set of so-called person markers ("Personalzeichen"; Schulze, 1982). These markers are inflected for both person and number, defining the two paradigms depicted in Table 1. The choice between the DIRECT and INDIRECT set of markers is a matter of lexically conditioned allomorphy which will be discussed below.

	DIRECT	INVERSION
1SG	-z(u)	-za
2SG	-n(u)	-va
3SG	-ne/-a	-t'u
1PL	-yan	-ya
2PL	-nan	-vạ
3PL	-q'un	-q'o

Table 1: Udi agreement markers

Apart from the verb, the agreement marker in Udi can attach to a variety of different hosts, including negation (1a), wh-phrases, and focussed constituents (1b).

(1) a. zu k'inigax te-z bese
I book.DAT not-1SG requested
'I didn't ask for a book.' (Harris, 1996, 212)

b. zu k'iniga-z bese
I book.DAT-1SG requested
'I asked for a book.' (Harris, 1996, 212)

c. * zu k'iniga-z te bese
I book.DAT-1SG not requested
(Harris, 1996, 212)

As observed by Harris (1992, 1996), whenever a marker of negation such as *te* is present in the clause, agreement obligatorily has to attach to it. Similarly, if the clause contains a wh-expression (2), the agreement marker must attach to this phrase.

(2) a. me išq’armux mano äizi-q’un karxesa?
this man.PL.ABS which village.DAT-3PL live
‘Which village do these men live in?’ (Harris, 1996, 210)

b. xinärmux ma-q’un taisa?
girl.PL.ABS where-3PL go
‘Where are these girls going?’ (Harris, 1996, 210)

c. šux-va ak’sa
who.DAT-2SG see
‘Who do you see?’ (Harris, 1996, 210)

d. ek’aluɣ-nu mia are?
why-2SG here came
‘Why have you come here?’ (Harris, 1996, 210; Dirr, 1928, 62)

e. et’e-a met’in t’ap’exa
why-3.SG he.ERG whipped
‘… [to see] why he whipped it.’ (Harris, 1996, 205; Dirr, 1928, 60)

While negation is always found in the vicinity of the verb in Udi, wh-expressions may be separated from the verb by other constituents, as in (2d,e). Still, the agreement marker must be right-adjacent to the wh-phrase.

Focussed constituents, typically occurring in a position immediately preceding the verb, pattern with wh-expressions and negation. Again, the agreement marker is attracted to the focussed constituent (3).

(3) a. … hünär rust’am-en-ne besa
heroic.deeds Rustam-ERG-3SG does
‘Rustam does heroic deeds.’ (Harris, 1996, 207; Pančviʒe, 1974, 238; folk tale)

b. me xinären tägsa k’inigiɣo-laxo-ne fikirbesa
this girl.ERG only book.PL.DAT-about-3SG thinks
‘Does this girl only think about books?’ (Harris, 1996, 208)

What these data show is that attachment to a focussed constituent is highly unselective as to the categorial status of the host, targeting argument NPs and PPs alike.

Furthermore, as illustrated by the examples in (4), the phrase which hosts the agreement marker need not always be adjacent to the verb, parallel to what has been observed for wh-phrases.

(4) a. sa q’oja kaft’ar-re pasčaɣun t’oɣol esa
one old woman-3.SG king beside comes
‘An old woman comes to the king.’ (Harris, 1996, 205; folk tale)

b. šin usin aytk’ayn, šot’in-q’a-n öküza xe tadi
who.ERG early speaks, he.ERG-SUBJ-3.SG ox.DAT water.ABS give
‘Whoever speaks first, let him give water to the ox.’ (Harris, 1996, 205; Pančviʒe, 1974, 149)

c. ạilen pis-ne ičux täsč'a
child.ERG badly-3.SG self.DAT behaves
'The child behaves (himself) badly.' (Harris, 1996, 205)

Without any focussed constituent (e.g. broad focus) or in Y/N questions, agreement is marked on the verb.

