ESSEX RESEARCH REPORTS IN LINGUISTICS

Volume 48

Contents

EDITORIAL NOTE

The papers collected in this volume were presented at the Second York-Essex Morphology Meeting (YEMM2) which was held on 27-28 November 2004 in the Department of Language and Linguistics, University of Essex.

YEMM2 was made possible through the financial and material support of the Department of Language and Linguistics at the University of Essex and the Linguistics Association of Great Britain (LAGB). We would like to express our gratitude to both these institutions for their generosity.

We would also like to thank Professor Martin Atkinson, head of the Department of Language and Linguistics, for his encouragement and help in organizing YEMM2, as well as Dr Sonja Eisenbeiss, series editor of the *Essex Research Reports in Linguistics* for her help with the publication of this volume.

Finally, YEMM2 would not have become such a valuable and enjoyable event without
the efforts and enthusiasm of all the participants and contributors in the workshop.

Ryo Otoguro

Gergana Popova

Andrew Spencer

ESSEX RESEARCH REPORTS IN LINGUISTICS

Essex Research Reports in Linguistics present ongoing research activities of the members of the Department of Language and Linguistics. The main purpose of the reports is to provide a quick publication outlet. The reports have pre-publication status. The contents, form and distribution of the reports lie in the hands of the authors, and copyright remains with the author(s) of the reports as well. Most of the reports will subsequently appear in revised form as research articles in professional journals or in edited books.

We are happy to exchange Essex Research Reports in Linguistics for working papers (or other publications) in linguistics and related disciplines. If you require extra copies of earlier issues of ERRL, please contact one of the author(s) directly.

Sonja Eisenbeiss
Department of Language & Linguistics
University of Essex
Wivenhoe Park
Colchester C04, 3SQ
UK

Email: *seisen@essex.ac.uk*

URL: *http://www.essex.ac.uk/linguistics/wkpaps*

1 The Location of Deponency*

Matthew Baerman (Surrey Morphology Group, University of Surrey)

1.1 Introduction

Deponency is a mismatch between morphosyntactic values and morphological form which was first described for Latin. The *deponents* of Latin are verbs which are morphologically passive but nevertheless behave syntactically as active verbs. For example, contrast the normal verb *amant/amantur* in (1) with the deponent verb *hortantur* in (2).

(1) quae ex se natos ita amant ad quoddam tempus
which.N.PL from self.ABL born.ACC.PL thus love.3PL to certain.ACC.SG time.ACC.SG
et ab eis ita amantur
and from them.ABL this love.3PL.PASSIVE

'...which [animals] thus love their offspring for a certain time and thus are loved by them.' (Cicero, *De Amicitia*, Chapter VIII)

(2) me=que hortantur ut magno animo sim
me.ACC=and exhort.3PL that great.ABL.SG spirit.ABL.SG be.1SG.SUBJ

'...and they exhort me to be of good courage' (Cicero, Epistulae ad Atticum, book 11, letter 6)[1]

In (1) the verb 'love' illustrates the regular alternation between the active form *amant* and the passive form *amantur*. In (2), the verb 'exhort', *hortantur*, has the same ending as the passive *amantur*, but is active (and transitive at that). The alternation in (1) is productive, available to any transitive verb, while deponent verbs such as *hortor* are an exceptional, lexically-specified class. This presents an obvious challenge to morphological description: passive morphology has a clear function for the majority of verbs, but in some cases it has the opposite function (or alternatively, no function). Further, since Latin passive morphologically is highly heterogeneous (varying for person, tense-aspect-mood, etc.) this mismatch must be systematic, and cannot be attributed to the quirky behaviour of a single affix. But in spite of the fact that deponency has been a familiar notion since Classical times, it is only recently that

*Thanks to Greville Corbett, Dunstan Brown and Andrew Hippisley for helpful discussion, as well as the audience at the second YEMM meeting. This work has been supported by the Economic and Social Research Council under grant number RES-000-23-0375.

[1]This example was found using the Perseus Lookup Tool via http://www.perseus.tufts.edu

Essex Research Reports in Linguistics 47, 1–19.
Ryo Otoguro, Gergana Popova and Andrew Spencer (eds.).
Department of Language and Linguistics, University of Essex,
Colchester, United Kingdom.

its ramifications for our model of morphology have been seriously considered (e.g. Börjars *et al.* 1996, Corbett 1999, Embick 2000, Sadler and Spencer 2001, Stump 2002, Kiparsky 2005, Bobaljik and Branigan forthcoming. As yet there is no consensus on what the correct formal representation of deponency should look like. In this paper, I consider the question of how deponency is defined in a lexical entry: in terms of morphosyntactic values, or in terms of forms? To this end I look at some aspects of the familiar Latin paradigm, focussing particularly on semi-deponents, where deponency is restricted to one portion of the paradigm. I then look at comparable phenomena from other languages to see if they resolve the unanswered questions.

1.2 The Latin Paradigm

An abridged synopsis of the Latin verbal paradigm is given in (3). Only third singular and/or masculine nominative singular forms are given, and only one conjugation class is considered (the third). Nevertheless, this brief sketch is sufficient to make the necessary points.

(3) Normal verb 'rule' and deponent verb 'follow'

		'rule'		'follow'	
		active	passive	active form	passive form
A	PRES IND	reg-it	reg-itur		sequ-itur
	IMPERF IND	reg-ēbat	reg-ēbatur		sequ-ēbatur
	FUT IND	reg-et	reg-ētur		sequ-ētur
	PRES SUBJ	reg-at	reg-ātur		sequ-ātur
	IMPERF SUBJ	reg-eret	reg-erētur		sequ-erētur
	IMP PRES	reg-e	reg-ere		sequ-ere
	IMP FUT	reg-itō	reg-itor		sequ-itor
	INF PRES	reg-ere	reg-ī		sequī
B	PERF IND	rēx-it	rēct-us est		secūt-us est
	PLUPERF IND	rēx-erat	rēct-us erat		secūt-us erat
	FUT PERF	rēx-erit	rēct-us erit		secūt-us erit
	PERF SUBJ	rēx-erit	rēct-us sit		secūt-us sit
	PLUPERF SUBJ	rēx-isset	rēct-us esset		secūt-us esset
	INF PERF	rēx-isee	rēct-us esse		secūt-us esse
C	PART PERF	——	rēct-us	——	secūt-us
D	SUPINE	rēct-um	——	secūt-um	——
	INF FUT	rēct-ūr-us esse	——	secūt-ūr-us esse	——
	PART FUT	rēct-ūr-us	——	secūt-ūr-us	——
	PART PRES	reg-ēn-s	——	sequ-ēn-s	——
	GERUND	reg-end-ī	——	sequ-end-ī	——

I have divided the paradigm into four sections (from 4A to 4D) according to the distribution of forms between active and passive.

(4) A. Active and passive are synthetic, and both formed from the present stem.

B. The active is synthetic, formed from the perfect stem, while the passive is periphrastic, consisting of the perfect participle (based on the supine stem) plus the copula/auxiliary 'be'.

C. There is only a passive form (the perfect participle).[2]

D. There are only active forms (based on the supine[3] or present stem).

Deponent verbs have only passive forms, where these are available (4A, 4B, 4C); otherwise, they have active forms (4D).

1.3 Analyses: rules versus underspecification

Two approaches have been taken to the description of deponency, treating it either as the result of a rule (Sadler and Spencer 2001, Stump 2002) or as the reflex of the static lexical specification of stems and affixes (Embick 2000, Kiparsky 2005).

The rule-based approaches depend on the distinction between a syntactic paradigm and a morphological paradigm; i.e. the paradigm of morphosyntactic values (the syntactic paradigm) and the paradigm of inflected forms (the morphological paradigm) are construed separately. By default, the two paradigms are congruent, but mismatches are possible, and deponency is a kind of mismatch. Sadler and Spencer (2001:91) express this through the rule of referral shown in (5):[4]

(5) If a lexeme L is marked [Class:Deponent] then for all feature sets σ,
if [Class:Deponent:Full] $\subset \sigma$ then [m-Voice:Active] $\Rightarrow$ [m-Voice:Passive]

That is, for deponent verbs, the active cells of the morphological paradigm take their form from the corresponding cells in the passive morphological paradigm. Stump (2002:173) expresses the mismatch in terms of paradigm linkage shown in (6):

(6) **Deponent**: [If L:[Primary root = R]
and MPF_{Latin}(<R, σ/{passive}>) = <W, σ/{passive}>,
then SPF_{Latin}(<L,σ >) = <W,σ >]

That is, the active values in a lexeme's syntactic paradigm are exceptionally linked to the corresponding passive cells in its morphological paradigm. Although the two proposals are technically distinct (Sadler and Spencer's (2001) rule involves the relationship between two morphological paradigms, while Stump's (2002) rule involves the link between a syntactic paradigm and a morphological paradigm), they are in the same spirit, based on the assumption that morphology is autonomous, and that the relationship between meaning and form must be stipulated.

[2]In a few verbs, the perfect participle has an active sense: *adol$\bar{e}$sco* 'grow up', *c$\bar{e}$no* 'dine', *placeo* 'please', *prandeo* 'have breakfast/lunch', *p$\bar{o}$to* 'drink', and, optionally, *j$\bar{u}$ro* 'swear' (Bennet 1942:78)

[3]By default, the future active participle (and the future infinitive which is formed from it) is based on the supine stem. In a few verbs, though, it has a distinct stem, e.g. the supine stems of *lav$\bar{a}$re* 'wash' and *parere* 'bring forth, bear' are *laut-* and *part-*, but their future active participles have the stems *lav$\bar{a}$t-* *t$\bar{u}$r-*and *parit-* *$\bar{u}$r-* (Bennet 1942:82).

[4]This is an abbreviated version of the rule they give, omitting reference to the semi-deponents, which are instead discussed below.

In reaction to such proposals, Kiparsky (2005) offers an alternative analysis based on the rejection of manipulative devices such as rules of referral or of paradigm linkage. In this model (in the framework of Lexical Morphology), the inflectional paradigm is construed as that which falls out from the compilation of the static lexical properties of stems and affixes, and deponency is seen as a by-product of underspecification. Kiparsky proposes that both verb stems and endings may be marked for the feature [±passive]. Specification for this feature must be compatible between stem and ending. The stem of a normal verb is underspecified for this feature, and so is compatible with both [+passive] and [−passive] endings. This is said to account for the distribution of forms in most of the paradigm (Kiparsky's examples are drawn from sections labelled A, B and D in (3). Kiparsky's account is summarized in (7):

(7) A. There are both [−passive] endings and [+passive] endings, yielding the opposition between active and passive forms.

B. There are [−passive] endings, which yield active forms. The periphrastic passive forms, by contrast, are underlyingly unmarked for [±passive], being associated both with underspecified endings and with an underspecified auxiliary. Thus, they could, in principle, function both in active and passive contexts. However, by the principle of ECONOMY (i.e., a morphologically simple form prevails over morphologically complex form), the synthetic forms with the [−passive] endings block the periphrastic forms in active contexts, so the underspecified periphrastic forms are restricted to passive contexts.

C. The endings are unmarked for [±passive]; the whole form is thus underspecified, and functions as active by default.

Deponent verbs, by contrast, are lexically specified as [+passive], and so combine only with [+passive] or underspecified endings. This accounts for the absence of active forms in A and B on the one hand, and on the other hand for the acceptability of active forms in D, since these endings are underspecified.

However, this model fails to account for the whole paradigm. There are two problems.[5] First, the assumptions underlying the account of B and D are not compatible. In

[5]There are two more problems which, though they do not relate directly to the issue at hand, do cast doubt on the putative advantages of Kiparsky's (2005) analysis.

First, Kiparsky asserts that his model of feature values accounts for the gap in the perfect passive paradigm. That is, it *explains* why the perfect passive uses periphrastic forms (which fill in the gap), while an approach such as Sadler and Spencer's (2001) or Stump's (2002) merely stipulates that this portion of the paradigm uses periphrastic forms. On the assumption that the value 'perfect' is a composite of the values 'present' and 'past', he argues

> If the 'periphrastic' perfect is a semantically complex category, then it follows that the morphologically simple synthetic perfect is a portmanteau which folds those categories together. And that makes immediate sense of the gap in the morphological paradigm in (14) [the missing perfect passive forms; MB]: the missing synthetic perfect passive ending would express the three features PRESENT, PAST, and PASSIVE, which would have made it the only *triple* portmanteau in (14). As is well-known, morphological complexity is one of the factors that typically determine the distribution of gaps in paradigms. A synthetic perfect passive is obviously not impossible, but *if* there is a gap in the paradigm, this is a likely place for it to occur. (Kiparsky 2005:126)

the account of B, it is assumed that an underspecified form can range over active and passive functions; the fact that the underspecified forms are restricted to the passive is due to blocking. But the forms in D are also construed as underspecified, and so should likewise range across both active and passive contexts. Instead, they are restricted to active contexts.

Second, nothing in this model explains the behaviour of C, the passive participle. This is the same form which serves as a component of the periphrastic forms in B, and so must be construed as underspecified for [±passive]. But it is restricted to passive contexts, just as the forms in B are. While in the case of B, this restriction was attributed to blocking by the synthetic forms, this does not hold for C, as there is no corresponding [+passive] form to block it. This second point is especially problematic for Kiparsky's model. The only readily available way to describe this behaviour is by some sort of rule, as described above: normally it is passive, but in deponent verbs it is active.[6] If a rule-based account is permitted in one portion of the paradigm, there is no obvious reason to exclude it from the rest of the paradigm.

It seems, then, that a rule-based approach is preferable. However, the perfect participle causes some difficulties here too. Both Sadler and Spencer (2001) and Stump (2002) propose that rules take the normal active paradigm and link its cells to the passive paradigm. But since there is no perfect participle form in the active paradigm, we have to assume that there is a cell (more precisely, a group of cells) with the value PERFECT ACTIVE PARTICIPLE in the syntactic paradigm of all verbs. But this cell has a formal realization only in the case of deponent verbs. That is, the majority of Latin verbs are defective for this value.

While this is a possible analysis, it is unfortunate that it depends on the accidental intersection of two deviations, defectiveness and deponency. That is, on the one hand, there are verbs which are defective for the perfect active participle. On the other hand, there are deponent verbs which, coincidentally, happen *not* to be defective for this value. But intuitively, the two facts are connected: the perfect active participle is itself a result of

---

This argument appears to be based on the assumption that the pefect active only expresses two values, presumably PRESENT and PAST, with the ACTIVE unspecified (as indeed Kiparsky assumes for the forms in (D). But this does not conform to Kiparsky's own analysis of the perfect active forms, whose endings are overtly specified as [−passive]; indeed, the fact that they are marked for the feature [±passive] is crucial to the blocking analysis. Therefore, the active forms must be triple portmanteau as well, and thus no less complex morphologically than passive forms would be, which obviates the claim that the location of the gap has been accounted for.

The second problem is connected with the existence of a few passiva tantum verbs, i.e. defectives that are only passive (Flobert 1975:409ff.), such as *plector* 'be punished', *c* $\bar{o}$*nfl*$\bar{\imath}$*ctor* 'be afflicted'. Morphologically, these are distinguished from deponents by their lack of active forms (the 'D' forms in (3)). Intuitively, it seems that what one wants to say about such verbs is that they are inherently passive. But if having [+passive] as part of a verb's lexical specification precludes its functioning as a passive, how are passiva tantum to be represented?

[6] Alternatively, one might maintain the Lexical Morphology account by dispensing with the blocking analysis of B, i.e. by assuming that the perfect participle form contained some affix marked [+passive]. In that case, the behaviour of B and C would be parallel to A, and blocking would not need to be invoked. The problem here is the existence of the supine in D, which is active only, but which is formally identical to the perfect participle (specifically, the neuter singular). This entails the assumption of two homophonous affixes, a [+passive] perfect participle affix and an underspecified or [−passive] supine affix. While this seems a possible analysis, the use of such covert distinctions would appear to belie the point of Kiparsky's model, namely that function is transparently derivable from form.

deponency. That is, deponent verbs are characterized by the conversion of passive values to active ones, even where the corresponding active value does not otherwise exist. This intuition can be incorporated into the rule by reversing its directionality. For example, Stump's rule in (6) states that the active cells of the syntactic paradigm are exceptionally linked to the passive cells of the morphological paradigm. By reversing this, we instead say that the passive cells of the morphological paradigm are exceptionally linked to the active cells of the syntactic paradigm. In Stump's formalism this might be modelled as in (8).

(8) Revised rule of paradigm linkage
Deponent: [If L:[Primary root = R]
and SPF_{Latin}(<L, σ >) = <W, σ >,
then MPF_{Latin}(<R, σ/passive>) = <W,σ >]

Seen in these terms, the *active* value of the perfect participle is a by-product of the rule of paradigm linkage responsible for deponent verbs, and need not be assumed as underlyingly present in normal verbs.

1.4 Semi-deponents in Latin

Latin semi-deponents introduce a further complication into the analysis. The semi-deponents are a small class which is deponent only in the perfect forms, as shown in (9):

(9) Semi-deponent 'dare'

	active form	passive form
PRES IND	aud-et	
IMPERF IND	aud-ēbat	
FUT IND	aud-ēbit	
PRES SUBJ	aud-eat	
IMPERF SUBJ	aud-ēret	
IMP PRES	aud-ē	
IMP FUT	aud-ētō	
INF PRES	aud-ēre	
PERF IND		aus-us est
PLUPERF IND		aus-us erat
FUT PERF		aus-us erit
PERF SUBJ		aus-us sit
PLUPERF SUBJ		aus-us esset
INF PERF		aus-us esse
PART PERF	———-	aus-us
SUPINE	aus-um	———-
INF FUT	aus-ūr-us esse	———-
PART FUT	aus-ūr-us	———-
PART PRES	aud-ēn-s	———-
GERUND	aud-end-ī	———-

In Sadler and Spencer's (2001) and Stump's (2002) models, semi-deponency is a particular instance of deponency, restricted to extensions of the value 'perfect'. For example, Stump (2002) offers the rule in (10):

(10) **Semideponent**: [If L:[Primary root = R],
σ is an extension of perfective, and
MPF_{Latin}(<R, σ/{passive}>) = <W, σ/{passive}>, then
SPF_{Latin}(<L, σ >) = <W, σ>]

The crucial point here is that the domain of deponency within the lexeme is expressed in terms of a morphosyntactic value. But things are different if we think in terms of the revised model offered above in (8). Recall that, on this view, deponency is defined in terms of the morphological form, not the value. This suggests that semi-deponency should also be defined in terms of the form. In this case, that would mean delimiting the domain of semi-deponency in terms of the stem rather than the morphosyntactic value, i.e. 'passive forms based on the supine stem are linked to cells in the active syntactic paradigm', provisionally represented in (11). (Note that this rule has no effect on the supine, which is linked to active cells in any case.)

(11) Revised rule for semi-deponents (taking liberties with Stump's notation)
Semideponent: [If L: [Supine stem = X]
and SPF_{Latin}(<L,σ >) = <W, σ >,
then MPF_{Latin}(<X, σ/passive>) = <W, σ >]

As far as the Latin data go, the difference between (10) and (11) is immaterial: the domain of semi-deponency coincides both with a morphosyntactic value and a stem. But they make different predictions. The rule in (10) suggests we might find deponency which affects only certain values, but does not correspond to a distinct stem, while the rule in (11) suggests we might find deponency which affects a definable stem that does not correspond to a morphosyntactic value. In the following section I look at some examples from other languages that are similar to Latin semi-deponents, in order to see if either of these predictions is born out. None of them involve quite the same morphosyntactic values as in Latin, but we can think of deponency in more general terms, as a kind of morphological mismatch between the expected and the actual grammatical properties of a particular morphological form. That is, the form is one which would typically be interpreted as having the property 'X, not Y', but in this particular instance has the property 'Y, not X'.

1.5 Case studies

1.5.1 Takelma

The following description is based on Sapir (1922). The morphological opposition which concerns us in Takelma involves the distinction between intransitive and transitive verbs.

Morphologically, the difference between these classes is reflected in their distinct person-number endings; the most salient differences are shown in (12) and (13).[7] Classes I and II are intransitive (12), whereby class I verbs usually describe actions, and class II verbs (typically derived from transitives) describe states or processes. Class III comprises transitives (13).

(12) Intransitive forms (Sapir 1922:161-166)

	class I 'run'		class II 'stop'	
	aorist	future	aorist	future
1SG	yowò-t^{h}eʔ	yù-t^{h}e:	hanàʔs-deʔ	hànʔs-de:
1PL	yowoy-ik^h	yu-gàm	hanàʔs-ik^h	hànʔs-igam
2SG	yowó-t^h	yu-dàʔ	hanàʔs-dam	hànʔs-daʔ
2PL	yowó-t^{h}p^h	yù-t^{h}baʔ	hanàʔs-dap^h	hànʔs-dabaʔ
1SG	yowò-ʔ	yù-ʔt^h	hanàʔs	hànʔs-da:

Note: t^h~d and k^h~g represent allophonic variation (fortis~lenis).

(13) Transitive forms (Sapir 1922:170-171)

	class III 'kill'	
	aorist	future
1SG>3	t'omom-àʔn	do:m-àn
1SG>3	t'omom-anàk	do:m-anagàm
2SG>3	t'omom-át^h	do:m-adàʔ
2PL>3	t'omom-át^{h}p^h	do:m-àt^{h}baʔ
3>3	t'omom	do:m-ànk^h

Note: only forms with a third person object are shown; forms with a first or second person object are not relevant for the present discussion.

Note also that there is a stem alternation between the aorist stem (used for the aorist) and the basic stem, used elsewhere (future, potential, inferential and imperative). Stem formation is quite diverse: Sapir lists 16 types of aorist-basic stem relationships (p. 96). Typically the aorist stem is an enlargement of the basic stem.

Deponency in Takelma involves the use of transitive person-number endings (with an implied third person object) by intransitive verbs. One type involves what we may call full deponency, i.e. the lexeme is deponent in all its forms, and so falls beyond the range of what concerns us at the moment, though it makes sense to consider it first: Sapir observes that the verb 'think' is intransitive, but takes class III endings, e.g. *gel-hewèha*ʔ*n* 'I think'; note that this contrasts with the genuinely transitive *gel-hewèhiwi*ʔ*n* 'I think of him' (p. 179).[8]

[7] The remaining paradigms are those for the inferential and present imperative, which are the same for all verb classes, and the future imperative, where class III patterns with I or II, depending in the person of the object. The potential is formed from the basic stem + aorist endings. Presumably, the various phenomena which affect the aorist do not affect the potential, bur Sapir is not explicit on this point.

[8] Sapir mentions two other verbs which he says belong to this type (p. 183, n. 1). These are in fact suppletive stems, discussed below in (17).

Otherwise, Takelma exhibits a wide variety of semi-deponent types. First, there is a set of intransitive verbs which extend the stem with *-n-* in the first person forms, and these extended forms take class III endings; elsewhere, intransitive (class II) endings are found:

(14) ‘work’ (Sapir 1922:182)

	aorist	future
1SG	hegwèhak^{hw}-n-aʔn	he^{e}gwàk^{hw}-n-an
1PL	hegwèhak^{hw}-n-anàk	he^{e}gwàk^{hw}-n-anagam
2SG	hegwèhak^{hw}-dam	he^{e}gwàk^{hw}-daʔ
2PL	hegwèhak^{hw}-dap^{h}	he^{e}gwàk^{hw}-dabaʔ
3	hegwèhak^{hw}	?

Note: no form is attested for the third person future of this verb.

Second, there are verbs which behave like those in (14), but only in the aorist:

(15) ‘be lean in one’s rump’, 1SG (Sapir 1922:182)

aorist	future
di:-k‘alàs-n-aʔn	di:-k‘àlsi-de:

Some verbs waver between this and normal intransitive endings (e.g. the alternative aorist form *di:-k‘alàs-de*ʔ ‘I am lean in my rump’).

Third, there are verbs which have class III endings throughout the aorist, but take intransitive endings elsewhere:

(16) ‘listen’, 1SG (Sapir 1922:183)

aorist	future
da:-sgek‘iy-àʔn	da:-sgèk‘i-t^{h}e:

Fourth, a few intransitive verbs have a distinct plural stem, where the singular stem takes intransitive endings, and the plural stem takes class III endings:

(17) ‘come to a stand’, aorist (Sapir 1922:95-96)

1SG aorist	1PL aorist
ba:-**sa:sàs**-deʔ	ba:-sal-**xòxigi**-nak^{h}

This stem alternation is optional; where it does not occur, normal intransitive endings are used throughout:

(18) ‘be seated’ (Sapir 1922:95-96)

1SG	1PL
šuʔ**wili:**-t^{h}eʔ	al-**xali:ya**-nàk^{h} *or* **šu**ʔ**wili:p**-ik^{h}

There is also a possible fifth type which is not lexically restricted, but rather occurs with any verbal lexeme. This involves an additional future tense form, namely a periphrastic future involving an auxiliary which is inflected as a transitive (viz. with class III endings). The periphrastic future is functionally distinct from the synthetic future, having a less rigid tone, and slightly greater intentive force, so we can provisionally give it a distinct name, 'future 2'. It is formed with the aorist stem of *gulug*^{w}- 'intend, desire', which takes transitive subject suffixes, plus the main verb stem, which takes the object suffixes, if any:

(19) 'I shall kill him' (Sapir 1922:185)

synthetic ('future 1')	periphrastic ('future 2')
do:m-àn	do:m gulugw-àʔn

Significantly, this construction takes class III endings with all verbs, including intransitives. Unfortunately, none of Sapir's examples illustrate this unambiguously, though he does give examples (pp. 185-186) of 'passives' (≈ impersonals) formed from decidedly non-agentive intransitives, e.g. *wè:giau gulugw-àn* 'it will be shined (= it was going to be daylight)', where *-an* is the passive suffix. These are significant because the passive is typically formed only from transitives; passive intransitives occur only in the periphrastic future, i.e. where intransitives have transitive morphology.

In summary, Takelma exhibits a number of different types of semi-deponency, and in every case the domain of deponency corresponds both to a morphosyntactic value and to a distinct stem, as outlined in (20).

(20)

type	value	stem
1	first person	stem with -n extension
2	first person aorist	aorist stem with -n extension
3	aorist	aorist stem
4	plural	plural stem
5	future 2	auxiliary *gulug*^{w}-

Thus, the situation is as ambiguous as in Latin: it could be either the morphosyntactic value or the stem which licenses deponency.