(5) a. vi baba ar-e-ne?
your father come-AOR-3SG
'Did your father come?' (Harris, 1992, 137)

b. ašlax b-e-ne.
matter do-AOR-3SG
'She took care of the matter.' (Harris, 1992, 137)

1.2 Morphological properties

We have already noted above that the set of Udi agreement markers is subject to a systematic allomorphy between two sets of paradigms (cf. Table 1). While most Udi verbs select their agreement markers from the DIRECT paradigm (6), a few verbs such as *ababaksun* 'know', *ak'sun* 'see', and *buqsun* 'love/want' (cf. Harris, 1984) choose the forms in the INVERSION set instead (7).

(6) a. zu a-r-e-zu k'wa
I.ABS hither-come-AOR-1SG home
'I came home.' (Harris, 1997, 1)

b. ɣar a-r-e-zu k'wa
boy.ABS hither-come-AOR-3SG home
'The boy came home.' (Harris, 1997, 1)

(7) a. Zu a-za-k'-sa šel läzätt'u pak.
I.ERG see$_1$-1SG-see$_2$-PRES good pretty garden.ABS
'I see a good, pretty garden.' (Harris, 1984, 247; Pančviʒe, 1974, 70)

b. Zu ek'a-za aba?
I.ERG what.ABS-1SG know
'What do I know?' (Harris, 1997, 1)

As argued in Harris (1984), selection of agreement markers is systematically related to the case marking patterns governed by the different verb classes. Udi is an ergative language where subject agreement typically targets an NP in the absolutive case if the verb is intransitive, but agrees with an ergative NP for transitive verbs. Direct objects in Udi may optionally be marked with dative case, instead of the usual absolutive, even if the verb subcategorises for another dative object. With a small set of verbs, the so-called inversion verbs, this case alternation is not restricted to the direct object, but may also apply to the (ergative) subject. Harris (1984) shows that it is exactly this class of verbs which also selects a particular set of agreement markers, i.e. the set labelled INVERSION in Table 1.

(8) a. Za a-za-k'-sa šel läzätt'u pak.
me.DAT see$_1$-1SG-see$_2$-PRES good pretty garden.ABS
'I see a good, pretty garden.' (Harris, 1984, 247; Pančviʒe, 1974, 70)

b. ɣarax te a-t'u-k'-sa xinär-a …
boy.DAT that see$_1$-3SG-see$_2$-PRES girl.DAT
'When the youth saw the girl …' (Harris, 1984, 248; Pančviʒe, 1974, 70)

What is crucial about the choice of agreement paradigm, is that it does not reflect the actual pattern of case assignment, but rather the lexical case assignment potential of the verb. Turning to wh-expressions (cf. (9) to (10)), however, we find that the agreement markers display the same allomorphic variation as if they were attached to the verb. Thus, even when the agreement marker attaches to a syntactic constituent outside the verb, its shape can only be determined on the basis of the verb's lexical properties.

(9) a. zu ma-z aš-besa
I.ERG where-1SG work-do
'Where do I work.' (Harris, 1992, 136)

b. ma-n aš-besa
where-2SG work-do
'Where do you work.' (Harris, 1992, 136)

c. me xinären ma-a aš-besa
this girl where-3SG work-do
'Where does this girl work.' (Harris, 1992, 136)

(10) a. zu ek'a-za aba?
I.ERG what.ABS-1SG know
'What do I know?' (Harris, 1997, 1)

b. un ek'a-va aba?
you.ERG what.ABS-2SG know
'What do you know?' (Harris, 1997, 1)

c. met'in ek'a-t'u aba?
she.ERG what.ABS-3SG know
'What does (s)he know?' (Harris, 1997, 2)

The same observation can, of course, be made regarding negation and focussed constituents.

Perhaps the most compelling piece of evidence in favour of a lexical derivation of Udi agreement markers is the possibility of these markers to occur infixed into the verb.