1.5.2 Chamorro

In Chamorro, what concerns us is a mismatch involving the morphological distinction between transitive verbs with a definite object and those with an indefinite object in actor voice construction. The morphological differences between actor voice forms are shown in (21) based on Topping (1973). Intransitive verbs fall into two lexically-specified classes, here provisionally called '1' and '2'. Class 1 verbs take *-um-* infixation with singular subjects, while class 2 verbs take the prefix *ma-* throughout their paradigm. Transitive verbs distinguish between definite and indefinite object forms; in turn, there are two sets of definite object forms, one for main clauses and one for subordinate clauses.[9]

[9]The distinction between main and subordinate clause follows Donohue and Maclachlan (1999). Topping (1973) instead distinguishes between neutral focus (Donohue and Maclachlan's (1999) actor voice

(21) Chamorro actor focus affixes

		intransitive		transitive	
		class 1	class 2	indefinite object	definite object
main clause	SG SUBJ	⟨um⟩	ma-	man-	ERG pronominal-
	PL SUBJ	man-	man-ma-	man-man-	ERG pronominal-
subordinate clause	SG SUBJ	⟨um⟩	ma-	man-	⟨um⟩
	PL SUBJ	man-	man-ma-	man-man-	⟨um⟩

Some illustrations of these oppositions are given below:

(22) Intransitive verb, class 1

a. *singular subject*
G⟨um⟩upu yo'.
flew⟨ACTOR⟩ I.ABS
'I flew.'

b. *plural subject*
Mang-gupu siha.
INDEF.OBJ-flew they.ABS
'They flew.' (Topping 1973: 83)

(23) Intransitive verb, class 2

a. *singular subject*
Ma-makmata yo'.
INTR-woke.up I.ABS
'I woke up.'

b. *plural subject*
Man-ma-makmata siha.
PL-INTR-woke.up they.ABS
'They woke up.' (Topping 1973: 83)

(24) Transitive verb, indefinite object

Man-li'e' yo' guma'
INDEF.OBJ-saw I.ABS house

'I saw a house.' (Topping 1973: 233)

(25) Transitive verb, definite object, main clause

Hu-li'e' i palao'an
1SG.ERG-saw the woman

---

in main clauses) and actor focus (Donohue and Maclachlan's (1999) actor voice in subordinate clauses). Donohue and Maclachlan (1999) discuss two subordinate clause contexts:

- infinitive complements:
  Hu-tanga l⟨um⟩i'e' Hawaii.
  1SG.ERG-desire see⟨ACTOR⟩ H.
  'I desire to see Hawaii.'
  (Cited from Topping 1973, 94)
- Wh-questions

  singular subject
  Ha- fahan i palao'an i chotda.
  3SG.ERG- bought the woman the banana
  'The woman bought the banana.'

  plural subject
  Hayi f⟨um⟩ahan i chotda.
  who bought⟨ACTOR⟩ the banana
  'Who bought the banana?'

  (Cited from Chung 1994)

They argue that verbs in Wh-questions can be construed as subordinate if Wh-questions are treated as clefts (with zero copula; thus 'Who is it who bought the banana?'). The third type, as represented by sentences such as (26) below, is not analyzed by them. However, given the paraphrase that Topping suggests for such sentences ('I am the one who saw the woman.'), the cleft analysis would presumably obtain for them too.

'I saw the woman.' (Topping 1973: 243)

(26) Transitive verb, definite object, subordinate clause

Guahu l⟨um⟩i'e' i palao'an
I.EMPHATIC saw⟨ACTOR⟩ the woman

'I saw the woman.'
'I am the one who saw the woman.' (Topping 1973: 243)

The main clause indefinite object form of the transitive verb *kanno'* 'eat' shows a deviation from this pattern. Instead of being formed by prefixation of *man-* (SG) or *man-man-* (PL), the suppletive root *chocho* is used instead. This behaves just as an intransitive verb, taking the infix ⟨um⟩ in the sungular and *man-* in the plural (yielding *mañocho*). This is apparently the only lexeme that behaves this way (Topping 1973:241).

(27) 'eat', definite object (normal behaviour)
Hu- kanno' i mansana
1SG.ERG ate the apple
'I ate the apple.'

(28) 'eat', indefinite object, main clause (deponent behaviour)
Ch⟨um⟩ocho yo' mansana
ate⟨ACTOR⟩ I apple
'I ate the apple.'

Thus, we can say that the verb 'eat' is deponent in that it is a transitive which displays the morphological behaviour of an intransitive (singular *-um-* infixation), and semi-deponent in that this behaviour is restricted to context where the object is indefinite. However, this coincides with the domain of the suppletive stem *chocho*. Again, the data are ambiguous. What is it that exactly licenses this behaviour: is it the value 'indefinite object' or is it the suppletive stem *chocho*?

1.5.3 Nimboran

The following discussion is based on Anceaux (1965). In Nimboran, what concerns us is the object marking on the verb. Person-number marking distinguishes between masculine object, plural object, and unspecified object; the last serves for all other object types, as well as for intransitives. This is illustrated in (29), using present tense forms. What will interest us are the masculine object forms, so the masculine object suffix (*-ra-* ~ *-re-*) has been isolated in the paradigm. Note also that stems in Nimboran potentially display alternations according to number (here *ŋgeduo-* SINGULAR/MINIMAL, *ŋgedou-* DUAL/AUGMENTED, *ŋgedoi-* PLURAL).

(29) ‘draw’, present tense

	unspecified object	masculine object	plural object
1SG	ŋgeduotu	ŋgeduo-ra-tu	ŋgedoudatu
2SG	ŋgeduote	ŋgeduo-ra-te	ŋgedoudate
3N SG	ŋgeduotum	ŋgeduo-ra-tum	ŋgedoudatum
3M SG	ŋgeduotam	ŋgeduo-ra-tam	ŋgedoudatam
1INCL MIN	ŋgeduomantam	ŋgeduo-re-mantam	ŋgedoudemantam
1DU	ŋgedouketu	ŋgedouk-ra-tu	ŋgedoidiatu
2DU/PL	ŋgedouketé	ŋgedouk-ra-te	ŋgedoidiate
3N DU	ŋgedouketum	ŋgedouk-ra-tum	ŋgedoidiatum
3M DU	ŋgedouketam	ŋgedouk-ra-tam	ŋgedoidiatam
1INCL	ŋgedouketam	ŋgedouk-ra-tam	ŋgedoidiatam
1PL	ŋgedoitiu	ŋgedoi-ra-tu	= 1du
3PL	ŋgedoitiam	ŋgedoi-ra-tam	= 3du

Deponency consists in the fact that a number of verbs use the masculine object form where the unspecified object form would be expected (Anceaux 1965:127-129), as in *krendiya-ra-tum* ‘it becomes entangled’ (p. 127).[10]

[10]Anceaux (1965) explicitly treats these verbs as identical to masculine object forms. Inkelas (1993:570, 574-576) implicitly treats this as an instance of accidental homophony between the masculine object marker and a particle (particles are, roughly speaking, semantically empty derivational elements — Anceaux simply refers to them as *elements* — that many verb roots select for). However, there is no obvious reason to reject Anceaux’s (1965) original equation of the two. Two points argue in its favour:

(a) Both display the same morphological alternation, *-re-* in the penultimate position, *-ra-* elsewhere, and it seems to cause a following future marker *-d-* to be realized as *-r-*, e.g. *ŋgedúo-**d**-u* ‘I will draw’ versus *ŋgedúo-rá-**r**-u* ‘I will draw him’ (Anceaux 1965:186, 202). Inkelas proposed that both share the underlying form *-rar-*, and that the variant forms are due to phonological rules (pp. 570-571). The catalyst for these rules is the final *-r*, which is never realized overtly, but

(i) causes deletion of a preceding *-a-* in prepenultimate syllables, and is subsequently deleted (through degemination?), with epenthetic *-e-* then inserted, and

(ii) assimilates the following *-d-* of the future marker, followed by degemination. If a purely phonological explanation for the variant forms is accepted, this makes it more plausible that there are two accidentally homophonous affixes. However, these rules are phonologically aberrant, and in part overlap with already established morphlogical alternations. In the case of ‘i’, while *e*-epenthesis is generally attested in Nimboran, there is no direct evidence for the other rules (resulting in pre-prenultimate *ar*-deletion). In the case of ‘ii’, there are a number of other affixes which induce the alternation *-d-* ∼ *-r-* in the future, and Inkelas assumes that this is a morphological process, not a phonological one (pp. 573-574). It seems just as plausible then to attribute the range of alternations to purely morphological rules.

(b) There are typological parallels from other languages that have intransitives with aberrant object-marking morphology, e.g. Kiowa (Watkins 1984:145), Amele (Roberts 1987:281-284), Basque (Hualde and de Urbina 2003:240-241), Mawng (Singer 2003) and Tiwi (Lee 1987:173).

Most of the examples are intransitive verbs, and presumably are unable to take an object ('go', 'die', 'tremble', 'swim', 'hold a pig-feast', 'laugh' etc.).[11]

In a number of verbs this behaviour is restricted to a portion of the total paradigm, i.e. they are semi-deponent. As Anceaux does not give complete paradigms, the following is based on his (quite detailed) description of the combinatory possibilities of the various morphological components; complete forms that *are* given by Anceaux are shown in boldface in the discussion below. Note also that 'unspecified object' and 'masculine object' are simply morphological labels in the paradigms below: all the forms are functionally unspecified object forms.

The verb 'come' seems to have spurious masculine object marking in the singular and dual (p. 128):[12]

(30) 'come', future tense

	unspecified object	masculine object
1SG		**puraru**
2SG		purare
3N SG		purarum
3M SG		puraram
1INCL MIN		puremandam
1DU		puŋkraru
2DU/PL	**puŋkedé** (PL)	**puŋkrare** (DU)
3N DU		puŋkrarum
3M DU		puŋkraram
1INCL AUG		puŋkraram
1PL	puiŋdiu	
3PL	puiŋdiam	

This example has the same ambiguity as those in the previous sections: the domain of deponency coincides with that of a value (plural) and that of a stem (the plural stem *puiŋ-*). (Note that, as a consequence of this distribution, second person has distinct dual and plural forms; typically the two are syncretic in Nimboran.)

The root *kri-* 'build' shows free variation between unusual and expected forms when there is no directional-locational suffix[13] (e.g. ***krik-ra-ru*** or *krikedu* 'build.1DU.FUT'); otherwise the expected unspecified object forms are found, e.g. *krike-ba-ru* 'build above.1DU.FUTURE', *krike-ŋa-ru* 'build below.1DU.FUTURE', *krike-sa-ru* 'build there.1DU.FUTURE', *krike-na-ru* 'build far away.1DU.FUTURE' etc. (p. 128). Here, the domain of deponency corresponds to a value (Anceaux translates the forms lacking a directional-locational suffix as 'here') and to a stem, in as much as an unsuffixed stem is distinct from a suffixed stem.

[11] On the other hand, some appear to be transitive ('tell', 'build'), whereby presumably the masculine object form is found even where the object is feminine or neuter in the singular or dual (i.e. the contexts where the unspecified object form is normally used). It is not clear from Anceaux's description what would happen in the case of a plural object.

[12] The *-r-* does not alternate with *-d-* in the future, which is otherwise typical of the masculine object marker, so it may well be that the *-ra-* in this verb should not be equated with the masculine object marker.

[13] 'Directional-locational' is Inkelas's term (1993:571); **?** uses the term 'position class'.

Two verbs take spurious masculine object marking in the iterative only (p. 129). The iterative has a distinct stem form, in as much as it is formed with an affix which Inkelas (1993:572-273) models as *-ŋkat-*, which is suffixed to the verb root or to the verb root + directional-locational suffix (if present). Curiously, both verbs have suppletive roots, one in the iterative and one elsewhere, and one of those roots is ∅. But in the case of one verb ('hear'), it is the ∅ root which is normal, and in the case of the other ('laugh') the ∅ root is deponent (31):

(31) distribution of suppletive ∅ roots

	unspecified object	masculine object form
'hear'	∅	ty-
'laugh'	kia-	∅

The root *ty-* shows further peculiarities, discussed below.

Some ∅ roots in fact form a class of their own with respect to semi-deponency (Nimboran is notorious for having, according to Anceaux (1965:184), twelve verbs with a ∅ root). Three ∅ roots, 'dream', 'bring' and the iterative of 'laugh' (discussed above) are sensitive to the presence of certain sets of the 16 directional-locational suffixes (p. 129); for convenience, following Anceaux (1965) and Inkelas (1993), we can refer to these suffixes simply by the labels '1' to '16'.[14] Deponent behaviour of ∅ roots is associated with suffixes 1-5 or with 6-16: with 'dream' and 'laugh' it is the forms with suffixes 1-5 which have spurious masculine object marking, while with 'bring' it is the forms with suffixes 6-16 that have spurious object marking (32).

(32) classification within ∅ roots

	unspecified object	masculine object form
'dream'	with suffixes 6-16	with suffixes 1-5
'laugh.ITERATIVE'		
'bring'	with suffixes 1-5	with suffixes 6-16

[14]Inkelas (1993) gives the suffixes as:

1	*none*	(default)
2	-ba-	'above'
3	-ŋa-	'below'
4	-sa-	'there'
5	-na-	'far away'
6	-ba-	'from here to above'
7	-se-	'from here to there'
8	-sa-	'from here to below'
9	-na-	'from here to far away'
10	-kaN-	'from above/far away to here'
11	-baN-	'from below to here'
12	-saN-	'from there to here'
13	-bena-	'from below/there/far away to above'
14	-sena-	'from below to there'
15	-kana-	'from above/below/there to far away'
16	-sana-	'from there/above/far away to below/there'

Suffixes 6, 8-10, 15 and 16 are also associated with vowel ablaut.

The iterative *ty-* root of 'hear' (discussed above (32)), does not in fact display spurious masculine object marking in all its forms. Along with a few other verbs, it takes an additional element which Anceaux gives as *-maŋ-* in the 1INCL augmented and the first plural (pp. 158-160). These forms lack spurious masculine object marking in 'hear' (33).

(33) 'hear', future tense

	unspecified object form	masculine object form
1SG		ty-re-katu
2SG		ty-re-kate
3N SG		ty-re-katum
3M SG		ty-re-katam
1INCL MIN		ty-re-maŋ-katam
1DU		tyk-re-katu
2DU/PL		tyk-re-kate
3N DU		tyk-re-katum
3M DU		tyk-re-katam
1INCL AUG	tyke-maŋ-katam	
1PL	tyi-maŋ-katu	
3PL		tyi-re-katam

Here the restriction on the range of deponency can be seen as corresponding to a value (first plural/augmented). Whether it corresponds to a stem is less clear. On the one hand there is a clear interaction between the additional element *-maŋ-* and spurious masculine object marking, the two being mutually exclusive.[15] However, as far as can be gleaned from Anceaux's material, *-maŋ-* is identical to the suffix found in the 1INCL minimal (which Inkelas (1993) renders as *-maN-* (e.g. p. 565), *N* being an underspecified nasal), which does not prevent masculine object marking.[16]

The final example is one where the domain of deponency can only be described in terms of the morphosyntactic values. In the root *iii-* 'divide' the deponent forms coexist alongside the expected ones in the dual, while in the first person plural only the expected forms are found (p. 128):

[15]But note that the ∅ roots 'dream' and 'laugh', discussed in (31) and (32), also take *-maŋ-* in the same contexts, and Anceaux's description (p. 129) suggests they retain spurious masculine object marking in all their forms (he does not cite relevant examples, though).

[16]Note further that an element which Anceaux identifies as *-maŋ-* is also used with a small number of verb roots to distinguish the otherwise syncretic 2PL from 2DU (pp. 165-166).

(34) ‘divide’, future tense

	unspecified object form	masculine object form
1SG		**iii-ra-ru**
2SG		iii-ra-re
3N SG		iii-ra-rum
3M SG		iii-ra-ram
1INCL MIN		iii-re-mandam
1DU	iii-ke-du	iii-k-ra-ru
2DU/PL	iii-ke-dé	iii-k-ra-re
3N DU	iii-ke-dum	iii-k-ra-rum
3M DU	iii-ke-dam	iii-k-ra-ram
1INCL AUG	iii-ke-dam	iii-k-ra-ram
1PL	iii-diu	
3PL		iii-ra-ram

This unusual behaviour can only be described in terms of the values involved, and does not correspond to any independently identifiable form or class of forms.

In sum, the role that can be attributed to stems in defining semi-deponency in Nimboran is variable. The examples described above range from (30), where deponency coincides with the plural stem of ‘come’, to (34), where the form of the stem clearly plays no role.

1.6 Conclusion

What do the examples in Section 1.5 tell us about thep location of deponency? Whatever their inherent interest, they are mostly as uninformative as the semi-deponents of Latin, with deponency restricted to particular stems, which themselves correspond to discrete morphosyntactic values. Of course, this is hardly surprising: while quirky stem alternations are certainly possible (as with Aronoff’s ‘morphomes’), they do tend to be realizations of discrete morphosyntactic values (e.g. for number, tense). Statistically, then, we should expect most examples to be ambiguous. The one apparently unambiguous example is Nimboran ‘divide’ (34), which points to semi-deponency defined in terms of the morphosyntactic values and *not* the stem. Is this sufficient to resolve the question? Probably not, being only a single lexical item from a single language. Clearly, more examples are needed.

Bibliography

ANCEAUX, J. C. 1965. *The Nimboran Language*. ’S-Gravenhage: Martinus Nijhoff.

BENNET, CHARLES E. 1942. *New Latin Grammar*. Boston: Allyn and Bacon.

BOBALJIK, JONATHAN D., and PHIL BRANIGAN. forthcoming. Eccentric agreement and multiple case checking. In *Ergativity*, ed. by Alana Johns, Diane Massam, and Juvenal Ndayiragije. Dordrecht: Springer.

Börjars, Kersty, Nigel Vincent, and Carol Chapman. 1996. Paradigms, periphrases and pronominal inflection: a feature-based account. In *Yearbook of Morphology 1996*, ed. by Geert Booij and Jaap van Marle, 155–180. Dordrecht: Kluwer.

Chung, Sandra. 1994. Wh-Agreement and 'referentiality' in Chamorro. *Linguistic Inquiry* 25.1–44.

Corbett, Greville G., 1999. Defectiveness, syncretism, suppletion, 'deponency': four dimensions for a typology of inflectional systems. Guest lecture at The Second Mediterranean Meeting of Morphology, September 1999, Lija, Malta.

Donohue, Mark, and Anna Maclachlan. 1999. What agreement in Chamorro? In *Proceedings of the Austronesian Formal Linguistics Association VI (Toronto Working Papers in Linguistics 16.2)*, ed. by Carolyn Smallwood and Catherine Kitto, 121–132, Toronto. Department of Linguistics, University of Toronto.

Embick, David. 2000. Features, syntax and categories in the latin perfect. *Linguistic Inquiry* 31.185–230.

Flobert, Pierre. 1975. *Les verbes déponents latins des origines à Charlemagne*. Paris: Société d'Édition "Les Belles Lettres".

Hualde, José Ignacio, and Jon Ortiz de Urbina. 2003. *A grammar of Basque*. Berlin: Mouton de Gruyter.

Inkelas, Sharon. 1993. Nimboran position class morphology. *Natural Language and Linguistic Theory* 11.559–624.

Kiparsky, Paul. 2005. Blocking and periphrasis in inflectional paradigms. In *Yearbook of Morphology 2004*, ed. by Geert Booij and Jaap van Marle, 113–136. Dordrecht: Kluwer.

Lee, Jennifer. 1987. Tiwi today: a study of language change in a contact situation. *Pacific Linguistics* C-96.

Roberts, John R. 1987. *Amele*. London, New York and Sydney: Croom Helm.

Sadler, Louisa, and Andrew Spencer. 2001. Syntax as an exponent of morphological features. In *Yearbook of Morphology 2000*, ed. by Geert Booij and Jaap van Marle, 71–96. Dordrecht: Kluwer.

Sapir, Edward. 1922. The Takelma language of southwestern Oregon. In *Handbook of American Indian Languages*, ed. by Franz Boas, volume 2, 1–296. Washington, DC: Government Printing Office.

Singer, Ruth, 2003. Restricted agreement (section of Confirmation report). Manuscript, University of Melbourne.

Stump, Gregory. 2002. Morphological and syntactic paradigms: Arguments for a theory of paradigm linkage. In *Yearbook of Morphology 2001*, ed. by Geert Booij and Jaap van Marle, 147–180. Dordrecht: Kluwer Academic Publishers.

Topping, Donald M. 1973. *Chamorro reference grammar*. Honolulu: The University Press at Hawaii.

Watkins, Laurel J. 1984. *A grammar of Kiowa*. Lincoln and London: University of Nebraska Press.

Matthew Baerman
Surrey Morphology Group
University of Surrey
Guildford, GU2 7XH, UK
m.baerman@surrey.ac.uk

2 On the Unity of 'Number' in Semantics and Morphology*

Daniel Harbour (Queen Mary, University of London)

2.1 Introduction

In an extended study of the relationship between morphologically complex agreement and semantically based noun classification (Harbour 2003a), I argued that the status quo in number theory—namely, that morphologists and semanticists concentrate on disjoint bodies of fact and develop correspondingly disjoint theories—is untenable: linguistics requires a unified morphosemantic theory of number. The current paper advances this case by observing that morphologists and semanticists have, despite divergent concerns, converged on a single discovery: the morphological notion of augmentation (Noyer 1992, Harbour 2003b) and the semantic notion of strict cumulativity (Krifka 1992) are near logical equivalents.

(1) a. A predicate, P, is **augmented**, Aug(P), if and only if
$$\exists x \exists y[P(x) \wedge P(y) \wedge x \sqsupset y]$$
(*i.e.*, it is satisfied by two individuals,[1] one containing the other).

b. A predicate, P, is **strictly cumulative**, Cum(P), if and only if
$$\forall x \forall y[[P(x) \wedge P(y)] \rightarrow P(x \sqcup y)] \wedge \exists x \exists y[P(x) \wedge P(y) \wedge x \neq y]$$
(*i.e.*, it is satisfied by the join of any individuals, minimally two, that satisfy it).

After outlining the quite disparate development and use of these concepts in morphology (section 2.2) and semantics (section 2.3), I demonstrate their near equivalence (section 2.4), as formulated in (2) and (3):

(2) a. A predicate, P, is **additive**, Add(P), if and only if
$$\forall x \forall y[[P(x) \wedge P(y)] \rightarrow P(x \sqcup y)]$$
(*i.e.*, it is satisfied by the join of any individuals that satisfy it).

b. A predicate, P, is **augmented***, Aug*(P), if and only if
$$\text{Add}(P) \wedge \text{Aug}(P)$$

*Thanks are due to Christian List, to David Adger, Susana Béjar and Philippe Schlenker, and to the organisers and audience of the York Essex Morphology Meeting.

[1] An individual is an atom or set of atoms, or, equivalently, a lattice point.

Essex Research Reports in Linguistics 47, 21–29.
Ryo Otoguro, Gergana Popova and Andrew Spencer (eds.).
Department of Language and Linguistics, University of Essex,
Colchester, United Kingdom.

(3) a. Augmentation entails additivity (for non-cardinality predicates, in morphologically relevant models).

b. Augmentation* entails strict cumulativity (in all models).

c. Strict cumulativity entails augmentation* (in all models).

Section 2.5 discusses the theoretical and practical significance of (3) for semantics and morphology, suggesting not only that we need a unified morphosemantic theory of number, but that some topics are immediately relevant to both semantic and morphological research. It also provides a new feature classification of agreement and pronominal categories, incorporating 'unit augmented' and 'paucal', and a typology of number systems, suggesting that traditional feature geometry is superfluous.

2.2 Augmentation

Augmentation originates in descriptions of pronominal and agreement systems found in languages of the Philippines and of Australia's Arnhem Land (Corbett 2000, whose exposition, pp. 166–169, is followed here). Its motivation lies in the rather odd view of such systems that results from use of the traditional descriptive categories 'singular', 'dual', 'plural', and so on. This was first noted by Thomas (1955) for Ilocano.

(4) **Table 1**
Ilocano pronominal forms (traditional categorization)

Person	Singular	Dual	Plural
1 inclusive		-ta	-tayo
1 exclusive	-ko	. . . .	-mi . . . .
2	-mo	. . . .	-yo . . . .
3	-na	. . . .	-da . . . .

Observe that there is only one specifically dual form. As this is for the first person inclusive 'you and I', the dual is to some extent 'forced' on the language—a singular inclusive is impossible. Yet, by adopting [±augmented], one can avoid positing this defective, semantically predictable dual:

(5) **Table 2**
Ilocano pronominal forms (revised categorization)

Person	[−augmented]	[+augmented]
1 inclusive	-ta	-tayo
1 exclusive	-ko	-mi
2	-mo	-yo
3	-na	-da

By way of illustration, consider the top two rows. For the first person inclusive, let us take P to be the predicate 'includes "I" and includes "you"'. Then, for *-tayo*, [1 inclusive +augmented], $\exists x \exists y[P(x) \wedge P(y) \wedge x \sqsupset y]$ means that the model has two individuals,

both containing 'I' and 'you', the one individual contained in the other. Minimally, then, the model includes {I, you, other} (as this contains an individual, {I, you}, that contains both 'I' and 'you'), though it may also contain {I, you, $other_1$, ..., $other_n$}, up to arbitrary *n*. This is the desired result. By contrast, for *-ta*, [1 inclusive −augmented], $\forall x \forall y[\neg P(x) \vee \neg P(y) \vee x \not\sqsupset y]$ means that there are no pairs containing both 'I' and 'you' that are in a containment relation; but, as {I, you} would be contained by any other individual satisfying P, {I, you} must be the only individual satisfying P. Again, this is the desired result.

For the first person exclusive, let us take P to be the predicate 'includes "I" and excludes "you"'. Then, for *-mi*, [1 exclusive +augmented], $\exists x \exists y[P(x) \wedge P(y) \wedge x \sqsupset y]$ means that the model has two individuals, both containing 'I' but not 'you', with one individual contained in the other. By the reasoning above, the model minimally includes {I, other}, the desired result. And for *-ko*, [1 exclusive −augmented], $\forall x \forall y[\neg P(x) \vee \neg P(y) \vee x \not\sqsupset y]$ means, by the reasoning above, that the model has a unique individual, {I}, satisfying P, again the desired result.

Observe that [±augmented] permits flexibility in cardinality. In Ilocano, [−augmented] sometimes entails cardinality 1, sometimes 2; [+augmented] sometimes entails cardinality 2 or more, sometimes 3 or more. Analogous to Ilocano, but more complex, is Rembarrnga (McKay 1978, 1979), which displays, in traditional terms, singular, dual, trial and plural. However, trial is restricted to first person inclusive (*cf.*, Ilocano's dual). Here, then [−augmented] can entail cardinality 1, 2, or 3. What is particularly elegant about this trial-free reanalysis is that the forms ending in *-bbarrah* occupy the same part of the 'paradigm' (as opposed to *ngakorrbbarrah* being trial and all other *-bbarrah* forms being dual):

(6) **Table 3**
Rembarrnga dative pronouns

Person	Minimal	Unit Augmented[2]	Augmented
1 inclusive	y~~u~~kk~~u~~	ngakorr**bbarrah**	ngakorr~~u~~
1 exclusive	ng~~u~~n~~u~~	yarr**bbarrah**	yarr~~u~~
2	k~~u~~	nakor**bbarrah**	nakorr~~u~~
3 masculine	naw~~u~~	barr**bbarrah**	barr~~u~~
3 feminine	ngad~~u~~	barr**bbarrah**	barr~~u~~

It is expressly with this flexibility of cardinality in mind that [±augmented] has been defined. Various versions have been offered: [±restricted] (Conklin 1962), [±others] (Matthews 1972), as well as Noyer's (*op. cit.*), which is the most robustly typologically tested to date. (1a) is just a notational variant of this, minus the condition that individual must be non-zero. Such a ban is crucial, as, without it, [−augmented], $\forall x \forall y[\neg P(x) \vee \neg P(y) \vee x \not\sqsupset y]$, is satisfied by the empty set (and one does not want first person inclusive non-augmented, say, to refer to non-entities). However, in another point of contact between the two theories of number, Krifka too must rule out zero elements, witness his $\neg \exists x \forall y[x \sqsubseteq y]$ postulate. As I am urging a unification of semantic and morphological treatments of number, it will suffice to stipulate the ban once for both.

[2]On the feature composition of 'unit augmented' ('minimal' plus one other), see section 2.5.

2.3 Cumulation

The term 'cumulative reference' originates with Quine and his treatment of the ontogenesis of reference: '*mass* terms like "water" ... have the semantical property of referring cumulatively: any sum of parts which are water is water' (1960, p. 91). It has come to play a significant role in semantics owing largely to Krifka's treatment of the interaction of nominal and verbal reference. For instance, Krifka (1992:33–36) shows that the telic~atelic distinction can be reduced to denial~assertion of strict cumulativity. Specifically, he defines the notion of having a set terminal point as the formal correlate of telicity and then shows that strict cumulativity entails atelicity, that is, non-existence of a set terminal point.

This formal result is important. It reduces telicity, a property of events only, to strict cumulativity, a property that applies equally to events and to objects. This permits a straightforward analysis of the dual fashion in which telicity can arise: either a predicate is inherently telic (e.g., *arrive*), or a predicate that is ordinarily atelic (e.g., *drink*) may become so if its object is non-cumulative (e.g., *a glass of wine*—non-cumulative because a glass of wine plus another glass of wine is no longer just a glass of wine).

(A minor difference should be noted between Krifka's definition of strict cumulativity (7) and that assumed here.

(7) $\forall x \forall y[[P(x) \wedge P(y)] \rightarrow P(x \sqcup y)] \wedge \neg \exists x[P(x) \wedge \forall y[P(y) \rightarrow x = y]]$

(7) is weaker than (1b), since only (7) is true of predicates that are true of no individuals: (1b) $\models \exists x \exists y[P(x) \wedge P(y) \wedge x \neq y]$; however (7) entails the same proposition only given the auxiliary assumption that the predicate is true of at least one individual: $\{(7), \exists x P(x)\} \models \exists x \exists y[P(x) \wedge P(y) \wedge x \neq y]$. This minor difference is irrelevant for current purposes.[3])

2.4 Equivalence

Clearly, the empirical concerns of the morphologists who devised the notion of augmentation were very different from those of the semanticists who devised the notion of cumulativity. Nonetheless, I now prove the statements in (3). Formally stated, they are:

(8) a. $Aug(P) \models_{\mathcal{M}} Add(P)$ (for any model, $\mathcal{M}$, relevant to morphology, and for any non-cardinality predicate, P)

b. $Aug^*(P) \models Cum(P)$

c. $Cum(P) \models Aug^*(P)$

To demonstrate (8a), I take non-cardinality predicates to be person and gender predicates. As the discussion of Ilocano illustrates, when morphologists are concerned with semantics (of pronominal or agreement categories), they are generally concerned with groups of people/things and whether they include the speaker, the hearer and/or others. If

[3]Krifka abbreviates the property in (7) as SCUM(P) and uses CUM(P) for (2a), which I have termed 'additivity' in order to avoid using 'strictly cumulative' and 'cumulative' in rapid alternation.

P is a predicate that denotes inclusion or exclusion of 'I' or 'you' from an individual of arbitrary size, then P obeys additivity, as the join of any two individuals of arbitrary size containing 'I', say, or excluding 'you' and 'I', is another individual of arbitrary size with the same property. Similarly, if two individuals consist entirely of feminine individuals or if they contain at least one masculine one, then their joins will also contain only feminine individuals, or at least one masculine one. We saw in section 2.2, that [+augmented] permits individuals of arbitrary size. So, in morphological models, $\mathcal{M}$, augmentation entails additivity, that is, Aug(P) $\models_{\mathcal{M}}$ Add(P), for any non-cardinality predicate, P.

To demonstrate (8b), observe that, if Aug*(P), then Aug(P), and so there are individuals, a and b, such that P(a) and P(b) and $a \sqsupset b$. From $a \sqsupset b$, it follows $a \neq b$. So, we can write $[P(a) \wedge P(b) \wedge a \neq b]$. By existential quantification, we have $Aug(P) \models \exists x \exists y[P(x) \wedge P(y) \wedge x \neq y]$. So, $Aug^*(P) \equiv Add(P) \wedge Aug(P) \models Add(P) \wedge \exists x \exists y[P(x) \wedge P(y) \wedge x \neq y] \equiv Cum(P)$.

To demonstrate (8c), observe that any model for Cum(P) will contain two individuals, a and b, such that P(a), P(b) and $P(a \sqcup b)$. Since $a \sqcup b \sqsupset a$, it follows that $[P(a \sqcup b) \wedge P(a) \wedge a \sqcup b \sqsupset a]$. So, by existential quantification, we have $Cum(P) \models \exists x \exists y[P(x) \wedge P(y) \wedge x \sqsupset y] \equiv Aug(P)$. As conjunctions entail conjuncts, $Cum(P) \models Add(P)$. So, $Cum(P) \models Add(P) \wedge Aug(P) \equiv Aug^*(P)$.

2.5 Ramifications

It is surely remarkable that morphologists concerned with agreement and pronoun inventories and semanticists concerned with the interaction of telicity with nominal and verbal reference should have converged on two such similar notions as augmentation and strict cumulativity. As augmentation* and cumulativity are logically equivalent, one can be dispensed with and the other adopted in morphology and semantics alike. Which?

I suggest that augmented* be adopted, as it induces a classification of pronominal and agreement categories in terms of its conjuncts, additivity and augmentation, that is superior to that induced by strict cumulativity. In particular, it illuminates the notions of 'unit augmented' in (6) and 'paucal'.

First, observe from the comments following (5) and from the proof of (8a) that the category augmented is [+additive] (in addition, of course, to [+augmented]). Combining [−additive] with [+augmented] yields, I suggest, a paucal. Paucals pick out groups with few members (Foley 1991:111, for instance, says the Yimas paucal generally ranges from 3 to 7); they are a plural-like category with an upper bound. Consequently, a paucal could identify {a, b, c, d} or {a, b, c}. As the former contains the latter, paucals are [+augmented]. However, they are [−additive], by the reasoning of the Sorites paradox: a few plus a few is not necessarily a few.

Now consider unit augmented forms. Second person unit augmented, say, identifies {you, other$_1$}, {you, other$_2$}, and so on. Clearly, none of these is contained in any other; so, unit augmented is actually [−augmented]. Moreover, unit augmented is non-additive, as {you, other$_1$} ⊔ {you, other$_2$} does not contain just 'you' and a unique other. Hence, unit augmented is [−additive −augmented]. We can summarize these results as:[4]

[4]I suggest that the fourth possibility, [+additive −augmented], is unacquirable. It would pick out, un-

(9) **Table 4**
Typology of agreement/pronoun categories

[±additive]	[±augmented]	Category
+	+	plural, augmented
−	+	paucal
−	−	unit augmented

The typology can be expanded to include the traditional singular and dual using the feature [±singular], which asserts that cardinality equals 1. Harbour (2003a:87–89) shows that singular is [+singular −augmented] and dual, [−singular −augmented]. Singular and dual cannot be inherently specified for [±additive]: singular is [−additive] for third person, [+additive] otherwise; dual is [+additive] for first person inclusive, [−additive] otherwise. Failure fully to crossclassify may make the feature inventory [±singular], [±additive], [±augmented] seem somewhat redundant—especially, also, as there are several ways to specify, for instance, plural: [−singular], [−singular +augmented], [−singular +additive +augmented], ... I suggest that this accurately reflects the partial overlap between number systems crosslinguistically. By activating different subsets of number features, we can characterize such overlapping systems as singular-dual-plural and singular-paucal-plural, say. I propose the following typology (examples from Corbett 2000).

(10) **Table 5**
Typology of number systems

Language	Categories	Features
Pirahã	no number distinctions	[ ]
Russian	singular, plural	[±sg]
Ilocano	minimal, augmented	[±aug]
Unacquirable[5]	non-additive, additive	[±add]
Upper Sorbian	singular, dual, plural	[±sg ±aug]
Bayso	singular, paucal, plural	[±sg ±add]
Rembarrnga	minimal, (unit) augmented	[±add ±aug]
Yimas	singular, dual, paucal, plural	[±sg ±add ±aug]

It follows from paucal's being a composite of features that no language can have paucal~non-paucal as its only number distinction: singular~non-singular must also be distinguished. Similarly, unit-augmented~non-unit-augmented cannot be a language's only number distinction. Corbett's study number systems suggests that these predictions are correct, rendering the tradition geometric structure, superimposed on number features (e.g., Noyer 1992, Harley and Ritter 2002), otiose.

der varying person features, {I, you}, {I}, {you} and would be undefined otherwise. This is equivalent to restriction of [±augmented] to first and second person, an instantiation of a general crosslinguistic pattern: languages frequently make a number distinction only for upper parts of the animacy hierarchy (Corbett 2000). If the generalization is part of UG, then I suggest that [+additive −augmented] would be 'misacquired' as an instance of it. So, its exclusion from (9) is justified.

[5]By (8a), all person/gender categories would be [+additive]; so, there would be no evidence that [±additive] alone is active.

To my knowledge, no feature classification has been previously given for paucal or unit augmented (beyond the unenlightening [±paucal]).[6] However, we lose this new insight, and the typologies it implies, if we adopt cumulative in place of augmented*. For, then $\exists x \exists y[P(x) \wedge P(y) \wedge x \neq y]$ replaces $\exists x \exists y[P(x) \wedge P(y) \wedge x \sqsupset y]$. Since both paucal and unit augmented satisfy the former condition, we lose any ready means of distinguishing them.

What follows from semanticists' adopting morphologists' augmented*? It would be a mistake, after the initial bout of search-and-replace subsides, simply to return to business as normal. The convergence of morphological and semantic research indicates that the fields are closer than their disparate subject matter suggests, close enough, in fact, to demand a unified theory of number, as argued more generally by Harbour (2003a). If this conclusion is correct, then it is likely that further points of contact or convergence will emerge and this will hold practical implications for morphologists and semanticists alike.

For morphologists, a practical implication is that, before one posits a new morphological feature, one should first search the semantic literature for kin concepts; many morphological concepts (paradigmatic dimensions) are likely not *sui generis*. This has just been illustrated for unit augmented and paucal, which are easily characterized once one has additivity and augmentation. (See Ojeda 1998 for a semantically rooted, morphologically insightful study of distributives and collectives.)

For semanticists, a practical implication is that paradigmatic distinctions are testing grounds for semantic concepts. This too was illustrated above. The paucal is especially interesting as its use is determined both by 'absolute size of the group being referred to' and by 'relative size, i.e. whether the group being referred to is contrasted with some larger group within which it is subsumed' (Crowley 1982:81, as cited by Corbett 2000:24). One and the same form is, then, ambiguous between *few* and *few of* readings.

A similar point of contact is provided by associative plurals. The Japanese morpheme *-tati*, for instance, creates 'a non-uniform plural whose extension can include entities that are not in the extension of the common noun to which *-tati* is attached' (Nakanishi and Tomioka 2004). Non-uniformity is illustrated by *tati*-modified proper names (ibid.):

(11) Taro-tati-wa moo kaetta
Taro-TATI-TOP already went home
'The group of people represented by Taro went home already'

The role of *Taro* in *Taro-tati* is strongly reminiscent of the role of first person in the first person plural: not multiple first persons, but multiple persons including the first. Similarly, *Taro-tati* does not mean multiple Taros, but multiple persons including Taro. Masculine gender is the same in some languages (Philippe Schlenker, p.c.) (e.g., Romance, Semitic): masculine plural agreement need not indicate multiple masculine things, but multiple things including a masculine one. This means that the interaction of personhood with plurality is not a quirk of person *per se* but is an instance of a more widespread

[6] Noyer (1992:198–199) attributes the difference between unit augmented and (normal) augmented to 'functional inference'. Though the current system more tightly constrains the meaning of unit augmentation, it does not derive that it is augmented by one: the 'dual augmented' satisfies the same feature specifications. I leave this issue open.

semantic phenomenon, of interest to morphologists and semanticists alike, one that cannot be satisfactorily treated by theory of number that is solely morphological or solely semantic.

Bibliography

Conklin, Harold. 1962. Lexicographic treatment of folk taxonomies. In *Problems in Lexicography*, ed. by F. W. Householder and S. Saporta, 119–141. Bloomington, Indiana: Indiana University Research Center in Anthropology, Folklore, and Linguistics.

Corbett, Greville. 2000. *Number*. Cambridge: Cambridge University Press.

Crowley, Terry. 1982. *The Paamese Language of Vanuatu*, volume 87 of *Pacific Linguistics*. Canberra: Australian National University.

Foley, William. 1991. *The Yimas Language of New Guinea*. Stanford, CA: CSLI Publications.

Harbour, Daniel. 2003a. *Elements of Number Theory*. Cambridge, MA: MITWPL. To appear (revised) as *Morphosemantic Number Theory*. Dordrecht: Springer. 2006.

——. 2003b. The Kiowa case for feature insertion. *Natural Language and Linguistic Theory* 21.543–578.

Harley, Heidi, and Elizabeth Ritter. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78.482–526.

Krifka, Manfred. 1992. Thematic relations as links between nominal reference and temporal constitution. In *Lexical Matters*, ed. by Ivan Sag and Anna Szabolcsi, 29–53. Stanford, CA: CSLI Publications.

Matthews, Peter. 1972. Huava verb morphology: Some comments from a non-tagmemic viewpoint. *International Journal of American Linguistics* 38.96–118.

McKay, Graham. 1978. Pronominal person and number categories in Rembarrnga and Djeebbana. *Oceanic Linguistics* 17.27–37.

——. 1979. Gender and the category *unit augmented*. *Oceanic Linguistics* 18.203–210.

Nakanishi, Kimiko, and Satoshi Tomioka. 2004. Japanese plurals are exceptional. *Journal of East Asian Linguistics* 13.141–179.

Noyer, Rolf. 1992. *Features, Positions and Affixes in Autonomous Morphological Structure*. Cambridge, MA: MITWPL.

Ojeda, Almerindo E. 1998. The semantics of collectives and distributives in Papago. *Natural Language Semantics* 6.245–270.

Quine, Willard Van Orman. 1960. *Word and Object*. Cambridge, MA: MIT Press.

Thomas, David. 1955. Three analyses of the Ilocano pronoun system. *Word* 11.204–208.

Daniel Harbour
Department of Linguistics
School of Modern Languages
Queen Mary, University of London
Mile End Rd, London E1 4KL, UK
harbour@mit.edu

3 Mismatch Phenomena from an LFG Perspective*

Ana R. Luís (University of Coimbra)
Ryo Otoguro (University of Essex)

3.1 Introduction

At the center of lexicalist theories of grammar such as Lexical Functional Grammar is the assumption that inflected verb forms must correspond to one and only one phrase structure node. Cross-linguistic evidence however seems to support the search for a more complex interface between morphology and phrase structure (cf. Wescoat 2002, Luís and Sadler 2003). In this paper, we examine inflected verb forms in European Portuguese and Hindi-Urdu that are challenging to the one-to-one correspondence between wellformed morphological strings and syntactic nodes. We show that in European Portuguese (EP) stem-affix strings can be broken up by separate syntactic word units, and that in Hindi-Urdu verbs inflect as distinct parts in the future tense while they are expressed synthetically in the phrase structure. We claim that these phenomena result from a mismatch between the morphological structure of words and their representation in the phrase structure.

In our treatment of the data, we provide an inflectional analysis within the realisational theory of Paradigm Function Morphology (PFM) (Stump 2001) and explore the interface between morphology and syntax within the framework of Lexical Functional Grammar (LFG) (Bresnan 2001). In our proposal, morphology and phrase structure are treated as parallel levels of representation with distinct wellformedness conditions and the structural correspondence between both levels is defined through mapping principles. Based on this division of labour, we formulate an account of the structural mismatches in EP and Hindi-Urdu by introducing minimal changes into the classical LFG mapping between morphology and c-structure mapping.

The paper is organised as follows. Section 2 offers a brief introduction to the formal framework of LFG, with particular emphasis being given to the level of c(onstituent)-structure. Section 3 introduces the theory of PFM and discusses the correspondence between morphology and c-structure. Section 4 and 5 describe the data and provide an LFG analysis of the mismatch phenomena in EP and Hindi-Urdu. The final section offers a short summary.

*Acknowledgements: We are grateful to Andrew Spencer, Louisa Sadler, Ron Kaplan, Mary Dalrymple, Joan Bresnan and Gergana Popova for fruitful discussions and comments at various stages of our work. Any remaining errors are our own.

Essex Research Reports in Linguistics 47, 31–45.
Ryo Otoguro, Gergana Popova and Andrew Spencer (eds.).
Department of Language and Linguistics, University of Essex,
Colchester, United Kingdom.

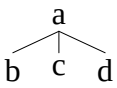
3.2 The c-structure in LFG

Morphology and configurational syntax constitute independent levels of linguistic structure in LFG and a strong division is assumed between word-internal structures, on the one hand, and structures between words, on the other. In what follows, we briefly survey standard lexicalist assumptions about both a) the wellformedness constraints that apply at the phrase structure level and about b) the c-structure representation of words.

3.2.1 Phrase structure trees

C(onstituent)-structure is the level at which linear ordering and hierarchical relations between words and phrases are represented through phrase structure trees. A tree diagram in LFG is said to be composed of nodes which satisfy relations of dominance and precedence (Partee *et al.* 1993:437–44, Kaplan 1995). Dominance results from the hierarchical grouping of the parts of the sentence into constituents and is expressed as vertical alignment between the nodes. Precedence is obtained from the left-to-right order of the constituents and is expressed as horizontal alignment. In (1), pairs $\langle a, b\rangle$, $\langle a, c\rangle$, $\langle a, d\rangle$ are in dominance relation *D,* i.e. node *a* dominates nodes b, c and d; and $\langle b, c\rangle$, $\langle b, d\rangle$, $\langle c, d\rangle$ are in precedence relation P, i.e. node *b* precedes nodes *c* and *d* and node *c* precedes node *d*.

(1)



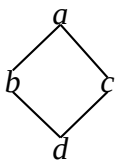
For a phrase structure tree to be wellformed it must meet the following wellformedness conditions (Partee *et al.* 1993:439–41):

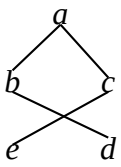
(2) a. **The Single Root Condition**:
In every well-formed constituent structure tree there is exactly one node that dominates every node.

b. **The Exclusivity Condition**:
In any well-formed constituent structure tree, for any nodes x and y, x and y stand in the precedence relation *P,* i.e., either $\langle x, y\rangle \in P$ or $\langle y, x\rangle \in P$, if and only if x and y do not stand in the dominance relation *D,* i.e., neither $\langle x, y\rangle \in D$ nor $\langle y, x\rangle \in D$.

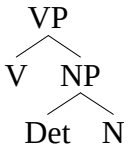
c. **The Non-tangling Condition**:
In any well-formed constituent structure tree, for any nodes x and y, if x precedes y, then all nodes dominated by x precede all nodes dominated by y.

(2a) is straightforward. (2b) prohibits two nodes from standing both in a dominance and precedence relation, as in $\langle a, b\rangle \in P \wedge \langle a, b\rangle \in D$; this ensures that no precedence relation holds between mother and daughter and no dominance relation holds between sisters. (2c) rules out trees in which the precedence relation between mother nodes is not preserved between daughter nodes, as in (3).

(3) a. 

b. 

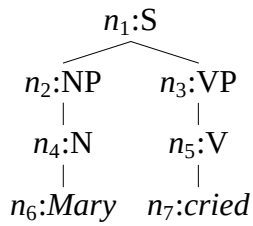
Phrase structure tress also provide information about the grammatical type of each constituent. Through the *labelling* of nodes, each node carries exactly one label that identifies the node's syntactic category and hierarchical position (i.e., bar-levels). A labelled phrase structure tree is given in (4).

(4)


These fundamental aspects of phrase-structure trees (i.e. c-structures in LFG), can be summarised in terms of the following tree-defining properties and relations (Kaplan 1987, 1995):

(5) N: set of nodes
L: set of labels
$\mathcal{M}$: $N \rightarrow N$ (dominance)
$\prec \subseteq N \times N$ (precedence)
λ: $N \rightarrow L$ (labelling)

N and L contain a set of nodes and labels respectively. Function $\mathcal{M}$ maps a daughter node onto its mother node, describing the dominance relation between the two. $\prec$ describes the precedence relation between two nodes. λ is a labelling function that associates a node with a label. With these mathematical formulae, a simple c-structure like (6a) can be described in terms of the set of equations given in (6b):

(6) a.


b. $N = \{n_1, n_2, n_3, n_4, n_5, n_6, n_7\}$
$L = \{\text{S, VP, NP, V, N, } \textit{Mary}, \textit{cried}\}$

$\mathcal{M}(n_2) = n_1$	$\mathcal{M}(n_3) = n_1$	$\lambda(n_1) = \text{S}$	$\lambda(n_2) = \text{NP}$
$\lambda(n_3) = \text{VP}$	$\mathcal{M}(n_4) = n_2$	$\lambda(n_4) = \text{N}$	$\mathcal{M}(n_5) = n_3$
$\lambda(n_5) = \text{V}$	$\mathcal{M}(n_6) = n_4$	$\lambda(n_6) = \textit{Mary}$	$\mathcal{M}(n_7) = n_5$
$\lambda(n_7) = \textit{cried}$	$n_2 \prec n_3$		

In addition, the wellformedness conditions given in (3) are further combined with language-specific constraints regulating the distributions of constituents. In English, for example, S dominates NP and VP, VP dominates V and optional NP, PP and so on. Such

distributional constraints can be stated by a set of equations as in (7a), but Phrase Structure (PS) rules such as (7b) are normally used:

(7) a. $\mathcal{M}(n_2) = n_1 \wedge \mathcal{M}(n_3) = n_1 \wedge \lambda(n_1) = \text{S} \wedge \lambda(n_2) = \text{NP} \wedge \lambda(n_3) = \text{VP} \wedge n_2 \prec n_3$

b. S → NP VP

(7b) is a constraint stating that a node labelled as S dominates a node labelled as NP and a node labelled as VP and that the NP linearly precedes the VP. The grammar of English is thought to contain a set of language-specific constraints including (7b). Another language may contain different PS rules like S → NP, NP, V where the order among two NPs and V is freely exchangeable. Finally, unlike in transformational grammars, there are no derivational operations changing one c-structure into another.

3.2.2 C-structure and the representation of words

In LFG, words have a special status and the role of morphology is to process morphological operations which create fully inflected words. Morphological operations (such as the combination of roots and affixes, or the change of stem forms, among other) are separated from syntactic ones, as defined by the principle of lexical integrity:

(8) morphologically complete words are leaves of the c-structure tree and each leaf corresponds to one and only one c-structure node. (Bresnan 2001:92)

One of the implications of the principle of lexical integrity is that the internal structure of a word must be invisible to configurational syntax ("morphologically complete words are leaves of the c-structure tree"). Elements smaller than a word cannot occupy a terminal node and only fully inflected words can be inserted into c-structure. Thus, in the c-structure representation terminal nodes are labelled with morphologically complete word forms. For example, in (6), the word forms *Mary* and *cried* are instantiations of the pre-terminal nodes N and V, respectively. The statement in (8) also postulates that the correspondence between morphological strings and terminal nodes must be isomorphic ("each leaf corresponds to one and only one c-structure node"). By assuming that only complete and well-formed words can conform to this ismorphism, one further implication of the lexical integrity principle is the idea that the morphological wellformedness of a string is dependent on a string's ability to appear under one and only one c-structure node.

3.3 The interaction between morphology and phrase structure

Bresnan (2001) observes that "while the relative order of words in sentences is extremely free [...], the relative order of stems and inflections in words (such as the case and tense markers) is fixed." Thus, the prime role of the one-to-one correspondence between morphological strings and c-structure terminals is to preserve the divison of labour between morphology and syntax and to stop wordformation from taking place in the syntax.

In this paper, a different approach to the morphology/c-structure mapping is adopted. The problem with the one-to-one correspondence between wellformed words and terminal nodes is the fact that there are morphologically complete stem-affix strings that do not correspond to one single terminal node, even though such strings are morphologically wellformed. The question we need to ask then is whether the morphological wellformedness of strings should be defined at the interface between morphology and syntax (as entailed by the principle of lexical integrity) or whether it should be defined inside the morphology. In this section, we take the view that wellformed strings are defined by the Paradigm Function of a language through the correct associations between morphological strings and complete sets of morphosyntactic features (or f-descriptions).

3.3.1 Morphological wellformedness

An important device in PFM is the Paradigm Function (PF) which takes as its argument a root of a lexeme (X) and a complete set of morphosyntactic features associated with that lexeme (σ), and delivers as output an inflected form of that lexeme. The PF therefore constitutes a wellformedness constraint over a morphologically complete inflectional string. Following Spencer (2004, ms), we define the PF as in (9):

(9) PF($\langle$LEXEME, $\sigma\rangle$) = $_{def}$
S: selection of stem
R: realisation rules specifying set of affixes
L: linearisation of affixes with respect to the stem

(9) illustrates the three subfunctions that make up the PF: S selects the stem, R yields the exponents and L linearises the exponents with respect to the stem. A more concrete example is given in (10) for the verb form Portuguese verb form *gostas* ('you.SG like'):

(10) σ = {(↑ TENSE) = PRES, (↑ SUBJ PERS) =$_c$ 2, (↑ SUBJ NUM) =$_c$ PL}[1]
PF($\langle$LIKE, $\sigma\rangle$) =$_{def}$
S: *gosta*
R: *s*
L: *gosta* ≺ *s*

The PF can also define the wellformedness of periphrastic expressions, as shown in (11), for *ātā hũ*'I come' — the habitual present of COME in Hindi-Urdu. The PF must select two stems, namely *ā* (the stem of the main verb) and *h* (the stem of the auxiliary). This is done by selecting two stems through the subfunction S (hence S-i and S-ii). Thus, the synthetic verb *gostas* in (10) and the analytic espression *ātā hũ* in (11) differ in the number of stems each one contains.