(11) … pasčaɣ-un ɣar-muɣ-on lašk'o-q'un-b-esa
king-GEN boy-PL-ERG wedding-3PL-do-PRES
'The king's sons married.' (Harris, 1997, 2, Dirr, 1928, 62)

Apart from complex verbs, where the agreement marker may be positioned between an incorporated element and a light verb (Harris, 1997), as illustrated in (11), the agreement marker may also appear infixed to monomorphemic verb roots (cf. Table 2), immediately preceding the root-final consonant. As should be evident from the contrast between the infixed transitive stems and the suffixed intransitives, the position of the agreement marker is much more likely to be conditioned by morphological factors than by prosodic ones.[2]

Infixed		Suffixed	
a-t'u-k'-sa	'sees (tr)'	ak'-ne-sa	'is visible (intr)'
bi-ne-t'-sa	'sows (tr)'	bit'-t'e-sa	'is sown (intr)'
bo-ne-x-sa	'boils (tr)'	box-ne-sa	'boils (intr)'
u-ne-k-sa	'eats (tr)'	uk-ne-sa	'is edible (intr)'
u̩-ne-ɣ-sa	'drinks (tr)'	u̩ɣ-ne-sa	'is drinkable (intr)'

Table 2: Infixed and suffixed agreement (Harris, 1997)

Despite this ability to penetrate the verb root in certain cases, attachment is still governed by the usual syntactic factors observed above: if e.g. a negative marker occurs in the sentence, infixed realisation is blocked (12b) and the agreement marker obligatorily attaches to negation (12c). The same observation can be made with respect to wh-expressions (compare (2c) and (7a)).

(12) a. Za gölö bu-za-q-sa bez a̩il-o̩ɣ-o̩x
me.DAT very love$_1$-1SG-love$_2$-PRES my child-PL-DAT
'I love my children very much.' (Harris, 1984, 248; Dirr, 1904)

b. * Manu ukalšey te bu-va-q'-sa
which food.ABS not love$_1$-2SG-love$_2$-PRES
(Harris, 1997, 3)

c. Manu ukalšey te-va buq'-sa
which food.ABS not-2SG love-PRES
'Which food do you like?' (Harris, 1997, 3)

To conclude the description of the empirical facts, we are faced with a dilemma: from a morphological point of view, both allomorphy and morphotactics suggest an affixal status for the agreement marker, while its placement properties would favour a treatment of the agreement marker as independent syntactic items.

[2]Note that this pattern poses a challenge for McCarthy and Prince (1993)'s theory of infixation, because here, unlike the famous Tagalog case, infixation does not reduce the number of codas. As infixation in Udi does not reduce prosodic markedness, it just has one alignment violation more against it.

2 Analysis

In the analysis I am going to propose, I will first present a formalisation of the case marking property of Udi verbs which will pave the way for a principled account of the allomorphic variations found in the set of agreement markers. Next, I will propose a lexical derivation of agreement markers assigning them a representation as a domain object on the verb's DOM list, distinct from that of the verbal head. I will show how the linear properties of the agreement affixes may interact with other morphological material, deriving, e.g. infixation into the stem. On the basis of the lexical representations so defined, I will formulate a minimal set of linearisation constraints that will capture the patterns observed.

2.1 Case marking

In her analysis of the case marking patterns of Udi verbs, Harris (1984) suggests that direct objects are subject to two optional case marking rules: one, which would assign them the expected absolutive case, and another one, which can freely assign the dative (or objective) case to an object. This proposal, which has been developed in the framework of Relational Grammar (RG), captures quite neatly, that neither ergatives, nor the absolutive subject of an intransitive are a target of case alternation. With inversion verbs, however, the (initial) subject is demoted to an indirect object, thereby accounting for the fact that these verbs also exhibit alternation with ergative subjects. She further claims that case marking rules may freely target initial and final representations.