[1] In previous work, we have used the standard PFM definition of σ, as a set of moprhosyntactic features. However, since PFM is being used here as the morphological component of LFG, it is also possible to define σ as a complete f-description (cf. Otoguro 2006). Other attempts at incorporating PFM into the LFG architecture include Sadler and Spencer (2001), Luís and Sadler (2003), Spencer (to appear), Otoguro (2003), Ackerman and Stump (2004), Luís and Otoguro (2004), among other.

(11) σ = {(↑ TENSE) = PRES, (↑ ASP PERF) = −, (%AGR NUM) = SG, (% AGR PERS) = 1, (%AGR GEND) = MASC}
PF(⟨COME, σ⟩) =$_{def}$
S: i. $\overline{a}$
ii. h
R: i. $t\overline{a}$
ii. $\tilde{u}$
L: i. $\overline{a} \prec t\overline{a}$
ii $h \prec \tilde{u}$

Each stem must undergo regular morphological operations. (Note that $\overline{a}t\overline{a}$is the imperfective singular masculine form of the main verb and $h\tilde{u}$ is the present first singular form of the auxiliary). We perform those operations by applying the sub-functions R (exponent specification) and L (linearisation): R-i realises the imperfective singular masculine ending of the main verb and R-ii realises the present first singular ending of the auxiliary; and L-i and L-ii position each ending to the right of the stem.[2]

Under this view of periphrasis, a complete f-description σ can be associated with more than one inflected word form, thus capturing the idea, formulated in Sadler and Spencer (2001), that the f-description is not paired with each word of the analytic expression, but with the analytic expression as a whole.

To sum up, then, the morphological wellformedness of strings is defined by the PF as the correct association between morphological expressions and their complete f-descriptions. This association holds for both synthetic and periphrastic expressions.

3.3.2 C-structure correspondence

Once we correctly pair an inflectional string and a f-description, the next question we need to ask is how to put morphological strings in correspondence with the phrase structure.

Before we consider the morphology/c-structure correspondence, let us return briefly to the wellformedness conditions addressed in section 6.2. One important observation to make is that the conditions enforced by the PS rules are only applicable to the containment relations above the X^0 level. Therefore, the dominance relation between a pre-terminal node and a terminal node such as N/*Mary* or V/*cries* requires a separate principle. This instantiation relation between N and *Mary* or between V and *cries* is regulated by the equality that holds between the category of the lexical item and the category of the node under which it appears: e.g., *cries* is labelled V in the lexicon, so that it can only be dominated by a V node in the c-structure.

Under the current proposal where *gostas* and $\overline{a}t\overline{a}\ h\tilde{u}$are fully inflected expressions of the lexeme LIKE and COME, respectively, we state the correspondence between morphological expressions and their c-structural positions by applying a labelling function to a pair of morphological string and σ. We abbreviate this application as follows:[3]

(12) a. *gostas* ⇒ *gostas*$_{\text{I}}$

[2] %AGR is a local name given to the agreement controller. See Otoguro (2006, to appear) for details.
[3] See Luís and Otoguro (2004), Luís and Otoguro (2005), Otoguro (2006) for details.

b. $\bar{a}t\,\bar{a}\;h\tilde{u} \Rightarrow \bar{a}t\,\bar{a}_{V}\;h\tilde{u}_{\mathrm{AUX}}$

In (12), the labels I, V and AUX define the category of the c-structure nodes under which each one of the stem-affix strings must be positioned (e.g., the label on $\bar{a}t\,\bar{a}$reflects the fact that in Hindi/Urdu these forms must be inserted under V).

One crucial aspect about the morphology/c-structure correspondence illustrated in (12) is that it does not map 'morphological strings' onto c-structure nodes. Instead, the algorithm in (12) places morphological tokens in correspondence with syntactic atoms . These additional layers of structure can be defined as follows:

(13) a. Morphological tokens are wellformed stem-affix strings defined by the Paradigm Function.

b. Syntactic atoms are leaves on c-structure trees; each leaf corresponds to one and only one terminal node.

Even though, in principle, one morphological token corresponds to one syntactic atoms (cf. (12)), the fundamental contribution of our approach is that these additional layers of structure allow us to capture morphology/c-structure mismatches which would otherwise be left unaccounted for. In sections 3.4 and 3.5, we show that the non-isomorphic correspondence between morphological tokens and syntactic atoms provides an insightful explanation for the conflicting morphological and syntactic properties of certain verb forms in EP and Hindi-Urdu. For EP we map one morphological token onto more than one syntactic atom (crucially, without being committed to the representation of incomplete morphological strings in the phrase structure) and for Hindi-Urdu we map two morphological tokens onto one syntactic node.

Summing up: we depart from classical LFG in that we do not insert morphological strings directly into the c-structure. Instead, at the interface between morphology and c-structure, we place morphological tokens in correspondence with syntactic atoms. The syntactic position of each syntactic atoms is defined by the mapping between the two levels via L .

3.4 European Portuguese

In this section, we examine data in which one morphological token (i.e., one complete stem-affix string) corresponds to more than one syntactic atom (i.e., single terminal node in the c-structure). An analysis analysis will be proposed based on the assumption that parts of words can be represented in the c-structure without violating the integrity of words.

3.4.1 The data

In EP, pronominal clitics can be placed post-verbally (as enclitics) or pre-verbally (as proclitics). In either position, citics display a significant number of affix properties, such as fusion (14), syncretism (15), and cluster-internal allomorphy (16):

(14) a. disse-**mo** (*me-o)
said-DAT.1.SG-ACC.3.SG.MASC
's/he said it to me'

b. ... que **mo**-disse (*me-o)
that DAT.2.PL-ACC.3.SG.MASC-said
'...that s/he said it to me'

(15) a. deu-**lho** (*lhe-o)
gave-DAT.3.SG/PL-ACC.3.SG.MASC
's/he gave it to him/them'

b. ... que **lho**-deu (*lhe-o)
that DAT.3.SG/PL-ACC.3.SG.MASC-gave
'...that s/he gave it to him/them'

(16) a. deu-**no-lo** (*nos-o)
gave-DAT.2.PL-ACC.3.SG
's/he gave it to us'

b. ... que **no**-**lo**-disse (*nos-o)
that DAT.2.PL-ACC.3.SG.MASC-said
'...that s/he said it to us'

Affix-like properties are also found inside the cluster: a) clitics are rigidly ordered (i.e., dative clitics must precede accusative clitics, although NPs are generally ordered in the opposite order) and b) show idiosyncratic co-occurrence restrictions (i.e., 1st and 2nd person clitics cannot co-occur inside the same cluster, even though there is no semantic nor syntactically principle ruling out such combination at the phrasal level). Based on such evidence, Luís (2004) takes the view that EP enclitics and proclitics should be analysed as the same affix unit and assigned the ability to be positioned before and and after the verb. Similar claims have been made about clitic pronouns in other Romance languages (Miller and Sag 1997, Monachesi 1999).

The inflectional status of clitic pronouns is captured in (17). The PF defines one complete stem-affix combination and where each verb-clitic string is the wellformed realisation of a pair $\langle$SEE, $\sigma\rangle$, where σ corresponds to a complete set of morphosyntactic features. The difference between (17a) and (17b) is that the clitic is linearised postverbally in (17a) and preverbally in (17b).

(17) PF defining *vê-me*/*me vê* ('seems me')

a. PF($\langle$SEE, {($\uparrow$ TENSE) = PRES, ($\uparrow$ SUBJ PERS) $=_c$ 3, ($\uparrow$ SUBJ NUM) $=_c$ SG}$\rangle$) $=_{def}$
S: *vê*
R: *me*
L: *vê* $\prec$ *me*

b. PF($\langle$SEE, {($\uparrow$ TENSE) = PRES, ($\uparrow$ SUBJ PERS) $=_c$ 3, ($\uparrow$ SUBJ NUM) $=_c$ SG, ($\uparrow$ REST) = +}$\rangle$) $=_{def}$
S: *vê*
R: *me*
L: *me* $\prec$ *vê*

However, as observed in previous work (Luís 2004), the difference between enclitics and proclitics is not just right/left linearisation with respect to the verb. At the level of c-structure, while enclitics behave like genuine verbal suffixes, proclitics exhibit properties that are not typical of verbal prefixes. The contrast is illustrated in (18)–(19). In (18), the enlitic must be adjacent to the host and it can trigger/undergo non-productive allomorphy.

(18) Levamo **-la** (not: levamos-a)
take -ACC.3.SG.FEM
'We well take her'

Proclitics, on the other hand, show a certain degree of syntactic separability. Adverbial words (given in italics) can intervene between the proclitic and the verb. [4] suggesting that proclitics are visible to the syntax:

(19) ...acho que ela **o** *ainda não* disse
...think that she ACC.3.SG.MASC yet not told
'...I think that s/he hasn't told it to him/her/them yet'

Hence, although both enclitics and proclitics are constructed as verbal affixes, enclitics attach to a verbal stem while proclitics select a phrasal position in the syntax (Luis 2004). We therefore propose the following PS rule to account for the adverbial position, between the verb and the proclitic :

(20) $\mathrm{I} \rightarrow \underset{\uparrow = \downarrow}{\mathrm{Cl}} \quad \underset{\downarrow \in (\uparrow \mathrm{ADJ})}{\widehat{\mathrm{Adv}}^*} \quad \underset{\uparrow = \downarrow}{\mathrm{I}}$

The ability for affixes to display phrasal behaviour has been previously attested in other languages. In effect, in the literature, formatives showing both morphological affix status and syntactic independence are often called phrasal affixes (Klavans 1985, Anderson 1992, Luís 2004). We therefore follow Luís (2004) and itake the view that EP proclitics are best regarded as phrasal affixes.

The question we need to address now is how to represent phrasal affixes at the level of c-structure.

3.4.2 Proposal

Under the assumption that phrasal affixes result from a mismatch between morphological and syntactic units, the labelling algorithm introduced in section 3.3.2 will allow us to capture the fact that one stem-affix string in the morphology is mapped onto two terminal nodes in the c-structure.

The correspondence between clitic-verb strings and terminal nodes is defined as in (20), where a) the morphological token is given in square brackets, b) H is the verb stem and c) 'x' and 'y' are inflectional affixes.

(21) $[x\text{-H-}y] \Rightarrow x_{\mathrm{Cl}} \ \text{H-}y_{\mathrm{I}}$

[4] One further difference between enclitics and proclitics is that proclitics can take a scope over a co-ordinated phrase .

The result of applying (20) to the output strings in (17) yields the labelled syntactic atoms given in (21).

(22) a. [*me-vê*] ⇒ *me* $_{\text{Cl}}$ *vê* $_{\text{I}}$
b. [*vê-me*] ⇒ *vê-me* $_{\text{I}}$

The morphological token in (21a) is mapped onto one syntactic atom of category I, while the morphological token in (21b) is mapped onto two syntactic atoms. Each category label constraints the position in which the atom occurs in the phrase structure. In EP, finite verbs are in an I^0 position. The proclitic is mapped onto a Cl category to capture the fact that it must be positioned under a non-projecting Cl node within I as stated in the annotated PS rule in (20). Thus, the insertion of syntactic atoms into the phrase-structure is regulated by standard phrase structure principles (e.g., immediate dominance, linearisation and instantiation) in combination with PS rules (for more details, cf. Luís and Otoguro 2004, 2005).

To conclude, the mapping between morphological tokens and syntactic atoms enables us to accommodate phrasal affixes within LFG c-structure without inserting the actual affix string directly into the c-structure thereby preserving the strict separation between morphology and syntax.

3.5 Hindi-Urdu

While the previous section has shown that wellformed stem-affix strings can be syntactically expressed by more than one syntactic atom, this section will examine verb forms in Hindi/Urdu in which one syntactic atom corresponds to more than one morphological token. Such data lends further empirical support to a non-isomorphic approach to the morphology/c-structure mapping in LFG.

3.5.1 The data

In Hindi/Urdu, verbs inflect for two sets of agreement features depending on tense, aspect and mood. Perfective and imperfective participles agree in gender and number, as in (23a–b), whereas subjunctive participles agree in person and number, as in (23c):

(23) a. vah ā̄yā̄.
he come.PERF.SG.M
'He came.'
b. (agar) vah ā̄tā̄.
(if) he come.IMPF.SG.M
'(if) he had come.'
c. (agar) vah ā̄e.
(if) he come.SUBJUNCT.3.SG
'(if) he comes.'

The list of agreement endings is given in (24).

(24) a.

	Imperfective		Perfective	
	Masc	Fem	Masc	Fem
Sg	-tā	-tī	-ā	-ī
Pl	-te	-tī	-e	-ī

b.

Subjunctive			
	1	2	3
Sg	-ũ	-e	-e
Pl	-ẽ	-o	-ẽ

Participles can also be followed by auxiliaries to realise tense, aspect and mood (McGregor 1995, Schmidt 1999). One of those auxiliary verbs is *hona* which inflects person and number in the present and for number and gender in the past. In periphrases, therefore, the lexical verb and the auxiliary can inflect for different sets of agreement features. In (25a), for instance, the perfective participle agrees in gender and number while the present auxiliary agrees in person and number (cf. perfective + past auxiliary in (25b)).

(25) a. Immediate past (perfective + present)

vah āyā hai.
he come.PERF.SG.M be.PRES.3.SG

'He has come.'

b. Remote past (perfective + past)

vah āyā thā.
he come.PERF.SG.M be.PAST.SG.M

'He had come/he came.'

Thus, one important observation about periphrastic tenses in Hindi-Urdu is that they can display both agreement patterns simultaneously, provided each set of features occurs on a different verb.

On the contrary, future verb forms, systematically display both sets of agreement features.

(26) ham āe-ge
we subjunct.1.PL-FUT.PL.MASC
'We will come.'

In (26), the verb stem *āe* is a subjunctive participle that agrees in person and number (cf. (24b), while the suffix which follows the stem agrees in number and gender (cf. (24a)). As the paradigm in (27) shows, each part of a single syntactic word effectively inflects for a different set of agreement features (subjunctive stem is in boldface).[5]

[5]The suffix is a grammaticalised form of the perfective form of *jā* 'go'. In its full form, it inflects as follows:

	Sg	Pl
Masc	gayā	gae
Fem	gaī	gaĩ

However, as a gramaticalised form, the auxilairy has lost its original meaning and undergone lexical reduction: *gayā* → *gā*, *gae* → *gā*, *gaī* → *gī*

(27) Future (*ān ā* 'come')

	Masc		Fem	
	Sg	Pl	Sg	Pl
1	**āũ**gā	**āẽ**ge	**āũ**gī	**āẽ**gī
2	**āe**gā	**āo**ge	**āe**gī	**āo**gī
3	**āe**gā	**āẽ**ge	**āe**gī	**āẽ**gī

The fact that both sets of agreement features can occur within the same verb form suggests to Otoguro (2006) that future verb forms do not behave morphologically like one synthetic verb forms (cf. (24)). Instead, the inflectional pattern in (27) seems to indicate that they behave morphologically like periphrastic expressions.

3.5.2 Proposal

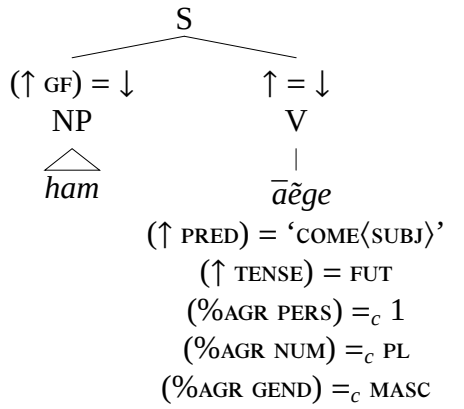
We will now show that assigning a periphrastic structure to the future verb forms enables us to account for their peculiar double-agreement. Our proposal entails the claim that future verb forms in Hindi-Urdu display a mismatch between their morphological structure (which is periphrastic) and their c-structure representation (which is synthetic).

In the morphology, we define the periphrastic structure of *āege* through the PF in (28). As alluded to before, periphrastic strings select more than one stem . Therefore, the PF in (28) defines two stems, namely *ā* (< *ān*) and *g* (< *hona*). The endings attached to each stem are specified by R and their linearisation is defined by L. Crucially, the former inflects for person and number and the latter for number and gender. Under this analysis, double-agreement marking on future verbs results from the combination of two stems with their appropriate endings.

(28) σ = {(↑ TENSE) = FUT, (%AGR NUM) =$_c$ PL, (%AGR PERS) =$_c$ 1, (%AGR GEND) =$_c$ MASC}
PF(⟨COME, σ⟩) =$_{def}$
S: i. *ā*
ii. *g*
R: i. *e*
ii. *e*
L: i. *ā* ≺ *e*
ii. *g* ≺ *e*

Having analysed future verb forms as comprising two complete and wellformed stem-affix strings, the next task is to place the periphrastic expression in correspondence with the c-structure. The fact that *āege* is positioned under the V^0 node (cf. (29) indicates that future forms are periphrastic in the morphology but synthetic in the syntax. We therefore need to account for the mismatch that exists between their morphological structure and the syntactic representation.

(29)



In Otoguro (2006), the labelling algorithm captures this type of mismatch by assigning only one category label to both morphological tokens. This is illustrated in (30a) where both stem-affix strings correspond to one syntactic atom and where the category label V reflects the verb's c-structural position. In contrast, the periphrastic expression *āy ā hai* in (30b) illustrates a case of one-to-one mapping between morphological tokens and syntactic atoms. Each morphological token is labelled with one category which indicates its position in the c-structure.

(30) a. [*āe*] [*ge*] ⇒ $\bar{a}ege_{V}$

b. [*āy ā*] [*hai*] ⇒ $\bar{a}y\bar{a}_{V}$ hai_{AUX}

Summing up, this section has shown that there are two types of periphrastic expressions in Hindi-Urdu: i) typical cases in which two stem-affix strings correspond to two independent terminal nodes, and ii) 'pseudo-periphrastic' expressions in which two stem-affix strings correspond to one independent terminal nodes. The latter are periphrastic in the morphology but synthetic in the syntax and their structural peculiarity is captured at the interface between morphology and the c-structure as a mismatch between morphological tokens and syntactic atoms.

3.6 Conclusion

This paper has examined the non-isomorphic correspondence between morphological strings and phrase structure nodes in EP and Hindi-Urdu. Empirical evidence has motivated the distinction between two types of 'word units': morphological tokens and syntactic atoms. Each one plays a different role in the grammar and is sensitive to different wellformedness constraints. The former is purely morphological and sensitive to the regular morphological principles, whereas the latter is the purely configurational and sensitive to regular c-structural principles. Our proposal has helped us solve the dual properties of EP proclitics by allowing us to represent phrasal affixes as a distinct syntactic atom. We have also provided an explanation for the partly periphrastic and partly synthetic properties of future verb forms in Hindi-Urdu. Under our account, they are morphologically periphrastic but syntactically synthetic. Finally, the formal modifications made into the LFG theory are in harmony with core assumptions about the morphological integrity of words and the division of labour between morphology and c-structure.

Bibliography

ACKERMAN, FARRELL, and GREGORY T. STUMP. 2004. Paradigms and periphrastic expression: A study in realizaion-based lexicalism. In *Projecting Morphology*, ed. by Louisa Sadler and Andrew Spencer, 111–157. Stanford, CA: CSLI Publications.

ANDERSON, STEPHEN R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

BRESNAN, JOAN. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

DALRYMPLE, MARY, RONALD M. KAPLAN, JOHN T. MAXWELL, III, and ANNIE ZAENEN. 1995. *Formal Issues in Lexical-Functional Grammar*. Stanford, CA: CSLI Publications.

KAPLAN, RONALD M. 1987. Three seductions of computational psycholinguistics. In *Linguistic Theory and Computer Applications*, ed. by P. Whitelock, M. M. Wood, H. L. Somers, R. Johnson, and P. Bennet, 149–181. London: Academic Press. Reprinted in Dalrymple *et al.* (1995).

—— 1995. The formal architecture of Lexical-Functional Grammar. In *Formal Issues in Lexical-Functional Grammar*, ed. by Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell, III, and Annie Zaenen, 7–27. Stanford, CA: CSLI Publications.

KLAVANS, JUDITH. 1985. The independence of syntax and phonology in cliticization. *Language* 61.95–120.

LUÍS, ANA, 2004. *Clitics as Morphology*. Doctoral dissertation. University of Essex.

——, and RYO OTOGURO. 2004. Proclitic contexts in European Portuguese and their effect on clitic placement. In *Proceedings of the LFG04 Conference*, ed. by Miriam Butt and Tracy Holloway King, 334–353, Stanford, CA. CSLI Publications.

——, ——. 2005. Morphological and syntactic well-formedness: The case of European portuguese proclitics. In *Proceedings of the LFG05 Conference*, ed. by Miriam Butt and Tracy Holloway King. Stanford, CA: CSLI Publications.

——, and LOUISA SADLER. 2003. Object clitics and marked morphology. In *Empirical Issues in Formal Syntax and Semantics*, ed. by C. Beyssade, O. Bonami, P. Hofherr, and F. Corblin, volume 4, 133–153. Paris: Presses Universitaires de Paris-Sorbonne.

MCGREGOR, R. S. 1995. *Outline of Hindi Grammar*. Oxford: Oxford University Press, third edition.

MILLER, PHILIP, and IVAN SAG. 1997. French clitic movement without clitics and movement. *Natural Language and Linguistic Theory* 15.573–639.

MONACHESI, PAOLA. 1999. *A Lexical Approach to Italian Cliticization*. Stanford, CA: CLSI Publications.

OTOGURO, RYO. 2003. Focus clitics and discourse information spreading. In *Proceedings of the LFG03 Conference*, ed. by Miriam Butt and Tracy Holloway King, 367–386, Stanford, CA. CSLI Publications.

——, 2006. *Morphosyntax of Case: A Theoretical Investigation of the Concept*. Doctoral dissertation. University of Essex.

——. to appear. Agreement and path specification in Icelandic. In *Papers from the 41st Regional Meeting of the Chicago Linguistic Society*.

PARTEE, BARBARA H., ALICE TER MEULEN, and ROBERT E. WALL. 1993. *Mathematical Methods in Linguistics*. Dordrecht: Kluwer Academic Publishers.

SADLER, LOUISA, and ANDREW SPENCER. 2001. Syntax as an exponent of morphological features. In *Yearbook of Morphology 2000*, ed. by Geert Booij and Jaap van Marle, 71–96. Dordrecht: Kluwer Academic Publishers.

SCHMIDT, RUTH LAILA. 1999. *Urdu: An Essential Grammar*. New York: Routledge.

SPENCER, ANDREW. 2004. Generalized Paradigm Function Morphology – a synopsis. *York Papers in Linguistics* .

——, ms. Generalized Paradigm Function Morphology. University of Essex.

——. to appear. Paradigm-based lexicalism: Negation in Japanese. In *Paradigms and Periphrases*, ed. by Farrell Ackerman, James Blevins, and Gregory Stump. Stanford, CA: CSLI Publications.

STUMP, GREGORY T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

WESCOAT, MICHAEL T., 2002. *On Lexical Sharing*. Doctoral dissertation. Stanford University.

Ana R. Luís
Department of Anglo-American Studies
University of Coimbra
3004-530 Coimbra, Portugal
aluis@ci.uc.pt

Ryo Otoguro
Department of Language and Linguistics
University of Essex
Wivenhoe Park, Colchester CO4 3SQ, UK
rotogu@essex.ac.uk

4 Udi Clitics: A Generalized Paradigm Function Morphology Approach*

Ana R. Luís (University of Coimbra)
Andrew Spencer (University of Essex)

4.1 Introduction

Subject person markers (henceforth PM) in Udi, a North East Caucasian language, can appear in any one of the following positions: at the right edge of a focused constituent, verb-finally, inside complex stems or inside monomorphemic verbstems. Harris (2000, 2002) accounts for the peculiar placement of PMs by treating word-final and edge-final PMs, in (1a), as enclitics, and word-internal PMs, in (1b), as endoclitics:[1]

(1) a. äyel-en p'ạ ẹs-**ne** aq'-e
child-ERG two appple-3SG.PM
'The child took two apples'

b. azq'e
a **z** q' e
take **1SGPM** take AORII
'I received'

Harris shows in detail that the enclitic uses of the PM conform to standard definitions of clitic (for instance the tests of Zwicky and Pullum (1983)). The claim that the verb form in (1b) contains an endoclitic necessarily entails the assumption that words can break up other words or, paraphrasing somewhat, that the syntax can see inside words, in clear violation of the Lexical Integrity Hypothesis, under which only affixes can appear word-internally. In this paper we argue that the conclusion doesn't follow if we change one important premise, namely, the assumption that the PM in their incarnation as clitics are placed by syntactic rules. On the contrary, we will argue that the enclitics are phrasal affixes and that phrasal affixes are morphological objects (affixes) placed with respect to syntactic categories rather than morphological categories.

*Much of the work reported here was conducted while AS was the recipient of a British Academy/Leverhulme Trust Senior Research Fellowship and while AL was the recipient of studentships from the AHRB/British Academy and the Portuguese Foundation for Science and Technology. We are very grateful to these bodies for their support.

[1]We put the PMs in bold face here and throughout.

Essex Research Reports in Linguistics 47, 47–59.
Ryo Otoguro, Gergana Popova and Andrew Spencer (eds.).
Department of Language and Linguistics, University of Essex,
Colchester, United Kingdom.

The analysis we will outline adopts the theory of Generalized Paradigm Function Morphology (Spencer ms., 2004) , and specifically the notion of 'extended paradigm function' explored in Luís and Spencer (2005) for European Portuguese. The analysis accounts for the distribution of person markers by aligning subject agreement markers either a) within a verb stem as (true) affixes or b) to the right of a focussed phrasal host, as clitics. This means that, as in European Portuguese, affixhood or clitichood is not an inherent property of the PM itself. Moreover, the PMs themselves will be treated as morphological objects no matter where they are placed. It is therefore a consequence of our analysis that complex verb stems are not broken up by syntactic objects, but by morphological objects. Given this analysis, the facts illustrated in (1) are derived within the morphology and the Lexicalist Hypothesis can be maintained.

The structure of the paper is as follows: we start by offering a quick overview of the verb structure in Udi. Section 3 surveys the facts about PM placement in Udi, based on Harris' (2002) detailed description. Section 4 summarises Harris' cliticisation approach while section 5 outlines our inflectional account in informal terms. Section 5 sketches the analysis within the theory of Generalized Paradigm Function Morphology.

4.2 Udi verb structure

To understand the placement of Person markers, we start by presenting the morphological structure of the Udi verb stem (cf. 2.1) and the complete sets of Udi Person markers (cf. 2.2).

4.2.1 Verb stems: simplex and complex

Udi is an agglutinative language in which the minimal verb form comprises a verbstem and a tense-mood.aspect suffix.[2] In (2), the verb form *aq'o* consist of the verbstem *aq'* and the future suffix *-o*.

(2) aq' -o
take FUTI
'(someone) will take'

An important property about Udi is that verbstems can either be simplex or complex. Simplex stems are monomorphemic and constitute the minority pattern in this language:

(3)

aq'-	'take'	bi-	'die'
ef-	'keep'	buq'-	'love'
baq'	'hold'		

Complex verb stems, on the contrary, combine a verb or light verb with an incorporated element, as in (4) (Harris 2002:65):

(4) a. lašk'o-b-
wedding-DO
'marry'

[2]The data used in this paper is taken form Harris (2002).

b. kala-bak
big-BECOME
'grow up'

c. oc'-k'
wash-LV

There are about six light verbs in Udi (*b* 'do', *bak* 'be, become', *p* 'say', *eg* 'come', *d*, *k*), three of which can be used as independent verbstems with independent meanings. Other verbs neither occur independently nor have their own meaning. We gloss the latter simply 'LV' (though it should be borne in mind that glosses for the other light verbs are essentially arbitrary). Incorporated elements can be nouns (4a), adjectives (4b), intransitive simplex verbstem (4c), borrowed verbs, adverbs, or unidentifiable elements (Harris 2002:65). In fact, most verbs in Udi are complex, with relatively small number of simplex stems (a matter we return to in section 5).

4.2.2 Person Markers

PM cross-reference the subject for person and number features and are generally not optional. There are three sets of PM,[3] namely: the Inversion set, the Question set and the General set, as shown in (5). Which one of the sets is selected in a given clause depends on the verb's category. The Inversion set is selected by verbs belonging to the inversion category (*buq'* love, want, *ak'*- see, *ababak*- know[4]); the set labeled Possession is used mainly with verbs denoting possession, while the General set occurs with all other verbs.

(5) Person markers in Udi (Harris 2002:28)

	General	Inversion	Possession
1Sg	-zu, -z	-za	-bez, -bes
2Sg	-nu, -n, -ru, -lu	-va	-vi
3Sg	-ne, -le, -re	-t'u	-t'a
1Pl	-yan	-ya	-beš
2Pl	-nan, -ran, -lan	-va, -v̥ḁn	-e̥f
3Pl	-q'un	-q'o	-q'o

The PMs in (5) exhibit shape variations which are mainly triggered by rules of phrasal phonology. For example, first and second singular person markers *-zu* and *-nu* undergo vowel elision before a consonant or at the edge of words. Also, third singular *-ne* and second singular *-nu* assimilate to a preceding [l] and [r] (and optionally to [d] and [t]) (Harris 2002:34, 67). However, one of the PM variants seems to be triggered by the tense of the verb rather than by phrasal phonology. Vowel ellision with third singular PM (belonging to the General set) takes place (optionally) if *-ne* immediately follows the subjunctive particle *-q'a-* or if *-ne* follows markers of the subjunctive II *-a-y-* (not subjunctive I). In this latter context, ellision is compulsory (Harris 2002:33). It is not

[3]There is also a partial set which is labeled 'Question' set (Harris 2002:28).

[4]Harris (2002:27) also suggests that these verbs have in common the fact that the subject is an experiencer (cf. also chapter 8 on the historical origin of the Inversion set).

clear whether this alternation ought to be regraded as a genuine case of allomorphy of the person marker.[5]

4.2.3 Summary

This section has shown that there are three complete sets of Udi Person markers and two main types of verb stems. In section 4 we return to the internal structure of verbal stems to address an interesting aspect about the evolution of verb stems, namely their etymology. The diachronic evolution of the Udi verb will help us understand why PM occur word-internally as they do.

4.3 Person Markers in Udi

We will now describe in more detail the positions in which PM occur and the conditions determining them. There are two basic types of placement patterns for Udi PM, verbal and phrasal.

4.3.1 Phrasal attachment

Phrasal attachment is perhaps the piece of evidence which most strongly suggests that PM constitute 'special clitics' (in the sense of Zwicky 1977). If there is a focused constituent in the clause, the PM encliticizes to it. Negative particles and question words are obligatoily in focus in Udi and attract therefore PMs, as shown in (6) and (7). If both a negative particle and a question word co-occur then the PM attaches to the negative particle (Harris 2002:119).

(6) nana-n te=***ne*** bụɣa-b-e (te=*ne*) p'ạ ačik'alšey
mother-ERG NEG-**3SG** find-DO-AORII two toy.ABSL
'Mother did not find two toys.'

(7) manu muz-in=***nu*** ayt-exa?
which language-INST-**2SG** word-SAY.PRES
'WHAT LANGUAGE are you speaking?'

The PMs can also attach to other focused arguments, as in (8).

(8) a. yaq'-a=***ne*** ba-st'a
road-DAT-**3SG** in-LV.PRES
'ON THE ROAD he opens it.'

b. merab-en ayt=***ne*** ef-sa
Merab-ERG word.ABSL-**3SG** keep-PRES
'Merab keeps his WORD.'

[5]There is not enough evidence to go into more details about the exact nature of the phonological alternations displayed by the PM. Also, the phonology of Udi has not been stated explicitly yet, as pointed out by Harris (2002:33). We will therefore leave the morphophonology unaddressed.

In clauses with zero copulas, the PM is hosted by the predicate nominal, unless the subject is in focus:

(9) nana k'wa=***ne***
mother.ABSL house.DAT-**3SG**
'Mother is at the house.'

These then are the four hosts to which PM attachment: a) a negative particle, b) a wh-word, c) focused arguments and d) a nominal predicate. How these four contexts interact in terms of precedence relations is explained and exemplified in great detail in Harris (2002). For the present purposes however it suffices to say that the position of the PM, shown in (6–9), results from the interaction between syntax and information-structure. The data has also shown that person markers in Udi appear in places in which full words cannot occur and exhibit promiscuity with respect to the words they attach to. Based on Zwicky and Pullum (1983) and various other studies (Harris 2002:114), Harris argues that properties such as these support the view that PM are special clitics.

4.3.2 Verbal host

There are also contexts in which Udi PMs seem to select their host. In this section, we focus on the attachment of PM to the verb, both verb-finally or verb-internally. They are final in the verb stem if the verb is in the future II, the subjunctive I, the subjunctive II, or the imperative.[6] these TAM categories, which take precedence over the phrasal contexts addressed in section 3.1, position the PM in verb-final position, regardless of whether there is a negative particle or a wh-word (Harris 2002:118–9)

(10) a. boš-t'-al-**ne**
in-LV-FUT-**3SG**
's/he will plant'

b. e/-a-**q'un**?
come-SUBJ**II-3PL**
'will they come?'

c. besp'-a-**nan**
kill-IMP-**2PL**
'You kill [her].'

As to the verb-internal placement of PM, we start with placement inside the complex verb stem. At the outset of this paper we alluded to the fact that PM can break up the components of a complex stem, occurring between the incorporated element and the light verb or verb root. As (11) shows, the PM occurs between morpeme boundaries:

[6]Verb-final placement also takes place with a very small set of suppletive roots consisting entirely of a single consonant (e.g., *b-esa-n* [make-PRES-3SG] 'she makes', *p-e-ne* [say-AORII-3SG] 'she said') and with irregular forms of other verbs (e.g. *aba-t'u* 'she knows' the shortened form of *aba-t'u-bak-e* [know-INV3SG-be-AORII]).

(11) a. eč-es-**ne**-st'a
bring-INF-**3SG**-CAUS.PRES
'(he) brought'

b. zer-ev-**ne**-k'sa
decorate-CAUS-**3SG**-LV-PRES
's/he arranges [the house]'

c. lašk'o-**q'un**-b-esa
wedding-**3PL**-DO-PRES
'married'

The placement in (11) cannot occur if the verb form is future II, the subjunctive I, the subjunctive II, or imperative (i.e., in the presence of the TAM categories resposible for verb-final positioning). In that case, the more general placement rule applies positioning PMs verb-finally:

(12) aš-b-al-**ne**
work-do-FUTII-**3SG**
's/he will work'

Examining the behaviour of PM inside monomorphemic stems, PM can also appear immediately before the final consonant in monomorphemic verb roots. In (13), various examples of intramorphemic placement are given. We follow Harris (2002) in glossing the parts of the discontinuous root twice with different subscript numbers.

(13) a. be-**ne**-ɣ-sa (verbstem *beɣ-*)
look$_1$-**3SG**-look$_2$-PRES
'look at'

b. a-**z**-q'-e (verbstem *aq'-*)
receive$_1$-**1SG**-receive$_2$-AORII
'received'

c. baš-**q'um**-q'-e (verbstem *bašq'-*)
steal$_1$-**3PL**-steal$_2$-AORII
'stole'

d. ba-**ne**-k-sa (verbstem *bak-*)
be$_1$-**3SG**-be$_2$-PRES
'is'

Intramorphemic placement however cannot occur if the stem consists entirely of a single consonant or a CV sequence, in which case PM occur verb-finally (Harris 2002:128):

(14) b-esa-**ne**
make-PRES-**3SG**
'he makes'

Also, specific lexical class trigger intermorphemic placement - between the verbstem and the present tense marker - but only in the intransitive.

(15) a. *bi-**ne**-x-sa* 'gives birth'

b. *bix-**ne**-sa* 'is born'

4.3.3 Summary

Verb-final placement, which is triggered by TMA properties of the verb, constitute the default position. Phrasal attachment occurs if there is a focused constituent in the clause; verb-internal placement takes place when neither the factors triggering verb-final nor phrasal attachment are available in the clause. Simplifying somewhat: intermorphemic placement is only possible with complex stems, containing an incorporated element; intramorphemic placement takes place with monomorphemic stems ending in consonant.

The account of the Udi facts proposed by Harris' (2000, 2002) within Optimality Theory treats the PMs as enclitics in word-final positions and as endoclitics in verb-internal position. However, an alternative approach is possible if we take seriously the idea that edge-final and word-final PMs are phrasal affixes (Klavans (1985), Anderson (1992), Luís and Spencer (2005)). Before sketching our analysis we compare and contrast the two approaches.

4.4 Harris (2002)

Harris (2000, 2002) derives the positioning of the PMs through specific alignment constraints:

a) enclisis to a verb is derived through the alignment constraint which aligns the left edge of the PM with the right edge of the inflected verb form (in the futureII, ending in *-al*; subjunctive I and II, imperative – all ending in *-a*);

b) enclisis to a focussed constituents follows from an alignment constraint which states that the left edge of the PM is aligned with the right edge of the Focused constituent.

c) intermorphemic placement (between IncE and light verb) is captured through an alignment constraint which aligns the left edge of the PM aligned with the right edge of the IncE;

d) intramorphemic placement is derived as 'infixation' incorporating McCarthy's and Prince (2005) insight that certain grammatical structures may incur minimal violations of constraints. Here, the alignment constraint requires that the right edge of the PM be aligned with the *right* edge of the verbstem. Since this entails overlapping segments the constraint will always be violated to some degree. The least violation is incurred when there is a mismatch of only one segment, which effectively means that the PM is moved to the left of the rightmost consonant. This is illustrated in Harris' Tableau 7.5 (Harris 2002:153), adapted here as (16), where '|' indicates the right edge of the verbstem and '+' the right edge of the PM:

(16)

	Candidates	ALIGN-PM-VERBSTEM
	a. ne+beɣ|-e	beɣ!
	b. b-ne+eɣ|-e	eɣ!
☞	c. be-ne+ɣ|	ɣ
	d. beɣ-ne+e	ne!

The claim that PM constitute special words positioned by rules of syntax crucially implies that PM violate the lexical integrity of the verbal host when they occur word-internally. Harris' analysis explicitly claims that rules of syntax clitics can see inside words.

The analysis works, but at typological cost: it depends crucially on an appeal endoclisis, i.e., the ability for words (clitics) to occur inside the verb root. As alluded to before, however, this is not an entirely uncontroversial claim, given that it is generally assumed that word-internal constituents are morphological elements not free words. If Harris' analysis proves to be the only one available, then it means that Udi constitutes a counter-example to the widely-held assumption within lexicalist grammar that a) the internal structure of words is not visible to the syntax and that b) the syntax cannot manipulate the structure of words.

4.5 An alternative proposal

In this section we show that the complex Udi data can be given a different interpretation, though one which is only minimally different from Harris' own account. One important property about person markers in Udi is indeed their local relationship with the verb. They not only realise morphosyntactic propoerties typically associated eith verbs, but they also display a high degree of locality with respect to the verb, either when they occur at the edge of it (by default) or inside it. We will therefore explore the idea that they constitute verbal affixes. How can we account for their behaviour within an inflectional analysis?

As will be shown in section 5, an inflectional analysis of verb-internal placement can easily explain why a given affix breaks up verbal stems/roots. Complex stems, even though they behave syntactically and semantically as one word (mostly with non-compositional semantics), are historically composed of two ancient morphemes. Cross-linguistically it is not rare for stems to be discontinuous and split up by inflectional morphemes. Spencer (2003) discusses the discontinuous stems of Athapaskan languages in this light and Spencer (ms.) gives further examples from the Siberian (Yeniseian) language Ket.

Intramorphemic placement raises a different question: if simplex verb stems are single morphemes how is it possible that PM can split them up too? However, even this apparently strange pattern can be incorporated into our affixational account given one additional assumption. Indeed Harris' own discussion hints at a natural solution which involves re-structuring the simplex verb stems.

> "All of the light verbs that form complex verbs consist of a single consonant, except for bak- 'become'. That is, in a typical complex verb, such as aš-ne-b-e, the PM is before the light verb, that light verb being the last consonant of the verbstem. (Harris 2002:213)"

We propose that the remaining simplex stems of the form (C)VC(C) (i.e. the majority) have been reanalysed as complex stems in the modern Udi lexicon. The small number of remaining stems which have a different shape then constitute the exceptions to endoclisis accounted for by Harris' Rule 6.

Now, assuming that verbal attachment (i.e., verb-final, intermorphemic and intramorphemic) can be derived purely within the morphology, the next important question we need to address is how to account for the phrasal attachment of PM which for Harris clear proof that the PMs are positioned by rules of syntax. Recall that PM attach to a focused constituent, regardless of the category of the host. In this respect, one could say, descriptively, at least, that PM behave partly like of genuine verbal affixes and partly like special clitics.[7]

European Portuguese exhibits a very similar patterning of clitics to Udi. In 'enclitic' contexts the Portuguese formatives behave like canonical suffixes, while in 'proclitic' contexts they behave like canonical clitics. In Luís (2004) and Luís and Spencer (2005) European Portuguese clitics are treated as identical formatives in each case but subject to different rules of placement or attachment depending on various morphosyntactic factors. Those works take the clitic contexts to be contexts of phrasal affixation in which morphological formatives are placed with respect to syntactically defined nodes rather than morphologically defined stems. In other words, in phrasal affixation morphology has access to syntactic structure.[8]

Likewise, the Udi PMs constitute (subject agreement) affixes, *irrespective* of where they are placed. The PMs are placed either as stem/root-suffixes (in verb-final and verb-internal position) or as phrasal affixes (in focused contexts). This proposal crucially relies on the assumption that the morphology a) defines the realisation of inflectional exponents over lexemes (not lexemic roots) and b) the definition of exponence (realization), domain (placement) and linearization are treated separately.

One consequence of our analysis is that verb forms in Udi are no longer broken up by syntactic objects (as suggested by Harris 2002), but by morphological objects (cf Luís and Spencer 2005 on 'mesoclisis' in European Portuguese). It is important to hightlight the fact that clitic constructions are not an instance of periphrasis, in which syntactic words realize morphosyntactic properties. Clitics have no syntactic representation (in particular, they aren't syntactic heads). PM aren't even non-projecting words in the sense of Toivonen (2003).

4.6 Generalized Paradigm-Function Account

Our analysis can be couched within the extension to Stump's (2001) theory of Paradigm Function Morphology (PFM) proposed in Luís and Spencer (2005) and Spencer (ms.). In PFM a paradigm function takes a pairing of a root and a set of features and delivers a fully inflected word form. In general the paradigm function is defined in terms of a sequence of realization rules which add successive affixes to the root. We can think of the paradigm function as a set of assembly instructions for word forms.

[7] Similar issues concerning the grammatical status of Udi clitics have been discussed in Crysmann (2000).

[8] Anderson's notion of phrasal affixation is generally used to support the view that all clitics constitute phrasal affixes, even if those so-called clitics behave like genuine word-level affixes. In our work, phrasal affixation is restricted to clitic phenomena in which clitics show no sign of being morphologically attached to a stem.

In Stump's original model the realization rules take the general form X ⇒ X-suffix/prefix-X, where the 'X' can be the lexeme's root or any intermediate affixed form. In the extended PFM model we separate out two distinct aspects of the realization rule, exponence (a definition of the form of each affix) and placement (what kind of stem it attaches to and in what direction). We code this idea by reformulating the notion of paradigm function. This is now a mapping from a pairing of the full representation of the lexeme and a set of features. The paradigm function defines the word form of a given lexeme corresponding to that feature set. In the extended model the paradigm function itself is split into four components, as seen in (17):

(17) Paradigm Function in Extended PFM

(i) Domain

(ii) Host

(iii) Exponence

(iv) Linearization

For affixation the Domain is the word and the Host is some appropriate stem. Like the realization rules of standard PFM, the rules introducing exponence are furnished with a 'rule block index'. This is an integer which is used to linearize affixes within a string. The default Linearization is for each affix to be placed in the order defined by the rule block index, but as Stump (2001) details there are numerous exceptions to this within ordinary affixation (see also Spencer 2003).

The modification of the paradigm function given in (17) allows us to collapse stem affixation with phrasal affixation (cliticization). The parameters in (17) are very similar to the parameters for cliticization proposed by Klavans (1985). Thus, in typical instances of phrasal affixation we have a clitic cluster, whose exponents are defined under (17(iii)) and linearized under (17(iv)). The domain parameter determines the phrase or prosodic category with respect to which placement takes place and the host specifies where within that domain attachment is made. A simple illustration of the difference between the two modes of attachment as defined by a paradigm function is given by the English plural (18) and the English possessive inflections (19):

(18) English plural:
PF(CAT, [Num: Pl]) =
Domain(CAT, [Num: Pl]) = word
Host(CAT, [Num: Pl]) = root(CAT)(by default) = *kat*
Exponence: Formative string(CAT, [Num: Pl]) = /z/ (default plural)
Linearization: /z/ = suffix (by default)

(19) English possessive: *[the cat on the mat's] whiskers*
PF(CAT, [Poss: Yes]) =
Domain(CAT, [Poss: Yes]) = NP headed by CAT
Host(CAT, [Poss: Yes]) = word form right edge of NP = *mat*
Exponence: Formative string(CAT, [Poss: Yes]) = /z/ (default plural)
Linearization: /z/ = suffix (by default)

Ignoring the question of 'endoclisis/infixing' for a moment, we now have the basis for a simple unified description of the Udi facts. Where some (non-verbal) phrase in the clause is in focus that phrase serves as the domain/host for attachment of the PM. A full account of this will require the paradigm function to have access to whatever feature is used to code non-default focus in Udi. For the sake of exposition we will simplify and just assume a feature [Default focus: {yes, no}], where default focus is on the verb. To some extent this means that PM placement is at least a partial exponent of this focus feature. Once we grant to the paradigm function the possibility of making explicit reference to the nature of focussing, the revised definition of the paradigm function will permit us to treat the PM as a focus-seeking enclitic or as a verbal affix.

A salient property of the Udi PM's is that they appear as infixes with certain classes of verb stem. Harris (2002) offers a description of this patterning within Optimality Theory which relies on making a phonological segmentation of the verb stem and placing the PM phonologically. However, Harris' meticulous tracing of the historical development of the structure of the verb stem suggests an alternative way of looking at the Udi verb lexicon (see especially Harris 2002: 211f). Very few Udi verbs are monomorphemic and those there are often have a form (C)VC. The typical form of the light verb element is just a consonant. Moreover, the LV is typically a cranberry form, devoid of any meaning (indeed, this is often true of both parts of a complex verb stem). We propose, therefore, that the verb lexicon has undergone a type of reanalysis under which (nearly) all verbs are taken to be morphologically complex: IncE + LV. Thus, even a simplex verb stem such as *aq'* 'take' is treated as morphologically complex, which we will represent as a_0 + q'$_{II}$. The subscripts represent morpheme template positional slots (see Spencer 2003 for discussion of discontinuous stems in Athapaskan). (We stress that nothing of substance hinges on this interpretation being correct, but it seems to provide a straightforward account of the facts).

Assuming such a reanalysis simplifies the statement of the infixation process. Where the paradigm function is defined over a feature set containing [Default Focus: yes], the Domain/Host parameter is defined as the verb stem, and the exponence defines the PM as occupying slot I (with obvious abbreviations), whether slot I intervenes between an etymologically complex stem, as in (20), or between the components of a reanalysed stem, as in (21):

(20) PF for *lašk'oq'unbbesa* 'they get married'
PF(⟨LAŠK'O, {3plSubj, Pres, DefaultFocus}⟩) =

Domain:	verb
Host:	verb stem = lašk'o$_0$ + b$_{II}$
Exponence:	z$_I$, esa$_{III}$
Linearization:	default
Output:	lašk'o-q'un-b-esa

(21) PF for *azq'e* 'I received'
PF(⟨AQ', {1sgSubj, AorII, DefaultFocus}⟩) =

Domain:	verb
Host:	verb stem = a_0 + q'_{II}
Exponence:	z_I, e_{III}
Linearization:	default
Output:	a-z-q'-e

All we have presented here, of course, is the sketch of a preliminary analysis, or, more accurately, re-interpretation of Harris' analysis. There are various exceptional patterns to account for with default focus in which infixation isn't found, including the various Future and Subjunctive examples and a variety of other morphosyntactic contexts. However, each of these special contexts is indeed special, in the sense that the context has to be characterized by additional features and thus will pre-empt the default intra-verbstem placement outlined above. Thus, as far as we can tell, those exceptional subcases will not materially affect our argument. Our account successfully unites the phrasal and stem affixational properties of these unusual markers, but it does so by treating both clitics and affixes as morphological objects introduced by a paradigm function. Thus, what is odd about Udi is not that a syntactic object is allowed to violate lexical integrity. Rather, we have an instance of exactly the same patterning as that found in European Portuguese and a number of other languages, in which a clitic has been grammaticalized as an affix in some morphosyntactic contexts but has remained a clitic in other contexts. The Generalized Paradigm Function Morphology model, with its extended notion of paradigm function, permits us to capture the dual behaviour of a single morphological formative by providing a unified treatment of clitics and affixes which correctly reflects the essentially morphological nature of these formatives.

Bibliography

Anderson, Stephen R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Crysmann, Berthold. 2000. Syntactic transparency of pronominal affixes. In *Grammatical Interfaces in HPSG*, ed. by Ronnie Cann, Claire Grover, and Philip Miller, 77–96. Stanford, CA: CSLI Publications.

Harris, Alice C. 2000. Where in the word is the Udi clitic? *Language* 76.593–616.

—— 2002. *Endoclitics and the Origins of Udi Morphosyntax*. Oxford University Press.

Klavans, Judith. 1985. The independence of syntax and phonology in cliticization. *Language* 61.95–120.

Luís, Ana, 2004. *Clitics as Morphology*. Doctoral dissertation. University of Essex.

——, and Andrew Spencer. 2005. A Paradigm Function account of 'mesoclisis' in European Portuguese. In *Yearbook of Morphology 2004*, ed. by Geert Booij and Jaap van Marle, 79–154. Dortrecht: Springer.

McCarthy, John J., and Alan Prince. 2005. Generalized alignment. In *Yearbook of Morphology 2004*, ed. by Geert Booij and Jaap van Marle, 177–228. Dordrecht: Kluwer Academic Publishers.

Spencer, Andrew. 2003. Putting some order into morphology: Reflections on Rice (2000) and Stump (2001). *Journal of Linguistics* 39.621–646.

——. 2004. Generalized Paradigm function morphology – a synopsis. In *York Papers in Linguistics Series 2: Papers from the First York-Essex Morphology Meeting*, ed. by Alexandra Galani and Beck Sinar, 93–106. Department of Language Science and Linguistics, University of York.

——, ms. Generalized Paradigm Function Morphology. University of Essex.

Stump, Gregory T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Toivonen, Ida. 2003. *Non-Projecting Words: A Case Study of Swedish Particles*. Dordrecht: Kluwer Academic Publishers.

Zwicky, Arnold M., 1977. On clitics. Bloomington: Indiana University Linguistic Club.

——, and Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English *n't*. *Language* 59.502–513.

Ana R. Luís
Department of Anglo-American Studies
University of Coimbra
3004-530 Coimbra, Portugal
aluis@ci.uc.pt

Andrew Spencer
Department of Language and Linguistics
University of Essex
Wivenhoe Park, Colchester CO4 3SQ, UK
spena@essex.ac.uk

5 Inflection and Derivation in a Second Language*

Kathleen Neubauer (University of Essex)
Harald Clahsen (University of Essex)

5.1 Introduction

In the psycholinguistic literature there is an ongoing debate that centres around the representation of morphologically complex words in the mental lexicon. Here, the role of a word's morphological structure in online processing has been of particular interest. The issue under discussion is whether single mechanism models or dual mechanism models are more adequate in accounting for the facts.

Single mechanism models assume that all morphological variants of a word are stored separately in the mental lexicon. The morphological structure of words is posited to play no genuine role in language production and comprehension. Connectionists claim that morphological structure is not explicitly represented in the mental lexicon. Morphologically simple as well as complex words are said to be stored in terms of distributed representations in associative memory and have connections among them (e.g. Rumelhart and McClelland 1986, MacWhinney and Leinbach 1991).

On the other hand, so-called dual route models claim that both morphological decomposition and full-form representation are used by the language processor. Proponents of the Dual Mechanism Model have claimed that the mental grammar has a dual structure (e.g. Pinker 1999, Clahsen 1999. A qualitative distinction has been drawn between two distinct representational mechanisms: storage of lexical items in rote memory (e.g. irregular inflection is believed to involve memory-based associations) and computation of default transformations by symbol-manipulation processes (e.g. regular inflection is said to involve symbolic rules). Word frequency and the phonological transparency of words play an important role in determining the most efficient processing route (Ullman 1999, Clahsen 1999). In an extension of the dual mechanism model to derivational morphology Clahsen *et al.* (2002) treat both productive inflection and derivation as the result of combinatorial operations, but associate productive derivation and irregular inflection with stored lexical entries.

An interesting proposal with regard to morphological processing in L2 has been made by Ullman (2001a, b). He proposes the declarative/procedural model of lexicon and gram-

*We thank Ingrid Sonnenstuhl and the Psycholinguistic Research Group in particular Claudia Felser and Sonja Eisenbeiss for helpful comments and suggestions.