Adopting Manning and Sag (1999)'s theory of ergativity, the basic insight of this proposal can easily be formalised in terms of three case marking rules, together with appropriate entries for direct and inversion verbs.

$$\left[\text{COMPS} \quad \left\langle \ldots \boxed{1}\text{NP}[str] \ldots \right\rangle\right] \rightarrow \left[\text{COMPS} \quad \left\langle \ldots \boxed{1}\text{NP}[dat] \ldots \right\rangle\right] \vee$$
$$\left[\text{ARG-ST} \quad \left\langle \boxed{1}\text{NP}[str], \ldots \right\rangle\right] \rightarrow \left[\text{ARG-ST} \quad \left\langle \boxed{1}\text{NP}[abs], \ldots \right\rangle\right] \vee$$
$$\left[\text{ARG-ST} \quad \left\langle \boxed{1}, \boxed{2}\text{NP}[str], \ldots \right\rangle\right] \rightarrow \left[\text{ARG-ST} \quad \left\langle \boxed{1}, \boxed{2}\text{NP}[erg], \ldots \right\rangle\right]$$

Figure 1: Udi case marking

With an ordinary transitive verb (Figure 2), the absolutive and ergative rules straightforwardly predict the basic realisations for structural case, assigning the absolutive to the least oblique argument on ARG-ST, whereas the next element will receive the ergative case. The specific linking of ARG-ST elements to the valence lists then derives the typical reversal of structural case in ergative languages (cf. Manning and Sag, 1999). Alternatively, case marking in Udi can proceed via

marking of the complements. Given the lexical representation in Figure 2, it is clear that the ergative subject can never be affected in direct verbs.

$$\begin{bmatrix} \text{SUBJ} & \langle \boxed{2} \rangle \\ \text{COMPS} & \langle \boxed{1} \mid \boxed{3} \rangle \\ \text{ARG-ST} & \langle \boxed{1}\text{NP}[\textit{str}], \boxed{2}\text{NP}[\textit{str}] \mid \boxed{3} \rangle \end{bmatrix}$$

Figure 2: Direct verbs

Inversion verbs however, observe a different linking pattern, where the two least oblique arguments are both linked to the COMPS list. While ergative and absolutive marking may still derive the standard case marking patterns by means of their position on ARG-ST, either of them, or both, can also be marked as datives.

$$\begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \langle \boxed{2}, \boxed{1} \mid \boxed{3} \rangle \\ \text{ARG-ST} & \langle \boxed{1}\text{NP}[\textit{str}], \boxed{2}\text{NP}[\textit{str}] \mid \boxed{3} \rangle \end{bmatrix}$$

Figure 3: Inversion verbs

Having developed an appropriate lexical representation for direct and inversion verbs, I shall now proceed towards an analysis of the agreement markers. Note that the HPSG variant of the original RG analysis is also much more precise as to what data structures the different case marking rules may operate on.

2.2 Lexical representation of Udi agreement

In order to model the allomorphic and morphotactic properties of Udi agreement markers I shall propose a representation in terms of morphological schemata (Riehemann, 1994), i.e. partial descriptions of possible words. The schema in Figure 4 relates the presence of an affixal formative (on MORPH) to the subcategorisation requirement for a third person singular subject, thereby modelling the syntactic effect of agreement.

Similarly, Figure 5 specifies an identical requirement for the highest complement, while at the same time it restricts the SUBJ value to the empty list. As a morphological reflex, the appropriate formative from the inversion set is introduced.

Following proposals by Kathol (1995) and Crysmann (1999), I assume that a word can introduce more than one domain object on its DOM list and, more specifically, that the agreement marker in Udi is represented as a separate domain object, distinct from the verbal head.[3] The order, in which these domain objects may sur-

[3]For the purposes of this paper I characterise the domain object holding the agreement marker as a non-predicative verb.