Essex Research Reports in Linguistics 47, 61–73.
Ryo Otoguro, Gergana Popova and Andrew Spencer (eds.).
Department of Language and Linguistics, University of Essex,
Colchester, United Kingdom.

mar which distinguishes between two brain memory systems: the so-called declarative memory and the procedural memory system. Both of these components are said to have distinct cognitive, computational and neural bases. The declarative memory is said to be an associative memory of distributed representations. In this memory system linguistic rules do not play any role. The procedural memory system can be classified as a rule system. It is claimed to subserve the fully productive morphological forms (defaults). The procedural memory system is thus said to underlie operations related to affixation. The declarative/procedural model posits a shift from procedural memory used in L1 to declarative memory in L2, i.e. it is predicted that L2 learners lexically memorize all complex word forms, including linguistic forms that are compositionally computed in L1. This hypothesis will be tested by investigating the processing of inflected and derived word forms by Chinese L2 learners of German.

5.2 Previous research

There is a considerably large body of research on morphological processing of inflected words in adult native speakers. Here, among the best studied phenomena are the English past tense and past participles. Previous research provides conflicting results.

An acceptability judgement task done by Kim *et al.* (1991) investigated the role of phonology, grammatical structure, and semantic extendedness in adult's generalisations of English past tense inflections. The subject's task was to rate regular and irregular past tense forms of denominal and extended endocentric verbs which were homophonous with irregular verbs. The results showed a preference for the regular past tense in the denominal condition, but a preference for irregular past tense forms for extended verbs. This was taken as evidence for the sensitivity of adult native speakers to formal grammatical structure of verbs. Marcus *et al.* (1995) investigated past participles in German. They employed a similar acceptability judgement task to investigate the generalization properties of a suffix in default circumstances. Denominal German verbs which were homophonous with existing strong verbs were presented in two conditions: as a denominal verb or as a verb root with an extended meaning depending on the context. The results showed a preference for the regularly inflected participles in the denominal condition but the irregular inflected participles were preferred when the verb's meaning was semantically extended. This was taken as evidence for the fact that when they inflect verbs German speakers take into account their morphological structure. Furthermore, it was shown that the *-t* suffix (regular) is used as a default when forming past participles.

Previous research seems to indicate that semantic transparency and productivity play an important role in morphological processing. In their study of derivational morphology in English Marslen-Wilson *et al.* (1994) investigated the processing of semantically transparent and opaque prefixed forms. They found evidence for morphological decomposition of semantically transparent forms. Sonnenstuhl and Clahsen (2003) studied derived word forms in German. Prefixed adjectives with the productive prefix *un-* and the less productive prefix *in-* were investigated in different experiments. Word-form frequency effects were found for both derivational forms which suggests full-form representations for fully productive and less productive prefixed forms. The results of the priming experiments

showed full priming for word forms prefixed with *un-* but reduced priming effects for *in-* forms. These results were taken as evidence for the decomposition of transparent and fully productive derivational forms but storage of derivational forms with limited productivity and transparency.

It has been argued that not only the degree of productivity but also possible phonological effects (neutral vs. non-neutral affixation) might influence various derivational processes. Alegre and Gordon (1999) investigated possible generalization effects of novel word forms in an acceptability judgement task. They found that the higher the similarity of novel words with non-neutral affixes to an existing word the higher its acceptability. This was not the case for novel words with neutral affixes. As far as adult L2-learners' morphological processing of complex words is concerned there is relatively little evidence available. Previous studies investigating the processing of inflectional morphology in L2-leaners have mainly concentrated on the formation of the English past tense. Consider, for example, Brovetto and Ullman (2001) who tested the predictions of the declarative/procedural model. They investigated the production of regular and irregular English past tense by Chinese and Spanish L2-learners of English. Frequency effects were found for both regular and irregular past tenses. This was interpreted in terms of memory-dependency in L2-speaker's processing of complex words. Note however, that this experiment did not contain enough filler items so that the subjects might have developed secondary strategies.

5.3 The present study

The main aim of the present study is to provide preliminary psycholinguistic evidence for the mental representation and the processing of morphologically complex words in advanced L2-learners of German. The main research questions to be addressed are the following:

- How are morphologically complex German word forms (inflected and derived words) processed and how are they stored in the L2-learner's mental lexicon?

- Is there a difference between how German native speakers and second language learners of German process and store morphologically complex words?

To address the above questions, two experiments investigate acceptability judgements of Chinese L2-learners of German and German native speakers when processing derived and inflected words. The main purpose of the experiments was to test for generalization properties of word formation processes. Evidence from previous acceptability judgement experiments supports the linguistic distinction between affixation and structured lexical entries. It has been shown that affixation-based processes can be applied to any kind of novel word whereas processes based on lexical entries can only be accessed by analogy and thus yield to similarity-based generalisations. Hence, similarity effects are taken as an indication for storage in associative memory.

5.4 Adjectival derivations with the prefixes *un-* vs. *in-*

The derivational affixes *un-* and *in-* in German are negative prefixes which form negated adjectives from their positive counterparts, for example: *bequem* – *unbequem* (‘comfortable- uncomfortable’), *akzeptabel* – *inakzeptabel* (‘acceptable- unacceptable’).

The prefix *un-* is of Germanic and the prefix *in-* is of Latinate origin. The prefixes *un-* and *in-* can be said to have the same semantic meaning: they both carry the meaning of simple negation. Both prefixes are bound negative affixes, i.e. they always attach to a base form. Derivations marked by one of these prefixes differ from each other with respect to their productivity and transparency. The prefix *un-* is highly productive and derivations marked by *un-* are phonologically transparent. In contrast, word forms derived by *in-* are usually of restricted phonological transparency and limited productivity. Further, the prefix *in-* shows phonologically conditioned variants such as *im-* before bilabials (e.g. *immobil* ‘immobile’), *il-* before /l/ such as *illegal* ‘illegal’ and *ir-* before /r/ (e.g. *irreal* ‘unreal’). This kind of assimilation of a nasal to a following consonant occurs in words prefixed with *in-* but not in words prefixed with *un-*. The prefixes *in-* and *un-* also show a distinct distribution pattern. Prefixation with *in-* is said to attach to Latinate adjectives, whereas *un-* is not restricted in that way. The negative *un-* can attach to simple and productively derived adjectives. Apart from very few exceptions, the prefix *un-* can substitute for *in-*, e.g. *inakzeptabel* and also *unakzeptabel*, but *in-* may not replace *un-*, i.e. **inbequem*. Note also that very frequent foreign adjectives are often prefixed with *un-*. (Duden 1995, Wolff 1984). The different distribution and properties of these two negative prefixes are often taken as evidence for the Level Ordering Hypothesis. With respect to their phonological properties Wiese (1996) categorises differently behaving affixes into class I and class II affixes. Class I affixes are said to either carry word stress and/or have the ability to influence the stress on their base. The prefix *in-* is classified as belonging to this class of affixes. As opposed to class I affixes, class II affixes never influence or attract word stress. Thus, Wiese treats the prefix *un-* as a class II affix. The different phonologically conditioned variants of *in-* are attributed to a nasal assimilation rule which is said to be only applicable to class I affixes but not to class II affixes.

Previous research on native speakers has shown that adjectival derivations with *un-* seem to be rule-based whereas derivations with *in-* are likely to be based on stored lexical entries (Sonnenstuhl and Clahsen 2003). In order to test for generalization properties of these two negative prefixes nonsense words which differed from each other with respect to their similarity to existing German adjectives were used. Properties of novel words are supposed to be ‘regular’, i.e. the default operation should be applied (e.g. Kilbury *et al.* 1992). Hence, using novel adjectives generalization properties of the derivational prefixes *un-* and *in-* as applied by Chinese L2-learners of German have been tested.

5.5 Method

5.5.1 Participants

22 Chinese learners of German (11 female, mean age 26.7years) voluntarily participated in the experiment. None of the Chinese L2-learners had learned German before the age of 17. Most Chinese L2 learners had learned basic German in their home country and all of them had had formal instruction in German at a German university. At the time of the experiment they were all studying at a university in Germany. The general level of proficiency in German of all L2-learners was assessed prior to the main experiments using different subtests of a standardised proficiency test (Allgemeiner Deutscher Sprachtest, Steinert 1978). With a mean test result of 70.4% the Chinese L2 learners can be considered as intermediate to advanced learners of German. In addition a control group of 22 German native speakers (11 female, mean age 27.9 years) also participated in the experiment.

5.5.2 Materials and Design

A modified German version of Alegre and Gordon's (1999) acceptability judgement experiment was used to see how the two prefixes *un-* and *in-* generalize to novel word forms. The novel adjectives used in this experiment differed in terms of their (phonological/orthographical) similarity to existing German adjectives. There were three types: novel adjectives which were not similar to any German adjective (non-rhyme condition), novel adjectives which were similar to existing German adjectives of Germanic origin which can only be negated using the prefix *un-* (Germanic rhyme condition). The Latinate rhyme condition involved novel words similar to existing adjectives of Latinate origin that can be prefixed with *un-* as well as with *in-*. The questionnaire materials included 36 experimental items, i.e. 12 experimental items per condition. All nonsense words had 2 syllables and contained phonotactically permitted sequences in German. To create nonce adjectives for the Latinate and Germanic rhyme conditions existing German adjectives of Latinate and Germanic origin were pair-wise matched for frequency of their base form and then only the onset of each word was changed to ensure a high similarity to the existing word.

Participants were presented with pairs of novel base and prefixed word forms. Each novel word was embedded in a simple context sentence. Each test item contained a context sentence presenting the base form of the nonsense adjective followed by two test sentences which were identical except that one contained the nonce word prefixed with *un-* and the other with *in-*. To reduce uncontrolled (semantic or formal) association the noun modified by the test adjective was a 1-syllable nonsense word. An example for a test item (including an English translation) is given in (1) below:

(1) a. Anna hat heute einen narseken Fneik gekauft.
'Anna bought a narsek Fneik today.'

b. Ihr alter Fneik ist nämlich schon ziemlich *innarsek*/*unnarsek*.
'Her old Fneik was already pretty *innarsek*/*unnarsek*'

The subject's task was to rate the naturalness of each derived form prefixed with *in-* and *un-* independently from each other on a 1-5 point scale with "1" meaning "very unnatural sounding" and "5" meaning "very natural sounding".

5.5.3 Predictions

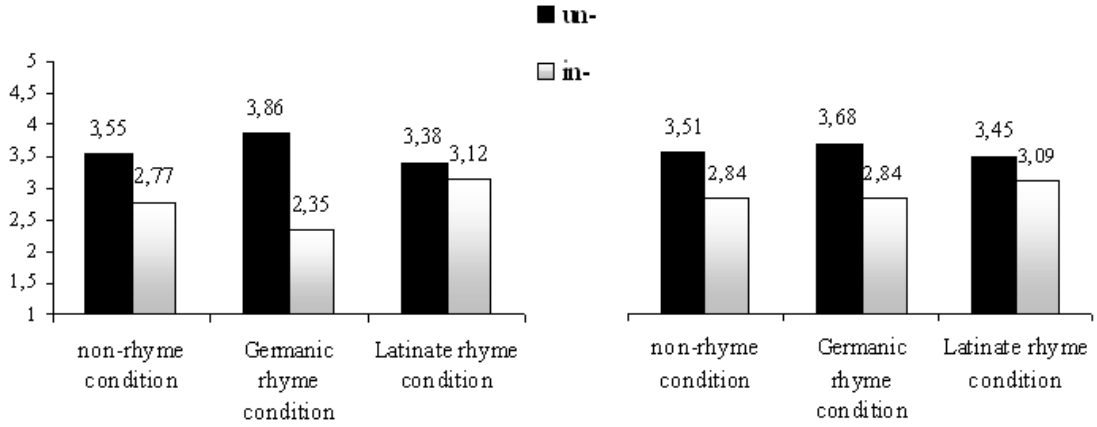
According to the dual mechanism model non-productive and non-transparent word forms are believed to be stored as wholes in the mental lexicon. Thus, for native speakers of German one might expect to find the strongest similarity effects for the less productive adjectival derivations with the prefix *in-*. For the non-rhyme condition it is predicted that native speakers prefer derivations with the prefix *un-*, since productive processes are believed to generalize to novel words. For the Germanic rhyme condition there should also be a preference for *un*-derivations. For the Latinate rhyme condition higher ratings for *in-* forms relative to the other two conditions are expected.

The declarative/procedural model claims that in late L2 acquisition the learner mainly relies on (declarative) memory. Thus, one would expect this memory dependence to show in the L2 learners' acceptability ratings. If, on the other hand, all words are stored in the L2-learner's mental lexicon and they can be accessed via analogical extension to an existing word a preference of *un*-derivations should be found in the Germanic rhyme condition.

5.5.4 Results and discussion

The mean ratings for all three conditions per language group are shown in Figure 1.

Figure 1: mean ratings per condition and language group



L1 speakers L2 learners

As Figure 1 shows, both native speakers and L2 learners of German preferred derivations with the prefix *un-* over the derived forms with *in-* in all three conditions. The L1 speakers' graph shows that word type has some effect on their acceptability ratings. This seems not to be the case for the L2 learners. The L2 learners' acceptability ratings for nonce adjectives prefixed with *un-* and *in-* are quite stable across conditions. The

main difference between the two language groups can be found in the Germanic rhyme condition. Statistically, there was a significant Word-Type × Prefix × Language Group interaction ($F(2, 42) = 5.28$, $p = 0.014$). Furthermore, there was a significant main effect of Prefix ($F(1, 42) = 26.23$, $p < 0.001$) and a main effect of Language Group ($F(1, 42) = 0.26$, $p = 0.06$). Looking at both language groups separately, a significant 2-way interaction of Word-type × Prefix ($F(2, 42) = 29.86$, $p < 0.001$) was found for the L1 speakers. The L2 learners showed a significant main effect of Prefix ($F(1, 21) = 6.36$, $p = 0.02$), but no interaction.

Overall, both native speakers and L2-learners of German showed a significant preference for negative adjectival derivations affixed with the Germanic prefix *un-* over derivations with the Latinate prefix *in-*. This preference is likely to be due to the higher frequency of the prefix *un-*. Note that the prefix *un-* is among the most frequent affixes in the language, that it is highly productive, and that it applies more widely in the language than the negative prefix *in-* (e.g. Duden 1995, Schnerrer 1982).

In addition, the Word-Type × Prefix interaction seen for the native speaker group shows that they are sensitive to the distributional differences between *un-* and *in-* forms. The absence of this interaction in the L2 group shows that the L2 learners seem to be insensitive to the differences between Latinate and Germanic words with respect to *in-* and *un-* prefixation.

5.6 Regular and irregular participles

Regular participles involve a *-t* suffix and no vowel change in the stem. The regular *-t* is said to serve as the default form and thus should apply to any kind of novel verb. In contrast, irregular participles often involve a stem change and are always suffixed with *-(e)n*. Irregular participle forms are largely unpredictable. A particularly interesting phenomenon for the distinction between regular and irregular inflection is the inflection of denominal verbs. Denominal verbs are always inflected following the regular pattern, even if the denominal verb is homophonous with an irregular verb. This is because irregularity is a property of verb roots. Hence, a noun root cannot have an irregular past tense or participle form associated with it (Kim *et al.* 1994, Marcus *et al.* 1995, Wunderlich and Fabri 1995).

Denominal verbs do not have lexical entries as verbs but rather involve category-changing affixation. When forming the past tense or a participle for these verbs access to lexical entries of verbs is blocked, even though they might sound similar to existing verbs. Denominal verbs form a particularly interesting case for studying L2 morphological processing because they are extremely infrequent in the language input. Also, the denominal rule is usually not explicitly taught in language classes (Marcus *et al.* 1995, Bandi-Rao 2002).

5.6.1 Predictions

With respect to the L1 speakers we should replicate the results of previous studies (e.g. Marcus *et al.* 1995). Hence, a preference for regular participle forms for denominal verbs,

but a strong preference for irregular participle forms if the verbs have an extended meaning to their usual senses would be expected.

If the Chinese L2 learners are aware of the morphological structure of complex derived verbs one would predict to find similar effects as for the native speakers. But if L2 learners are not sensitive to the grammatical structure of words and all regular and irregular morphologically complex word forms are memorized in the mental lexicon, as predicted by Ullman's declarative/procedural model, a preference for irregular participle forms in both the denominal verb condition and the extended verb condition would be predicted, as all items are constructed in analogy to existing irregular verbs.

5.6.2 Participants

The participants were the same as in the experiment presented previously. The Chinese L2-learners underwent a vocabulary pre-test to test their knowledge of the verbs to be used in the experiment. According to these pre-test results 21.8% (144 out of 660 answers) of the total data of the main experiment had to be excluded because only verbs which were known to the participants were taken into consideration.

5.6.3 Materials and Design

A modified version of Marcus *et al.*'s (1995) acceptability judgement experiment was used. This paper-and-pencil test contained 15 experimental items. 15 denominal German verbs which were homophonous with existing strong verbs in German were chosen. Each test item consisted of a context paragraph and two test sentences. The grammatical structure of the verb used was indicated to the participants in the context paragraph. In the context paragraph the word was either presented as the noun or as a verb with a semantically extended meaning of the usual sense of the verb. In the test sentences the word was used as the root of a prefixed verb in its participial form. The two test sentences were identical except that at the end of the sentence one test sentence contained the regular participle, e.g. formed with a *t*-suffix and no stem change and the other test sentence contained the irregular participle form with the *-(e)n* suffix and with a stem change. All sentences were simplified lexically and structurally to ensure an easy understanding by the L2 learners. Two examples of used test items including translations are given in (2a and 2b):

(2) a. denominal verb condition (meaning from Fliege ('fly')):

Der Insektenfreund Mark Möller sammelt verschiedenste Arten von Fliegen. Er bekommt ständig neue Fliegenexemplare von Freunden und Bekannten. Möller klebt alle toten Fliegen in sein Insektenbuch ein.
Gerade hat er wieder eine neue Seite in seinem Buch *befliegt*/*beflogen*.
('The insect lover Marc Miller collects different kinds of flies. He constantly gets more flies from friends. Miller puts all dead flies in his books of insects. Just now he has *be-flied*/*be-flown* another page of his book.')

b. extended verb condition (meaning from fliegen ('to fly')):

> Die Sportskanone Sven ist leidenschaftlicher Surfer. Jeden Sommerurlaub verbringt er an einem anderen Strand, um mit seinem Surfbord gegen den Wind die Meereswellen zu befliegen. Dieses Jahr versucht er sich an der Küste Mexikos.
> Im Laufe der Jahre hat er schon viele Wellen mit seinem Surfbord *beflogen*/*befliegt*.
> ('The sportsman Sven is a passionate surfer. He spends every summer holiday on a different beach in order to windsurf with his surfboard against the sea waves. This year he is having a try at the Mexican coast.
> Through the years he has already *be-flied*/*be-flown* many waves with his surfboard.')

Each subject saw a verb either as a denominal verb or as a verb with a semantic extension of its usual sense. Each of the four test versions contained about half of the verbs as denominal verbs and the other half as extended verb roots. Also, each participant received half of the test sentences with the regular verb form before the irregular one and the other half with the irregular verb form before the regular form. In order to avoid subjects to develop a strategy when rating the experimental items, 15 filler items were also added resulting in a total of 30 test items.

The filler items included existing regular and irregular German verbs used in their standard senses. The formal presentation of the filler items was identical to the experimental ones. The 30 test items were presented in a pseudo-randomized order. The subjects were asked to first read the context paragraph very carefully. Then the subjects' task was to rate the naturalness of each of the two participle forms per test item on a 1-5 point scale. The rating scale was explained to the participants with "1" meaning "very unnatural sounding" and "5" meaning "very natural sounding". The participants were further instructed to pay special attention to the context paragraph and to base their ratings on this context. They were also asked to judge the two test sentences independently.

5.6.4 Results and discussion

Figure 2: mean ratings for both conditions per language group

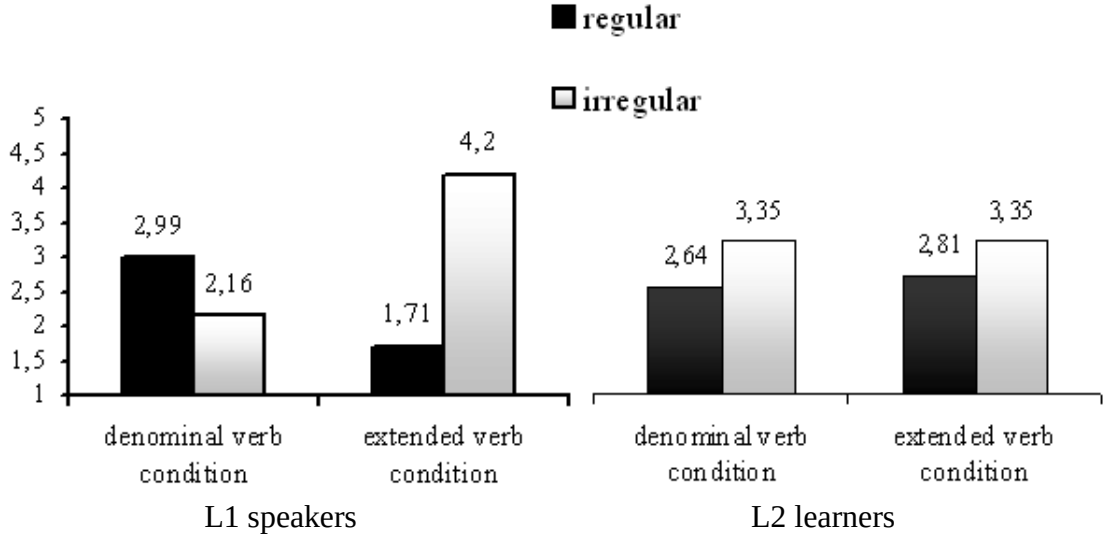


Figure 2 shows clear contrasts between the L1 speakers and L2 learners. Native speakers rated the regular participle forms higher than the irregular ones in the denominal condition, but lower in the extended verb condition. In contrast, the Chinese L2 learners rated the irregular participle forms higher than the regular forms in both conditions. There was a significant interaction of Verb-Type × Participle Form × Language Group ($F(1, 42) = 51.42$, $p < 0.001$) as well as a significant main effect of Language Group ($F(1.42) = 4.37$, $p = 0.043$) which confirm that L2 learners and native speakers of German exhibit a distinct pattern in their acceptability ratings.

Moreover, there was a significant interaction of Verb Type × Participle Form ($F(1, 21) = 76.39$, $p < 0.001$) for the L1 group which suggests that their acceptability ratings for both participial forms depend on the type of verb presented to them. They gave preference to the regular participle forms for verbs that are derived from nouns even when their homophonous verb root counterparts usually take irregular participle forms. In contrast, for the Chinese L2 learners the type of verb presented seems to have very little effect on their ratings for regular participles and absolutely no effect on ratings for irregular participle forms. This was confirmed statistically by a main effect of Participle Form ($F(1, 21) = 6.45$, $p = 0.019$).

5.7 Summary and Conclusion

To sum up the results, the experiment investigating derivational morphology found a frequency effect for the prefix *un-* for both native speakers and L2 learners of German. Furthermore, a similarity effect was found for the native speakers but not for the Chinese L2 learners. Native speakers but not L2 learners seem to be sensitive to the word type and prefix distribution. The degree of similarity of a nonce word to existing words does not seem to play an important role for the L2 learners. This might be due to the fact that in the L2 learners' native language (Chinese) there is no distinction between words of Germanic and words of Latinate origin. The L2 learners might not be too familiar with this distinction and its consequences for the use of the different prefixes.

The results of the second experiment on inflectional morphology show that native speakers of German are, as expected, sensitive to the morphological structure of verbs. Thus, the results from previous experiments investigating the treatment of denominal verbs by native speakers could be replicated.

This stands in sharp contrast to the results from the Chinese L2 learners. The L2 learners treated both verbs that appeared in a semantically extended meaning and denominal verbs the same and preferred irregular participle forms in both cases. The observed similarity-based generalizations to existing irregular verbs in German might be an indication of storage of these verbs in the L2 learner's mental lexicon and the L2 learner's memory dependency when performing the experimental task. The results might also mean that the Chinese L2 learners are not sensitive to the morphological structure of these verbs, i.e. they could be unaware of the fact that the verbs used in the denominal condition are verbs which are (transparently) based on nouns. But further research is needed to determine why the L2 learners responded as they did in this experiment. Do L2 learners know the denominal constraint or not? Do they have a regular rule available or do they store regular

and irregular word forms in the mental lexicon as predicted by the declarative/procedural model?

Overall, the results of both experiments have revealed differences between German native speakers and Chinese L2 learners of German. L1 and L2 speakers exhibited a distinct pattern of behaviour in both experiments.

Bibliography

ALEGRE, M., and P. GORDON. 1999. Rule-based versus associative processes in derivational morphology. *Brain and Language* 68.347–354.

BANDI-RAO, S. 2002. Inflecting denominal verbs – the role of semantics. In *Proceedings of the BUCLD 26*, 48–59. Boston University.

BROVETTO, C., and M. T. ULLMAN. 2001. First vs. second language: A differential reliance on grammatical computations and lexical memory. In *CUNY 2001 Conference on Sentence Processing*.

CLAHSEN, HARALD. 1999. Lexical entries and the rules of language: A multidisciplinary study of German inflection. *Behavioural and Brain Science* 22.991–1060.

——, I. SONNENSTUHL, and J. P. BLEVINS. 2002. Derivational morphology in the german mental lexicon. In *Morphological Structure in Language Processing*, ed. by H. Baayen and R. Schreuder, 125–155. Mouton de Gruyter.

Duden, 1995. Duden, Grammatik der deutschen Gegenwartssprache. Hrsg. und bearb. von G. Drosdowski in Zusammenarbeit mit P. Eisenberg et al.

KILBURY, J., P. NAEGER, and I. RENZ. 1992. *New Lexical Entries for Unknown Words*. Theorie des Lexikons. Arbeiten des SFB 282. Düsseldorf: H. Heine Universität.

KIM, J. J., G. F. MARCUS, S. PINKER, M. HOLLANDER, and M. COPPOLA. 1994. Sensitivity of children's inflection to grammatical structure. *Journal of Child Language* 21.172–209.

——, S. PINKER, A. S. PRINCE, and S. PRASADA. 1991. Why no mere mortal has ever flown out to centre field. *Cognitive Science* 15.173–218.

MACWHINNEY, B., and J. LEINBACH. 1991. Implementations are not conceptualization: Revising the verb learning mode. *Cognition* 40.121–157.

MARCUS, G. F., U. BRINKMANN, H. CLAHSEN, R. WIESE, and S. PINKER. 1995. German inflection: the exception that proves the rule. *Cognitive Psychology* 29.189–256.

MARSLEN-WILSON, W., L. K. TYLER, R. WAKSLER, and L. OLDER. 1994. Morphology and meaning in the English mental lexicon. *Psychological Review* 101.3–33.

PINKER, S. 1999. *Words and Rules*. New York, NY: Basic Books.

RUMELHART, D. E., and J. L. MCCLELLAND. 1986. On learning the past tense of English verbs. In *Parallel Distributed Processing*, ed. by J. L. McClelland, D. E. Rumelhart, and the PDP Research Group, volume 2, 216–271. Cambridge, MA: MIT Press.

SCHNERRER, R. 1982. Funktionen des Wortbildungsmorphems un- in der deutschen Gegenwartssprache. *Beiträge zur Erforschung der deutschen Sprache* 2.22–51.

SONNENSTUHL, I., and H. CLAHSEN, 2003. Prefixed word forms in the mental lexicon. Unpublished manuscript. University of Essex.

STEINERT, J. 1978. *Allgemeiner Deutscher Sprachtest*. Braunschweig: Westermann.

ULLMAN, M. T. 1999. Acceptability ratings of regular and irregular past-tense forms: Evidence for a dual-system model of language from word frequency and phonological neighbourhood effects. *Language and Cognitive Processes* 14.47–67.

—— 2001a. The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research* 30.37–69.

—— 2001b. The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism* . Special Issue: "Neurolinguistics of Bilingualism".

WIESE, R. 1996. *The Phonology of German*. Oxford: Clarendon Press.

WOLFF, S. 1984. *Lexical entries and word formation*. Bloomington, Ind.: Indiana University Linguistics Club.

WUNDERLICH, D., and R. FABRI. 1995. Minimalist morphology: An approach to inflection. *Zeitschrift für Sprachwissenschaft* 14.236–294.

Appendix

Experimental Material: Adjectival derivations with *un-* and *in-*
List of nonsense adjectives used:

- *Non-rhyme condition:*
  utab, pomset, fistur, kenvar, hurdaf, predos, plerim, spirenk, blensir, gamlis, narsek

- *Germanic rhyme condition:*
  kewusst, sunkel, peiter, bauber, nequem, zeise, fauter, peikel, kieder, fetreu, kanger, prage

- *Latinate rhyme condition:*
  pirekt, sonkret, seal, somplett, nerfekt, fabil, sporrekt, gormell, fiskret, pegal, suman, hobil

Experimental Material: German participles

List of denominal verbs and their meanings as used in the experiment

verspinnen	cover with spyders (from *Spinne* 'spider')	make something up in the wrong way (from *spinnen*)
besitzen	supply with seats (from *Sitz-* 'seat')	sit around on (from *sitzen-* 'to sit')
befliegen	put flies onto (from *Fliege-* 'fly')	windsurf
verblasen	cover with blisters (from *Blase-* 'blister')	use up by blowing (from *blasen-* 'to blow')
verklingen	put knew knife blade on (from *Klinge-* 'blade')	play to point of breakdown
bescheinen	supply with documents (from *Schein-* 'document')	shine onto
bepfeifen	put pipes onto (from *Pfeife-* 'pipe')	whistle for someone (from *pfeifen-* 'to whistle')
bereiben	supply with graters (from *Reibe-* 'grater')	rub around on (from *reiben-* 'to grate')
verscheren	cover with scissors (from *Schere-* 'scissors')	trim in the wrong way (from *scheren-* 'to cut/trim')
verschlingen	put bandage on (from *Schlinge-* '…')	gulp (from *schlingen-* 'to gulp')
vertragen	cover with stretchers (from *Trage-* 'stretcher')	ruin by carrying too much (from *tragen-* 'to carry')
verwiegen	cover with cradles (from *Wiege* 'cradle')	weight wrongly (from *wiegen* 'to weight')
befallen	supply with traps (from *Falle* 'trap')	infest
beliegen	supply with cots (from *Liege* 'cot')	lie around/down on
verwachsen	cover with wax (from *Wachs* 'wax')	grow in the wrong way

Kathleen Neubauer
Department of Language and Linguistics
University of Essex
Wivenhoe Park, Colchester CO4 3SQ, UK
kneuba@essex.ac.uk

Harald Clahsen
Department of Language and Linguistics
University of Essex
Wivenhoe Park, Colchester CO4 3SQ, UK
harald@essex.ac.uk

6 Integrating Nominalisations into a (Generalised) Paradigm Function Model Moprhology*

Gergana Popova (University of Essex)

6.1 Introduction

The aim of this paper is to explore some properties of nominalisations using Bulgarian data, and to propose a tentative formalisation of the morphological mechanisms involved within the framework of Paradigm Function Morphology (a model of morphology which stems from the work of Matthews (1972), Anderson (1992), Aronoff (1994), and is very thoroughly formalised in Stump (2001)).

The paper focuses on the difference between productive deverbal nominals with event semantics and verbal argument structure on the one hand, and other nouns on the other. Nouns with event semantics are often called event nouns, or complex event nominals following Grimshaw (1990). Morphologically, they are transpositions. The conceptual importance of transpositions has been discussed in, for example, Beard (1995) and Spencer (1999). Transpositions challenge any morphological model because they have features that are traditionally said to belong to inflection and features that have been considered a central property of derivation. So, for example, Bulgarian deverbal nouns with the suffix *-ne* derived via transposition head a noun phrase, and take noun-related morphological categories like gender and definiteness, but at the same time display event semantics and have the same meaning as the verb they derive from. These nouns also appear to take the same arguments as the verbs they derive from, which makes them more like verbs and less like other nouns. The formation of these nouns is regular and predictable, which is why traditional descriptions include them in the verbal paradigm. It is difficult to classify the formation of these nouns with inflectional phenomena, however, as it involves a change in syntactic category, which is arguably a property typical of derivational processes.

A second aim of this paper is to describe (some of) the properties of complex event nominals (CENs) in Bulgarian. Bulgarian CENs have not been described in detail in modern linguistic frameworks, and it is of course difficult to be exhaustive within the limited scope of this work. The focus here is on nominals derived via the suffix *-ne* (almost exclusively CENs), which are action nominals. There is work available on cognate

*I am grateful to Bas Aarts, Andrew Spencer and Ryo Otoguro for valuable comments and suggestions, as well as to the audience at the second YEMM meeting. This research has been supported by an ORS grant ORS/2000015006 and a University of Essex studentship.

Essex Research Reports in Linguistics 47, 75–93.
Ryo Otoguro, Gergana Popova and Andrew Spencer (eds.).
Department of Language and Linguistics, University of Essex,
Colchester, United Kingdom.

nominalisations in other Slavic languages (Polish and Russian) which will be drawn upon. Hopefully, this paper can serve as a starting point for more thorough comparisons between the Slavic languages.

The paper is organised as follows: the next section gives a brief overview of the mechanisms available in Bulgarian for deriving nouns from verbs. Section 6.3 looks in more detail at the distinction between complex event nominals (CENs), simple-event nominals (SENs), and result nominals (RN), which originates from the work of Grimshaw (1990) on English and establishes the validity of this distinction for Bulgarian. It also looks more specifically at the properties of Bulgarian CENs, and explores their argument taking properties. Sections 6.4 and 6.5 look at the interaction of these argument taking properties and the categories of aspect and lexical class. Section 6.6 offers an account of nominalisations within the framework of Paradigm Function Morphology (Stump 2001), making use of some proposals to extend the framework (Spencer 2004) and introducing some other modifications. A brief description of some of the formal mechanisms used is presented in section 6.6.1. The final section is a summary, and points to directions for further research.

6.2 Nominalisations in Bulgarian

The most productive way to derive nouns from verbs in Bulgarian is suffixation of the suffix *-ne* to an imperfective verb. (Descriptions of the morphology of noun derivation can be found in Tilkov *et al.* (1983) and Radeva (1991), Stojanov (1977)). In part due to its productivity and generality, this process has been considered part of the verbal paradigm in traditional descriptions of the language. Derivation of *ne*-nominalisations is illustrated in the table below.

	Underlying verb		Deverbal nominal	
1	piša	'write.IMP'	pisa-ne	'writing'
2	zamest-va-m	'act as deputy.IMP'	zamest-va-ne	'acting as deputy'
3	zamestja	'act as deputy.PERF'	*zamest-ne	
4	prekâs-va-m	'interrupt.IMP'	prekâs-va-ne	'interruption'
5	prevključ-va-m	'switch.IMP'	prevključ-va-ne	'switching'
6	otpâtu-va-m	'depart'	otpâtu-va-ne	'departure'
7	transform-ira-m	'transform.IMP'	transform-ira-ne	'transformation'

Bulgarian is a language with a highly grammaticalised category of aspect with two values: perfective and imperfective. Some imperfective verbs are morphologically non-derived, like the verb *piša* 'write' in the first row of the table[1]. Non-derived imperfective verbs derive nominalisations with *-ne* like all other imperfective verbs. In the majority of cases an imperfective verb is derived from a perfective one via suffixation with the suffix *-va-* or one of its allomorphs (*-uva-, -ava- (-'ava-), -a- (-'a-)*). As a result, a vast number of pairs of verbs exists in the language which differ only with respect to their aspectual values. Only the imperfective members of these pairs allow derivation of *ne*-nominals, as

[1] Suffixation with *-ne* is said to be based on the aorist stem of the verb. This explains the difference in the shape of the verbal stem and the noun stem in this example. The morphophonological alternations involved will largely be ignored, since they are not important to the points made in the paper.

illustrated in rows 2 and 3 of the table. Further examples of *ne*-nominalisations based on derived imperfective verbs are given in rows 4 and 5. One exception to the imperfective-base-only regularity is noted in Stojanov (1966). He brings to attention verbs like *otpâtuvam* 'depart' (see row 6) as examples of perfective verbs that allow *ne*-nominalisation. Without going into the details of the linguistic analysis, suffice it to say that although Stojanov considers these verbs to be perfective, many grammarians and lexicographers point out that they are really biaspectual (references can be found in Stojanov's work). Biaspectual verbs in general allow nominalisation with *-ne*, as is shown in row 7 with an uncontroversially biaspectual verb.

Very similar to the suffix *-ne* is the suffix *-nie,* indeed the two have a common historical origin. Forms derived with *-nie* have either survived from earlier stages of the development of the language, or have been borrowed from Russian, but are no longer productively derived (see Tilkov *et al.* (1983)). Stojanov (1977) notes that the nouns in *-nie* are formed from perfective as well as from imperfective verbs, for example *sâbera* (PERF) 'gather' – *sâbra-nie* 'gathering';[2] *nakaža* (PERF) 'punish' – *nakaza-nie* 'punishment', *stradam* (IMP) 'suffer' – *strada-nie* 'suffering'. Often there are pairs of verbs, one formed with *-ne* and one with *-nie* on (lexically) identical stems. Usually, the verb with *-nie* is a result verb and the verb with *-ne* is a process verb. Nominals in *-nie* however are far fewer and are often lexicalised, so such pairs are not a ubiquitous phenomenon.

There are a number of other nominalisation patterns available to the language, some of which will be illustrated briefly in later sections.

6.3 Complex event nominals, simple event nominals and result nominals in Bulgarian

Grimshaw (1990), in a study of argument structure and nominalisations, draws a distinction between nominals that have argument structure (complex event nominals or CENs) and nominals that don't (simple event nominals and result nominals), which she correlates with the presence or absence of verbal structure respectively. Bulgarian confirms the distinction between complex event nominals on the one hand and simple event nominals and result nominals on the other. The tests proposed by Grimshaw have been applied to Russian in Schoorlemmer (1995) and Bredenkamp *et al.* (1998), and for Polish in Rozwadowska (1997). The same tests can be applied successfully to Bulgarian data as well, and identify the following differences between CENs and other nominalisations:

- CENs ((1)), but not result nominals ((2)) can be modified by phase verbs like *započvam* 'begin' or *prodalžavam* 'continue'

  (1) Izrazjavaneto na čuvstvata mu započna predi dva dni.
  expression of feelings his started before two days
  'His expressing his feelings started two days ago'.

[2] Again, morphophonological differences are due to the fact that *-nie* takes the aorist stem.

(2) *Izraženieto na liceto j prodalži dva časa.
expression on face her continued two hours
'The expression on her face lasted for two hours.'

- CENs ((3)), but not result nominals ((4)) can be modified with durative and completive adverbials like *v prodalženie na dva dni* 'for two days' or *za dva dni* 'in two days':

(3) Sreštaneto s čuždenci v prodalženie na dva dni go iztošti.
meeting with foreigners in duration of two days him exhausted
'Meeting foreigners for two days exhausted him'.

(4) *Sreštata s čuždenci v prodalženie na dva dni go iztošti.
meeting with foreigners in duration of two days him exhausted
'Meeting foreigners for two days exhausted him.'

- CENs, like the verbs from which they derive, allow modification for manner (see (5) and (6)):

(5) Toj piše pisma spokojno i uvereno.
he writes letters calmly and confidently
'He writes letters calmly and confidently'.

(6) Spokojnoto i uvereno pisane na pisma mu pomaga.
calm.THE and confident writing of letters him help
'The calm and confident writing of letters helps him.'

- CENs allow modification with adjectives like *frequent* ((7)) and *permanent* ((8)). The nominals modified in this way can be interpreted to denote iterated events without themselves changing their singular number. Result nominals, if they allow such modification at all, need to appear in the plural (see (9)).

(7) Čestoto čukane go iznervi.
Frequent.THE knocking him nervous made
'The frequent knocking made him nervous'.

(8) Postojannoto zvânene beše neprijatno.
Constant.THE ringing was unpleasant
'The constant ringing was unpleasant.'

(9) Čestite udari po vratata go iznervixa.
Frequent.PL.THE knocks on wall-THE him nervous made
'The frequent knocks made him nervous'.

- CENs allow for event control ((10)), whereas result nominals don't ((11)).

(10) Nalaga se sâbiraneto na sobstvenicite za da se reši problema
demanded REFL gathering of owners for to REFL solve problen
s pokriva.
with roof

'The gathering of the owners in order to solve the problem with the roof is mandatory'.

(11) *Nalaga se sâbranieto na sobstvenicite za da se reši problema
demanded REFL gathering of owners for to REFL solve problen
s pokriva.
with roof

'The gathering of the owners in order to solve the problem with the roof is mandatory'.

The differences between CENs and other nominals described in this section relate to the fact that the former have event semantics, while the latter don't. These differences are further correlated with the fact that CENs take arguments, whereas other nominals don't. The next section is devoted to the argument taking properties of CENs.

6.3.1 The argument structure of complex event nominals in Bulgarian

According to Grimshaw (1990), the argument taking property of CENs is determined by their ability to suppress the external argument of the verb from which they derive. This would predict that nominalisation should be possible from transitive or passivised verbs only. It has been shown in Schoorlemmer (1995) for Russian and in Rozwadowska (1997) for Polish that CENs from intransitive verbs are also possible in these languages, which contradicts the theory that nominalisations are akin to passivisation. In Bulgarian too CENs are freely derivable from intransitive verbs, and in that case the external argument is not suppressed, though its syntactic expression changes.

Bulgarian nouns (result nouns, as well as CENs) are most typically followed by a prepositional phrase headed by the preposition *na*, which can receive a variety of interpretations, similar to the Polish genitive postnominal NP described in Rozwadowska (1997). For example, in (12) below John is the owner of the car; in (13) John can be the one who gives the present, or its intended recipient; in (14) John might be a member of the team, or a fan, or the owner of the team. In intransitive CE nominals the sole argument gets mapped onto the *na*-PP, as is shown in examples (15) and (16) below:

(12) Kolata na Ivan e v garaža.
Car of John is in garage-THE

'John's car is in the garage.'

(13) Podarakât na Ivan e mnogo xubav.
Present.THE of John is very nice

'John's present is very nice.'

(14) Otbor*ât* na Ivan e na pârvo mjasto.
team.THE of John is on first place
'John's team is first'.

(15) Pristiganeto na Marija v osem časa obârka planovete ni.
Arrival.THE of Maria at eight o'clock spoiled plans ours
'Maria's arrival at eight o'clock spoiled our plans.'

(16) Spaneto na Ivan prodâlži osem časa.
Sleep of John lasted eight hours
'John's sleep lasted eight hours.'

In CENs derived from transitive verbs the linking of arguments to the syntactic positions available is as follows: the nominal may express all the arguments of the verb (noted also in Steinke (1999)), in which case the object of the verb is mapped onto the first postnominal *na*-PP, the recipient or the oblique is mapped onto the second *na*-PP and the subject or the agent is mapped onto the *ot*-PP, which is the Bulgarian equivalent of the English *by*-phrase. This is illustrated in (18) below (the verbal equivalent is in (17)). On the other hand, the nominal may choose not to express all the arguments available to it, in which case the following configurations are possible: (i) the direct object may be expressed, as well as the recipient or oblique, but not the subject (see (19)), (ii) or the object and the subject might be expressed, but not the oblique, as in (20), or (iii) the object may be expressed, but neither the oblique nor the subject as in (21). What is not possible is to express the oblique only as in (22), or the subject only as in (23), or to try to express the subject as a *na*-PP ((24) is ungrammatical on the intended reading of John as the agent).

(17) Ivan podari knigata na Petâr.
John gave (as present) book to Peter
'John gave Peter the book as a present.'

(18) Podarjavaneto na knigata ot Ivan na Petâr beše iznenada.
Giving of book by Ivan to Peter was surprise
'John's giving the book to Peter as a present was a surprise.'

(19) Podarjavaneto na knigata na Petâr beše iznenada.
Giving of book to Peter was surprise
'Giving the book to Peter was a surprise'.

(20) Podarjavaneto na knigata ot Ivan beše iznenada.
Giving of book by Ivan was surprise
'John's giving the book as a present was a surprise.'

(21) Podarjavaneto na knigata beše iznenada.
Giving of book was surprise
'The giving of the book was a surprise.'

(22) *Podarjavaneto na Petâr beše iznenada.
Giving to Peter was surprise
'Giving to Peter was a surprise'.

(23) *Podarjavaneto ot Ivan beše iznenada.
Giving by Ivan was surprise
'John's giving was a surprise.'

(24) *Podarjavaneto na knigata na Ivan beše iznenada.
Giving of book by John was surprise
'(intended) John's giving the book was a surprise.'

These data indicate that the CE nominalisations take one real complement (marked in Bulgarian with the preposition *na*), since minimally they need this one complement in order to participate in grammatical constructions. Schoorlemmer (1995) also acknowledges one real argument for CE nominalisations, which on her analysis is assigned structural case by the noun. In her work this explains why nominalisations of intransitive verbs map the subject of the underlying verb onto this position, instead of licensing a *by*-phrase. However, the assumption of a structural case does not in itself explain why in nominalisations of transitive verbs it is not the subject, but the object that is mapped onto this argument position, and why the oblique never gets assigned to it.

Another possible explanation for the patterns observed above, as was mentioned already, is to appeal to a demotion of the subject in a process akin to passivisation. More generally, the historical facts suggest that there is a close relationship between the *ne*-nominalisation and the past passive participle. (This is the derivation suggested in Nandriş (1959:153) and Stojanov (1966:40).) As mentioned before, however, linking nominalisations to passive formations raises the wrong expectation that intransitives will not nominalise at all, and on this basis a derivation of CENs subsequent to passivization has been rejected in Rozwadowska (1997) and Schoorlemmer (1995).

Morphologically, however, a link between the *ne*-nominals and the past passive participle is difficult to deny, and it has been suggested in Sadler *et al.* (1997) that the cognate Russian nominals are derived from the same stem as the Russian past passive participle. A crucial argument for this arising from the Russian data is that when the past passive participle is formed with adding *-t* rather than *-n* to the verbal stem, the nominal also contains *-t* and not *-n*. For example, the past passive participle of the Russian verb *prozit'* 'live through' is *prozit-ij* and the nominalisation from this verb is *prozit-'jo*. (Sadler *et al.* 1997:195). In Bulgarian, however, a similar covariation in form does not appear to be a necessity. Examples can be found where the past passive participle is also formed with *-t* rather than *-n*, but the nominal takes *-n*. For instance, the past passive participle of the verb *brâsna* 'shave' is *brâs-n-at* 'shaven', but the nominalisation is *brâsne-ne* 'shaving'. In addition, in Bulgarian past passive participles are formed both from perfective and imperfective stems, and are not formed from intransitive verbs, whereas *ne*-nominalisations are formed from imperfective verbs only and from both transitive and intransitive verbs. A derivation of the nominalisation from the past passive participle would mean that many nouns will have to derive from non-existent forms. It seems more logical therefore for Bulgarian not to pursue a derivation of the nominals related directly to that of the past passive participle or to the passive forms of the verb.

I would like to suggest that instead we should derive *ne*-nominals from verbs, preserving the verbal argument structure, but should assign to CENs their own pattern of linking arguments. Whereas verbs in Bulgarian have an accusative behaviour, CEN nominalisa-

tions seem to have an ergative behaviour.[3] Onto their most internal complement they map the direct object when there is one, or the subject when there is no direct object.

This would predict that the *by*-phrase would be licensed in transitive constructions, as noted in Rozwadowska (1997). She builds upon the analysis of Williams (1987) of the English *by*-phrase as an ergative marker. Williams (1987) insists that a distinction of nouns into argument taking and non-argument taking is not necessary if we assume that the *by*-phrase contains an ergative case marker. Bulgarian data show, however, that non-CE nominals map the arguments of the underlying verb differently. In non-CE nominals the subject of the underlying verb can be mapped onto a postnominal *na*-PP in preference to the underlying object, as shown in (25). Indeed, these nominals cannot map the agent onto a *by*-phrase at all, and in addition they sometimes absorb one of the verb's arguments. For example the noun *podarâk* 'present' in the example below does not allow mention of what the present was, i.e. it has absorbed the argument position occupied by *kniga* 'book' in the previous examples. The noun can also appear on its own without any satellites. Instead the satellite PPs are treated more like adjuncts.

(25) Podarakât na Ivan za Petâr beše iznenadvašt.
Present.THE of Ivan for Peter was surprising
'Ivan's present for Peter was surprising.'

One telling example of the difference between the postnominal *na*-PP position in CENs and in other nouns is the difference in interpretation in minimal contexts. Thus the most natural interpretation of example (26) (with a CEN) is that parents are being respected (by their children), whereas the most natural interpretation of (27) with a (result nominal) is that the parents are the ones that feel respect (say, for their children).

(26) Uvažavaneto na roditelite e sâštestven element ot vâzpitanieto.
Respect of parents is crucial element of upbringing.THE
'Respect for parents is a crucial element of upbringing'.

(27) Uvaženieto na roditelite e sâštestven element ot vâzpitanieto.
Respect of parents is crucial element of upbringing.THE
'Parents' respect is a crucial element of upbringing'.

Unless we widen the observation of Williams (1987) to suggest ergativity of CENs independent of the *by*-phrase, there is no way to predict the difference of interpretation above.

There is some evidence to suggest that CENs care not only about the syntactic functions of the arguments of the underlying verb, but also about what case they are marked by.

It seems that only the direct object of the verb which is not marked by a preposition (in Bulgarian), or is not lexically marked for case (in Russian) can fill in the argument slot of the derived CEN. According to Schoorlemmer (1995), if a verb assigns oblique case to an internal argument, then the case assignment persists after the nominalisation, see example (28) and (29) (number 70 and 71 in Schoorlemmer (1995:324))

[3]Bulgarian nominalisations seem very similar to a class of nominalisations in Inuit, described in Manning (1996:98)

(28) Zloupotrebljat' vlast'ju
abuse.INF power.INSTR

(29) Zloupotrebljenie vlast'ju
abuse power.INSTR
'abuse of power'

A parallel example can be given for Bulgarian. Thus, the verb *zloupotrebjavam* 'abuse' takes a prepositional object with the preposition *s* 'with' (illustrated in (30)), and the CEN derived from it does so too (illustrated in (31)).

(30) Zloupotrebjavam s vlastta
abuse.1SG with power

(31) Zloupotrebjavaneto s vlastta
abuse with power
'abuse of power'

Note however that (contrary to appearances) in this case the complement position of the noun does not seem to have been filled. Both in Bulgarian and in Russian we can still express the subject of the verb as a complement of the noun (with a *na*-PP in Bulgarian, as in (32) and as NP-GEN in Russian as in (33).

(32) Zloupotrebjavaneto s vlastta na Ivan
abuse-1sg with power of John

(33) Zloupotreblenie Ivana vlast'ju
abuse John.GEN power.INSTR
'John's abuse of power'

The argument taking properties of Bulgarian CENs can be illustrated even more vividly by their ability to take a direct argument without a preposition, as in (34).[4]

(34) Narodât ziveeše v napregnato očakvane velikata promjana.
People lived in tense expectation great change
'The people lived in tense expectation of the great change'.

Bojadžiev *et al.* (1998) note on the basis of example (34) that *ne*-nominalisations, like verbs, must be allowed to assign case (within a framework that permits the notion of *abstract case*).

[4]Interestingly, this argument taking pattern is excluded when the noun bears a definite article, see (i) below (both (34) and (i) are from Maslov (1982:287), see also Tilkov *et al.* (1983:62). I will have nothing to say about this here.

(i) *Narodât ziveeše tixo, vapreki očakvaneto velikata promjana.
People lived quietly, despite expectation.THE great change
'The people lived in tense expectation of the great change'.

The generalisation emerging from the discussion above is that Bulgarian, like many other languages, makes a distinction between complex event nominals on the one hand and simple event nominals and resultatives on the other. Complex event nominals have properties which indicate that these nominals have event semantics. These nominals also inherit the argument structure of the verb, but unlike verbs, in the mapping of their arguments they exhibit an ergative surface pattern.

Before adopting this analysis, one question to ask is whether aspect and *Aktionsarten* (lexical class) are somehow implicated in the argument taking properties of CENs. This question is significant because aspectual characteristics have been used to explain the argument taking properties of complex event nominals (for example for English in Grimshaw (1990)). It also has repercussions for how we model the morphology of *ne*-nominalisations, more specifically, whether we take aspect to be one of the morphosyntactic properties of *ne*-nominalisations and whether it needs to be related to the aspect of the underlying verb.

6.4 *Ne*-nominalisations and aspect

The morphology of *ne*-nominalisations seems to suggest that these nouns might have inherited the aspect of the verb from which they are derived. As described in section 6.2, *ne*-nominalisations derive from imperfective verbs. Intutively, given that CENs have eventive semantics, it seems plausible for them to inherit aspect from the verb. It appears that the Slavic languages differ in this respect. Schoorlemmer (1995:308ff) presents convincing arguments that Russian CENs don't have aspect, unlike Polish CENs, which do. Although Bulgarian derives *ne*-nominalisations from imperfective verbs only, and Russian derives the cognate forms from both imperfective and perfective verbs, the two languages seem to be the same in that they don't assign aspect to the nominals. The following data support this hypothesis for Bulgarian (some of the proposed tests are similar to the ones in Schoorlemmer (1995))

It was mentioned before that CENs permit modification with adverbials like *for an hour* and *in an hour*. Such kinds of adverbials have been used extensively in the literature to distinguish between imperfective and perfective aspect. It is usually said that imperfective eventualities combine with the durative *for an hour*, while perfective eventualities combine with the completive *in an hour*. If CENs had aspect we would expect some CENs to allow only the first of these adverbials, and others only the second, but not both. Examples (35a) and (35b) below show, however, that we can felicitously use both adverbials with the same nominal, provided we supply the appropriate context. This is different from the behaviour of Slavic verbs.