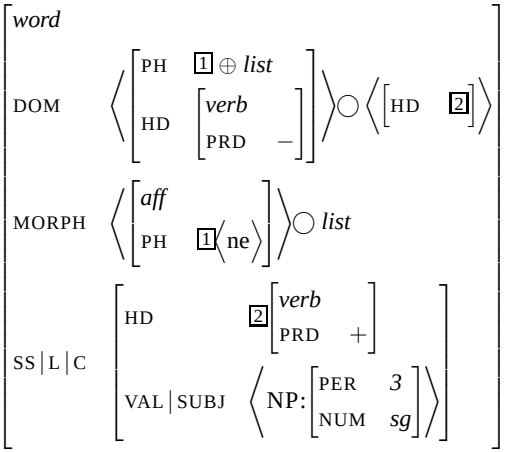

Figure 4: Direct agreement inflection

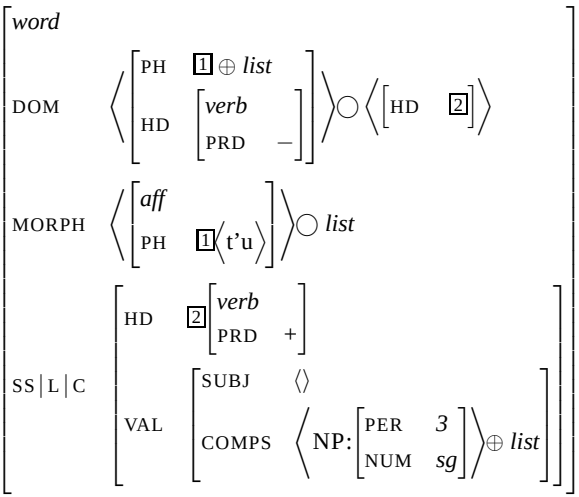

Figure 5: Inverse agreement inflection

face is lexically underspecified. This representation provides the key mechanism to resolve the morphosyntactic paradoxon we established in the preceding section.

As we have seen in Table 2, transitive stems may be discontinuous, allowing the agreement marker to be positioned before the stem-final consonant. Again, we can state the potential discontinuity by means of a word-formation schema, mapping the morphophonological contribution of the stem to domain objects. Figure 6 licenses complex words, where a stem-final consonant is mapped onto a final domain object, the rest of the stem's phonology, however, is restricted to be initial on some domain object, whose categorial information is structure-shared with the categrial information of the entire sign.

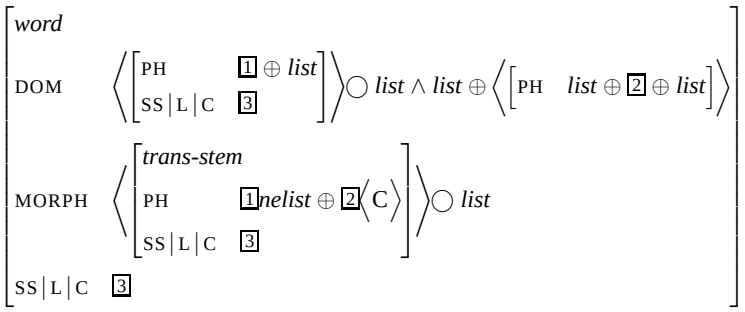

Figure 6: Discontinuous transitive stems

For inflected verbs, the schema in Figure 6 will be unified with one of the schemata in Figures 4 or 5. As these schemata restrict the DOM list to contain exactly two domain objects, only two situations may arise: either the first part of the stem is initial on the left domain object, or it is initial on the right domain object. In the former case, the agreement marker is coerced to appear as the initial string on the right domain object, combining with the discontinuous stem-final consonant (⟨[PH ⟨a⟩],[PH ⟨t'uk'-⟩]⟩), while in the latter case, the agreement marker is represented as the initial string on the left domain object, yielding a contiguous representation of the stem phonology (⟨[PH ⟨t'u⟩],[PH ⟨ak'-⟩]⟩). Correspondingly, the categorial specification of the DOM list will be ⟨[PRD +], [PRD −]⟩ in the first case, and ⟨[PRD −], [PRD +]⟩ in the second.

2.3 Syntactic distribution

On the basis of the lexical introduction of agreement markers as separate non-predicative verbal domain objects, linearisation constraints can now be formulated which derive the syntactic attachment of agreement.

As depicted in Figure 7, the relative order of the agreement marker and the verb is restricted to immediate precedence in the case of suffixation to the verb, while only ordinary precedence is required for those cases where the agreement marker attaches to a syntactic host on the left. If, however, the domain object corresponding to the agreement marker indeed precedes the verb, there must be an appropriate

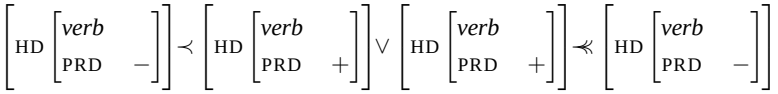

Figure 7: Placement properties of the agreement marker

licensor in the clause, i.e. negation, a wh-phrase, or a focussed constituent, to which the agreement marker could attach.

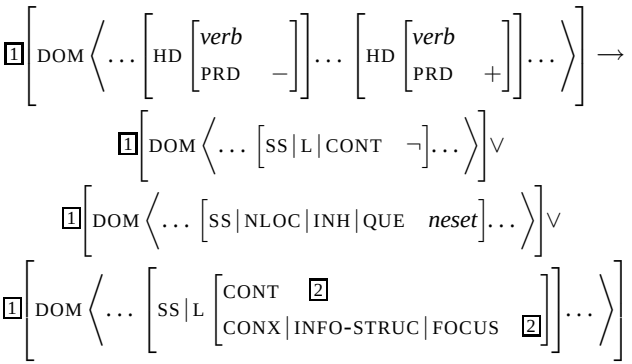

Figure 8: Licensing for preposed agreement

The preposing licensors, as characterised in Figure 8, are identified on the basis of SYNSEM information only. The representation of wh-expressions as domain objects with a non-empty QUE value immediately captures both simple wh-words as well as pied-piped wh-phrases, like the one in (2a), while it correctly ignores the homophonous relative marker. For the representation of focussed constituents, I adopt the representation suggested by Przepiórkowski (1999), who integrates Vallduví's theory of information structure with the concept of order domains. If a licensor, so described, is indeed present in the clause, the agreement marker has to appear adjacent to it. This effect can be obtained by constraints such as those in Figure 9.

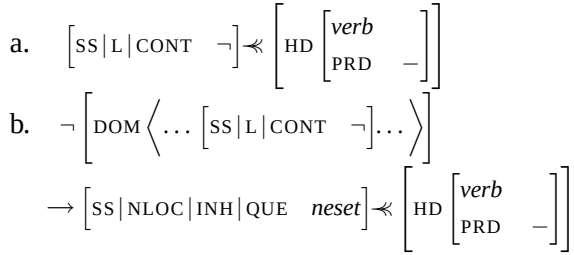

Figure 9: Attachment constraints (partial)

As we have seen in section 1.1, negation always attracts the agreement marker, preempting attachment to all other potential hosts. Thus, the attachment constraint for wh-phrases contains an additional restriction regarding the absence of negation from the clause.

3 Conclusion

I have argued that the agreement marking in Udi presents a challenge for linguistic analysis in that its allomorphy and its morphotactics (in particular infixation) suggest a lexical treatment, yet the distribution of these markers requires a certain degree of syntactic transparency. I have proposed an analysis in terms of Linearisation HPSG which tries to reconcile these two seemingly paradoxical properties by assuming that the agreement marker is lexically introduced as a separate domain object on the predicate. This basic assumption is able to capture a number of important observations: first, the uniform introduction of the agreement marker by means of a lexical schema for verbs immediately accounts for the fact that there is only one agreement marker present in any clause, regardless of where it is attached. Second, the uniformity of the marker including its systematic allomorphy lexically triggered by the verb is a direct consequence of the lexically-based linearisation account.