(35) a. Pâtuvaneto v prodalženie na dva dni ja umori.
travel in duration of two days her tired
'Travelling for two days tired her.'

b. Pâtuvaneto do Varna za šest časa ja umori.
Travel to Varna in six hours her tired
'Travelling to Varna in six hours tired her'.

Nominalisations can be used both with a durative meaning, as in (36a), and with a completive meaning, as in (36b).

(36) a. Razpisvaneto na pismata prodalži edin čas.
signing of letters.THE went on one hour
'The signing of the letters went on for an hour.'

b. Razpisvaneto na pismata priključi za edin čas.
signing of letters.THE finished for one hour
'The signing of the letters finished in an hour'.

Ne-nominalisations are equivalent to both perfective and imperfective verbs. The same form in (38a) and (38b) nominalises both the perfective verb in example (37a) and the imperfective one in (37b).

(37) a. Maria razpisa dokumenta i preblednja.
Maria signed.PERF document.THE and paled
'Maria signed the document and paled.'

b. Maria razpisvaše dokumentite v negovo prisâstvie.
Maria signed.IMP documents.THE in his presence
'Maria was signing the documents in his presence'.

(38) a. Razpisvaneto na dokumenta ja nakara da preblednee.
Signing of document.THE her made to pale
'Signing the document made her pale.'

b. Razpisvaneto na dokumentite stavaše v negovo prisâstvie.
Signing.THE of documents was happening in his presence
'The signing of the documents was taking place in his presence.'

6.5 CENs and the lexical class of the verb

There is evidence to suggest that nominalisations are sensitive to the lexical class of the verb. For example, Spencer and Zaretskaya (1998) and Spencer and Zaretskaya (2001) show that generally in Russian it is not possible to form CENs (or situation type nominals in the terminology of the authors) from state verbs (with the exception of verbs of configuration existence and psychological state predicates). Rozwadowska and Spencer (2001) confirm that the same is true of Polish.

It seems that the lexical class of a predicate is an important predictor of whether it will allow a CEN. Closer to our concerns here is the issue of whether the lexical class of the verb (and in this sense its aspectual composition) can explain the argument linking properties of the CEN derived from it, and more specifically whether CENs derived from distinct lexical classes link their satellites in a different way.

Rozwadowska (1997) shows that although psychological predicates in Polish do derive nominalisations, these nominalisations behave as if the predicates were intransitive with respect to licensing a *by*-phrase, whatever their actual argument taking potential.

It seems, however, that psychological predicate nominalisations need not necessarily behave differently from other CENs. Both in Polish and in Russian intransitive psych verbs nominalisations map their object onto the postnominal NP-GEN, as expected, which is shown in (39). Rozwadowska's claim is that the same happens in nominalisations from transitive psychological predicates, i.e. they also map the subject into the postnominal genitive NP, and fail to realise the subject with the Polish equivalent of the *by*-phrase. One can find examples in Russian, however, where nominalisations of transitive psychological predicates behave like nominalisations of other transitive verbs, see (40) (Russian examples adapted from Spencer and Zaretskaya (1998)).

(39) a. Soldaty golodali vo vremja vojny.
soldiers starved in time of.war
'Soldiers went hungry during the war.'

b. golodanie soldatov vo vremja vojny.
starvation soldiers.GEN in time of.war
'Soldiers' wartime starvation.'

(40) a. Učeniki ne znali pravil.
students not know rules
'The students didn't know the rules.'

b. Neznanie pravil učenikami ix ne opradvyvaet.
not.lnowing rules.GEN students.INSTR them not absolve
'The students' lack of knowledge of the rules does not absolve them.'

Similar examples can be derived for Bulgarian, also with the expected argument linking, see (41) below.

(41) a. gladuvaneto na vojnicite prez vojnata.
starvation of soldiers during war
'The soldiers' starvation during the war.'

b. Nepoznavaneto na pravilata ot učenicite ne gi opravdava.
not.knowing of rules by students not them absolve
'The students' lack of knowledge of the rules does not absolve them.'

The conclusion to be drawn from these data is that the precise correlation of lexical class and argument linking calls for futher research, in the very least to explain what differences there might be in the cognate forms in the different Slavic languages. For now, however, it will suffice to say that the proposed formalisation will not refer directly to the lexical class of the verb.

6.6 A paradigm-based approach to *ne*-nominalisations

Bulgarian *ne*-nominalisations are mostly CENs. They are also transpositions, in that they don't modify the meaning of the verb they derive from. Indeed, *ne*-nominalisations denote events, just like verbs, and inherit the argument structure of the verbs they derive

from. Where *ne*-nominalisations differ from verbs is in the syntactic realisation of their arguments, and in this respect they differ from other nouns as well.

Other types of nominals (simple event and result nominals) don't have event semantics, but have referential semantics only. They may also express all or some of the arguments of the verb they derive from, but the syntactic realization they have for their arguments is different from that of CENs, and they may absorb some of the verb's arguments.

All nominals, though, behave like nouns, in that they can take the definite article and make morphological distinctions for number and gender, they take PPs (in Bulgarian), rather than bare noun phrases,[5] they can be modified by adjectives, and so on.

On the other hand, in most cases all nominals will not care about some of the properties that a verbal stem will have, like inflection class,[6] or whether the verb belongs to a class that inflects regularly for, say, tense, etc.

To sum up, we need to distinguish three relationships between the information borne by the verbal stem and the information borne by the nominalisation: in the first place, there is information that the verbal stem and the nominal derived from it share or, in other words, information that the nominal inherits from the verbal stem. Second, there is information that the verbal stem possesses, which is of no consequence to the nominal, i.e. information that gets suppressed (cf. Beard (1995)). Third, there is information that is modified in the process of nominalisation or, put differently, the process of nominalisation may introduce information that is in addition to, or contradictory to, the information present on the verbal stem. The different types of nominalisation will differ with respect to how these possibilities are balanced. Transpositions keep most of the information of the verbal stem, and in this respect they also most resemble inflectional morphological processes, which simply furnish the values for the morphosyntactic properties of a given root or stem. Other kinds of nominalisation modify the semantics of the verb to a much greater extent. Paradigm Function Morphology (PFM) as formalised in Stump (2001) does not allow for the possibility that as a result of the application of a paradigm function some properties associated with the morphological form might get changed. The modifications to PFM contained in Spencer (2004) presuppose a different approach to the informational content of a lexical entry. The next section offers a very brief overview of the PFM formalism. Section 6.6.2 contains a proposal for expressing the formation of CENs and RNs in PFM.

6.6.1 (Generalised) Paradigm Function Morphology

In the morphological model formalised in Stump (2001) word-forms of lexemes are derived via a paradigm function (PF) which is defined as a series of applications of realisational rules (RRs) and has the format in (42):

[5] As we have seen, there is an exception to this, which will be ignored for the time being.

[6] Russian allows a glimpse into an interesting interaction of nominalisation and inflectional class, see Sadler *et al.* (1997). Some generalisations in Bulgarian derivational morphology also suggest that certain derivational patterns might include preference for a particular inflectional class, see Radeva (1991).

(42) $PF(< X, \sigma >) = RR_n \ldots (RR_2(RR_1(<X, \sigma>)))$
where σ is a complete and well-formed set of morphosyntactic properties and X is the root of a lexeme.

The realisational rules themselves are ordered in blocks according to the order of their application and have the format in (43):

(43) $RR_{n,\tau,C}(< X, \sigma >) =_{def} < Y, \sigma >$
where n is the number of the block to which the rule belongs;
τ is a subset of σ and is also the set of properties that the rule realises; and
C is the category of lexemes to which this rule is appropriate.

For example, the realization rule that derives plural nouns may have the shape in (44a) and the rule that derives the plural of DOG might look like (44b):

(44) a. $RR_{1,\{pl\},N}(< X, \sigma >) =< Xs, \sigma >$
b. $RR_{1,\{pl\},N}(< dog, \sigma >) =< dogs, \sigma >$

Another kind of rule relevant to the proposal presented later are the stem-selection rules which have the format in (45):

(45) $RR_{n,\tau,C}(< X, \sigma >) =_{def} < Y, \sigma >$
where Y is one of the stems appropriate for the lexeme with root X.

The roots in these rules are not associated with information. This is modified to an extent in Generalized PFM. Generalised PFM refers to a number of additions and modifications to the model proposed in Spencer (2004). They include a more articulated lexical entry, which includes information about the stems of a lexeme, its syntax and its semantics. The paradigm function is defined over lexemes, rather than roots. The realisation rules (generation of phonological strings) are divorced from linearisation (the placement of these strings relative to the stem).

6.6.2 Deriving nominalisations

The lexical entry of the verb

I will assume here that the stem from which *ne*-nominals are derived is a verbal stem. For each verbal stem (at least) the following information will have to be available: syntactic category, semantic representation, argument structure, syntactic realization of arguments, and the linking between the two. Since PFM usually expresses properties associated with forms as sets, we can define for each lexeme a set of properties ρ which it gets assigned in the lexicon, and which are separate from the set of morphosyntactic properties σ which a lexeme is associated with in order to derive the full set of its wordforms.

Below is the possible representation of the lexical entry of the verb *DAM* 'give':[7]

[7]The abbreviations are to be interpreted as follows: CAT – category, INFL-CL – inflection class, SEM – semantics, ARG-ST – argument structure, SYNT-STR – syntactic structure, SUBJ – subject, OBJ - object, OBL – oblique.

$$\left\langle \text{DAM}, \left\{ \begin{array}{l} \text{CAT: verb} \\ \text{INFL-CL: e-conj} \\ \text{SEM: EVENT(x)} \\ \text{ARG-ST: } \boxed{1}\text{ giver, } \boxed{2}\text{ givee, } \boxed{3}\text{ given} \\ \text{SYNT-STR: } \boxed{1}\text{ SUBJ, } \boxed{3}\text{ OBJ, } \boxed{2}\text{ OBL} \end{array} \right\} \right\rangle$$

Such an enriched representation of a lexical entry has been proposed as part of a more wide-ranging extension of PFM by Spencer (2004). The boxed numbers in the lexical entry are meant to express coindexation. DAM is a ditransitive verb, it expresses a relation between three participants: a giver, a givee, and something given. The most frequent linking pattern for this verb is to have the giver in the subject position, the given in the object position, and the givee in the oblique.

The *ne*-nominalisation function

The *ne*-nominalisation function will associate the following information with the verb (expressed as before as a set υ of values and features):

$$\left\{ \begin{array}{l} \text{CAT: noun} \\ \text{GENDER: neuter} \\ \text{SEM:} \\ \text{ARG-ST:} \\ \text{SYNT-STR: } \textit{na}\text{-}\boxed{3}\text{ or } \boxed{1}\text{NP, p-NP, p-NP} \end{array} \right\}$$

There are features in the set of properties υ that are not present in the set of properties ρ (for example GENDER) and there are features for which the opposite is true (INFL-CL). There are features which are defined in both sets, but have different values (SYN-STR), and finally there are features that are present in the set υ, but whose values are not specified in that set (for example ARG-ST). The properties of the *ne*-nominalisation derived from DAM will be determined by the two sets according to the following rules: if a feature is present in the set ρ, but absent in the set υ, this feature is ignored. If a feature is present in the set υ, but its value is unspecified, it takes the value it has in the set ρ, if a feature is specified in both sets but with conflicting values, then the value in the set ρ is ignored.

This should ensure that the *ne*-nominal will ignore the inflectional class of the verb it derives from, will inherit its argument structure and semantics from the verb, but will determine its own syntactic realisation of the semantic arguments, and will specify features that verbs just don't possess, like gender. The syntactic realisation of arguments is determined by the coindexation of the NP governed by the preposition *na* with either the given/object of the verb, or with the giver/subject of the verb. These are also meant to be in a hierarchical order, i.e. the subject will get mapped to this position only when there is no object available to be mapped onto it.

Let's assume that a function called *FORM-NE* associates the stem and its properties with the nominal properties in the set υ. How the actual form of the *ne*-nominalisation will be spelled out, however, needs to be specified separately by realisation rules. To decide on the form of the rules we need to look at the segmentation of *ne*-nominalisations. Most

typically, *ne*-nominals segment as shown in table 6.6.2:

prekâs 'interrupt'	*va*	*ne*
prevlkjuč 'switch'	*va*	*ne*
zamest 'act as deputy'	*va*	*ne*
root	?	nom. suffix

The final segment is the nominalising suffix, the initial element is the root. What is called into question is the nature of the segment *-va-* (or its allomorphs). This suffix is usually said to derive imperfective verbs from perfective bases. However, as we have seen, we have no reason to attribute the category of aspect to *ne*-nominalisations. There is instead good evidence to think of the forms with *-va-* as being verbal stems. One reason for this is precisely the fact that they cannot be associated with any one semantic value across the categories where they appear as formatives, namely the derived or secondary imperfective verbs, the *ne*-nominalisations, or the imperfective participles. In addition, the stem with *-va-* appears in derived nominals where it is even clearer that we cannot hypothesize the presence of aspect, and not even the presence of event semantics. Agentive nominals with the suffix *-ač* (as in *prekâsvač* 'switch' from the verb *prekâsvam* 'to interrupt'), for example, select for the stem with *-va-*, even though we will not wish to attribute to them any verbal properties.

The suffix *-va-* then will be added to the verbal root (or zero stem) by a stem formation rule. Another rule needs to add the suffix *-ne*.

One possible formulation of the realisation rules is as follows:

$FORM - NE(X) =_{def} \mathrm{RR}_2\mathrm{RR}_1(X)$
$\mathrm{RR}_1(X) = Y$ where Y is *va*-stem derived from X
$\mathrm{RR}_2(X) = Xne$

The function that derives a result nominal will be different in terms of the interaction of information between the verbal lexeme and the set of properties introduced by the nominalisation function. The function deriving nominals like *podarâk* 'present', for example, might associate with the verbal lexeme the following information:

$$\left\{\begin{array}{l}\text{CAT: noun}\\ \text{GENDER: neuter}\\ \text{SEM: PRESENT(given)}\\ \text{ARG-STR: giver, givee}\\ \text{SYNT-STR: PP, PP}\end{array}\right\}$$

A nominal like this, though derived from the verb *podarjavam* 'give as a present', will inherit practically nothing from the verb, though perhaps the *giver* and the *givee* in its argument structure need to be co-indexed with the relevant roles in the verbal entry.

6.7 Conclusions and directions for future research

This brief examination of the argument taking properties of deverbal nominals in Bulgarian suggests that there are important links to be made with the phenomenon of ergativity.

It is not sufficient to say with respect to these nominals that the cognate of the *by*-phrase is a marker of ergativity, indeed it is wrong to attribute to the *by*-phrase the property of being a licensor of ergative argument mapping, as in Williams (1987), since in Bulgarian we can talk of ergative behaviour in its absence.

In (at least some) Slavic languages it is also misleading to talk of nominalisation as a passivising process. In this respect the arguments raised in Schoorlemmer (1995) and Rozwadowska (1997) apply with equal force to Bulgarian. Synchronic morphological data also suggests that it will be less illuminating to derive Bulgarian nominalisations with the suffix *-ne* on the same stem as the past passive participle, thus making Bulgarian nominals even further removed from passive formations than is perhaps the case for Russian.

Though argument mapping properties of nominals and event structure have been linked, for example in Grimshaw (1990), there does not seem to be enough ground to assign Bulgarian CENs their own aspectual values, or even to divide them into classes with respect to the *Aktionsart* of the underlying verb. It seems, however, that the Aktionsart of the verb can in some cases predict the (im)possibility of deriving a CE nominal.

In formalising the process of nominalisation in PFM the observation has been made that this process involves a complex interaction between the information associated with the underlying verb and the information associated with the nominal. CENs will differ from other nominals in terms of the amount of information that they will inherit from the verb. They also differ from other nominals in terms of the syntactic expression of the argument structure shared with the verb. CENs also inherit the verb's event semantics.

More research is needed to clarify whether the ergative behaviour of CENs can be closely correlated with any of their other properties. An investigation needs to be undertaken into the behaviour of satellites within the noun phrase, for example possessive pronouns and adjectives, pronominal possessive clitics, etc., since there seem to be interesting differences between Polish, Russian, and Bulgarian in this respect. The correlations of argument mapping and case might also be in an interesting area of future work, so as to identify the differences and similarities in argument mapping between similar languages with a morphological category of case (Polish and Russia), and without one (Bulgarian).

Bibliography

Anderson, Stephen R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Aronoff, Mark. 1994. *Morphology by Itself: Stems and Inflectional Classes*. Cambridge, MA: MIT Press.

Beard, Robert. 1995. *Lexeme-Morpheme Base Morphology*. Albany: State University of New York Press.

Bojadžiev, Todor, Ivan Kucarov, and Jordan Penčev. 1998. *Sâvremenen bâlgarski ezik*. Sofia: Izdatelska kâšta "Petâr Beron".

Bredenkamp, Andrew, Louisa Sadler, and Andrew Spencer. 1998. Investigating argument structure: The Russian nominalization database. In *Linguistic Databases*, ed. by John Nerbonne, number 77 in CSLI Lecture Notes, 137–159. Stanford, California: CSLI Publications.

Grimshaw, Jane Barbara. 1990. *Argument Structure*. Cambridge, Massachusetts and London, England: The MIT Press.

Manning, Christopher D. 1996. *Ergativity: Argument Structure and Grammatical Relations*. Stanford, California: CSLI Publications.

Maslov, Jurij Sergeevič. 1982. *Gramatika na bâlgarskiia ezik*. Sofia: Nauka i izkustvo.

Matthews, Peter Hugoe. 1972. *Inflectional Morphology: A Theoretical Study Based on Aspects of Latin Verb Conjugation*. Cambridge: Cambridge University Press.

Nandriş, Grigore. 1959. *Old Church Slavonic Grammar*. University of London: The Athlone Press. Reprinted with corrections 1965.

Radeva, Vasilka. 1991. *Slovoobrazuvaneto v bâlgarskija knižoven ezik*. Sofia: Universitetsko izdatelstvo "Sv. Kliment Oxridski".

Rozwadowska, Bożena. 1997. *Towards a Unified Theory of Nominalisations: External and Internal Eventualities*. Wroclaw: Wydawnictwo Uniwersytetu Wroclaskiego.

——, and Andrew Spencer. 2001. Psychological predicates in Polish: Event participants, nominalizations and mapping. In *Current Issues in Formal Slavic Linguistics*, ed. by Gerhild Zybatow, Uwe Junghanns, Grit Mehlhorn, and Luka Szucsich, number 5 in Linguistik International, 536–545, Frankfurt am Main. Peter Lang.

Sadler, Louisa, Andrew Spencer, and Marina Zaretskaya. 1997. A morphomic account of a syncretism in Russian deverbal nominalisations. In *Yearbook of morphology*, ed. by Geert Booij and Jaap van Marle, 181–215. Kluwer Academic Publishers.

Schoorlemmer, Maaike. 1995. *Participial Passive and Aspect in Russian*. OTS Dissertation Series. Utrecht University: Onderzoeksinstituut voor Taal en Spraak.

Spencer, Andrew. 1999. Transpositions and argument structure. In *Yearbook of morphology 1998*, ed. by Geert Booij and Jaap van Marle, 73 – 101. Dordrecht/Boston/London: Kluwer Academic Publishers.

——, 2004. Generalised paradigm function morphology. Manuscript.

——, and Marina Zaretskaya. 1998. Stative predicates in Russian and their nominalizations. In *Essex Research Reports in Linguistics*. University of Essex.

——, and Marina Zaretskaya. 2001. The stative middle construction in Russian - a covert category. In *Current Issues in Formal Slavic Linguistics*, ed. by Gerhild Zybatow, Uwe Junghanns, Grit Mehlhorn, and Luka Szucsich, number 5 in Linguistik International, 546–553, Frankfurt am Main. Peter Lang.

Steinke, Klaus. 1999. Slovoobrazuvane i valentnost. In *Aktualni problemi na bâlgarskoto slovobrazuvane*, ed. by Julia Baltova, Kornelija Ilieva, and Margarita Tetovska-Troeva, 79–84. Sofia: Universitetsko izdatelstvo "Sv. Kliment Oxridski" and Akademično izdatelstvo "Prof. Marin Drinov".

Stojanov, Stojan. 1966. Otglagolni saštestvitelni s nastavka *-ne* ot svâršeni glagoli. *Ezik i literatura* 21.39–48.

——. 1977. *Slovoobrazuvaneto v bâlgarskiia ezik. I. Imena*. Sofia: Daržavno izdatelstvo "Narodna prosveta".

Stump, Gregory T. 2001. *Paradigm Function Morphology. A Theory of Inflectional Form.*. Cambridge: Cambridge University Press.

Tilkov, Dimitâr, Stojan Stojanov, and Konstantin Popov (eds.) 1983. *Gramatika na sâvremenniia bâlgarski knižoven ezik. Morfologija.*, volume 2. Sofija: Izdatelsto na Bâlgarskata akademija na naukite.

Williams, Edwin. 1987. English as an ergative language: The theta structure of derived nouns. In *Papers from the 23rd Annual Regional Meeting of the Chicago Linguistic Society*, volume 23, 366–375. Chicago Linguistic Society.

Gergana Popova Department of Language and Linguistics
University of Essex
Wivenhoe Park, Colchester CO4 3SQ, UK
gdpopo@essex.ac.uk

The following volumes of ESSEX RESEARCH REPORTS IN LINGUISTICS have appeared since 2000:

47:2004	Robert D. Borsley	*On the Nature of Welsh VSO Clauses*
-/-	Beatriz Mariz Maia de Paiva	*Exploring the Relationships between Theories of Second Language Acquisition and Relevance Theory*
-/-	Rebecca Clift	*Discovering Order*
46:2004	Harald Clahsen	*Linguistic Perspectives on Morphological Processing*
-/-	Melanie Ring and Harald Clahsen	*Distinct Patterns of Language Impairment in Down Syndrome and Williams Syndrome: The case of syntactic chains*
-/-	Michaela Wenzlaff and Harald Clahsen	*Finiteness and Verb-Second in German Agrammatism*
45:2003	Hahne, A., Mueller, J. and H. Clahsen	*Second language learners' processing of inflected words: Behavioral and ERP evidence for storage and decomposition*
-/-	Marinis, T., Roberts, L., Felser, C. and H. Clahsen	*Gaps in second language sentence processing*
44:2003	Johnson, W. and D. Britain	*L vocalisation as a natural phenomenon*
-/-	Matsumoto, K. and D. Britain	*Contact and obsolescens in a diaspora variety of Japanese: The case of Palau in Micronesia*
43:2003	Clahsen, H., M. Ring and C. Temple	*Lexical and morphological skills in English-speaking children with Williams Syndrome*
-/-	Wenzlaff, M. and H. Clahsen	*Tense and agreement in German agrammatism*
-/-	Clahsen, H., M. Hadler and H. Weyerts	*Frequency effects in children's and adults' production of inflected words*
-/-	Papadopoulou, D. and H. Clahsen	*The role of lexical and contextual information in parsing ambiguous sentences in Greek*
41:2002	Britain, D.	*The British History of New Zealand English?*
-/-	Britain, D.	*Phoenix from the ashes?: The death, contact and birth of dialects in England*
-/-	Britain, D.	*Surviving 'Estuary English': Innovation diffusion, koineisation and local dialect differentiation in the English Fenland*
40:2002	Felser, C., L. Roberts, R. Gross	*The processing of ambiguous sentences by first and*

	and T. Marinis	*second language learners of English*
-/-	Clahsen, H., I. Sonnenstuhl and J. Blevins	*Derivational morphology in the German mental lexicon: A dual mechanism account*
39:2002	Marinis, T.	*Acquiring the left periphery of the Modern Greek DP*
-/-	Marinis, T.	*Subject-object asymmetry in the acquisition of the definite article in Modern Greek*
-/-	Papadopoulou, D. and H. Clahsen	*Parsing strategies in L1 and L2 sentence processing: a study of relative clause attachment in Greek*
38:2001	Britain, D.	*Dialect contact and past BE in the English Fens*
-/-	Britain, D.	*If A changes to B, make sure that A exists: A case study on the dialect origins of New Zealand English*
-/-	Matsumoto, K. and D. Britain	*Conservative and innovative behaviour by female speakers in a multilingual Micronesian society*
37:2001	Atkinson, M.	*Defective intervention effects, die!*
-/-	Atkinson, M.	*Putting the X on TH/EX*
-/-	Felser, C.	*Wh-copying, phases, and successive cyclicity*
36:2001	Clahsen, H. and C. Temple	*Words and Rules in Children with Williams Syndrome*
-/-	Radford, A. and E. Ramos	*Case, Agreement and EPP: Evidence from an English-speaking child with SLI*
35:2001	Patrick, P.	*The speech community*
-/-	Figueroa, E. and P. Patrick	*The meaning of kiss-teeth*
34:2000	Atkinson, M.	*Minimalist visions*
-/-	Radford, A.	*Children in search of perfection: Towards a minimalist model of acquisition*
-/-	Hawkins, R.	*Persistent selective fossilisation in second language acquisition and the optimal design of the language faculty*
-/-	Atkinson, M.	*Uninterpretable feature deletion and phases*
-/-	Felser, C.	*Wh-expletives and secondary predication: German partial wh-movement reconsidered*
-/-	Borsley, R.D.	*What do 'prepositional complementisers' do?*
-/-	Clahsen, H., I. Sonnenstuhl, M. Hadler and S. Eisenbeiss	*Morphological paradigms in language processing and language disorders*

33:2000	Radford, A.	*Minimalism: Descriptive perspectives*
32:2000	Clift, R.	*Stance-taking in reported speech*
-/-	McDonough, J. and S. McDonough	*Composing in a foreign language: an insider-outsider perspective*
-/-	Arnold, D.	*Corpus access for beginners: the W3Corpora project*
-/-	Patrick, P. and S. Buell	*Competing creole transcripts on trial*
31:2000	Weyerts, H., M. Penke, T. F. Muente, H.-J. Heinze and H. Clahsen	*Word order in sentence processing: An experimental study of verb placement in German*
-/-	Nakano, Y., C. Felser and H. Clahsen	*Antecedent priming at trace positions in Japanese long-distance scrambling*.
30:2000	Al-Wer, E.	*Education as a speaker variable in Arabic variationist studies*
-/-	Al-Wer, E.	*Jordanian and Palestinian dialects in contact: vowel raising in Amman*
-/-	Al-Wer, E.	*Perspectives on two Arabic emphatic sounds: A merger or an artificial split?*
-/-	Cook, V.	*The innateness of a Universal Grammar principle in L2 users of English*
29:2000	Matsumoto, K. and D. Britain	*Hegemonic diglossia and pickled radish: symbolic domination and resistance in the trilingual Republic of Talau*
	Britain, D.	*The difference that space makes: an evaluation of the application of human geographic thought in sociolinguistic dialectology*
28:2000	Britain, D. and A. Sudbury	*There's sheep and there's penguins: 'Drift' and the use of singular verb forms of BE in plural existential clauses in New Zealand and Falkland Island English*