The treatment of morphologically conditioned infixation by means of discontinuous stems and their mapping to the PHON values of different domain objects is actually quite similar to the analysis suggested for discontinuous agreement markers in Fox by (Crysmann, 1999). The linearisation properties of the Udi agreement marker also share a significant degree of similarity with the placement properties of pronominal affixes in European Portuguese: while in Udi, negation, wh-expressions, and focussed constituents enforce the attachment of the agreement marker, a similar set of triggers also attracts the clitic cluster in Portuguese, triggering a preposing effect.

References

Crysmann, B. (1999), "Morphosyntactic paradoxa in Fox," in G. Bouma, E. Hinrichs, G.-J. Kruiff and R. Oehrle, eds., *Constraints and Resources in Natural Language Syntax and Semantics*, Studies in Constraint-Based Lexicalism, pp. 41–61, Stanford, CSLI publications. 2, 8, 12

Crysmann, B. (to appear), "Clitics and coordination in linear structure," in B. Gerlach and J. Grijzenhout, eds., *Clitics from Different Perspectives (Working Title)*, Linguistik Aktuell/Linguistics Today, John Benjamins. 2

Dirr, A. (1904), "Grammatika udinskogo jazyka," *Sbornik materialov dlja opisanija mestnostej i plemen Kavkaza* **33**. 6

Dirr, A. (1928), "Udische Texte," *Caucasica* **5**, 60–72. 3, 5

Harris, A. (1984), "Case marking, verb agreement, and inversion in Udi," in D. M. Perl-

mutter and C. G. Rosen, eds., *Studies in Relational Grammar 2*, pp. 243–258, University of Chicago Press, Chicago. 1, 4, 5, 6, 7

Harris, A. (1992), "The particle *-a* in Udi," in H. I. Aronson, ed., *The Non-Slavic Languages of the USSR*, pp. 135–156, Chicago, Chicago Linguistic Society. 1, 2, 4, 5

Harris, A. (1996), "Focus in Udi," in H. I. Aronson, ed., *NSL 8: Linguistic Studies in the Non-Slavic Languages of the Commonwealth of Independent States and the Baltic Republics*, pp. 201–220, Chicago, Chicago Linguistic Society. 1, 2, 3, 4

Harris, A. (1997), "What's in a word? the problem of endoclisis in Udi," paper presented at the Eastern States Conference on Linguistics (ESCOL), Yale University, CT. 1, 4, 5, 6

Kathol, A. (1995), "Linearization-based German syntax," Ph.D. thesis, Ohio State University. 1, 8

Manning, C. and I. Sag (1999), "Dissociations between argument structure and grammatical relations," in A. Kathol, J.-P. Koenig and G. Webelhuth, eds., *Lexical and Constructional Aspects of Linguistic Explanation*, Studies in Constraint-based Lexicalism, CSLI, Stanford. 7

McCarthy, J. and A. Prince (1993), "Generalized alignment," in G. Booij and Y. van Marle, eds., *Yearbook of Morphology*, pp. 79–153, Kluwer, Dordrecht. 6

Pančviʒe, V. (1974), *Uduri enis gramat'ik'uli analizi*, Mecniereba, Tbilisi. 3, 4, 5

Przepiórkowski, A. (1999), "Negative polarity questions and Italian negative concord," in V. Kordoni, ed., *Tübingen Studies in Head-Driven Phrase Structure Grammar*, Arbeitspapiere des Sonderforschungsbereichs 340 Bericht Nr. 132, Seminar für Sprachwissenschaft, Universität Tübingen. 11

Riehemann, S. (1994), "Morphology and the hierarchical lexicon," manuscript, CSLI, Stanford. 8

Schulze, W. (1982), *Die Sprache der Uden in Nord-Azerbajdžan*, Harrassowitz, Wiesbaden. 2