5

Dissociations Between Argument Structure And Grammatical Relations

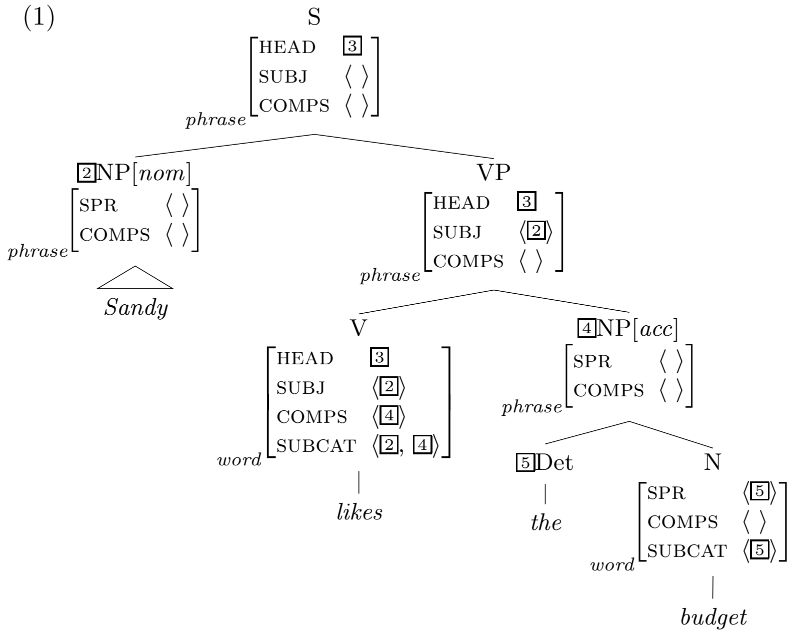
Christopher D. Manning and Ivan A. Sag

In Pollard and Sag (1987) and Pollard and Sag (1994:Ch. 1–8), the subcategorized arguments of a head are stored on a single ordered list, the SUBCAT list. However, Borsley (1989) argues that there are various deficiencies in this approach, and suggests that the unified list should be split into separate lists for subjects, complements, and specifiers. This proposal has been widely adopted in what is colloquially known as HPSG3 (Pollard and Sag (1994:Ch. 9) and other recent work in HPSG). Such a move provides in HPSG an analog of the external/internal argument distinction generally adopted in GB, solves certain technical problems such as allowing prepositions to take complements rather than things identical in SUBCAT list position to subjects, and allows recognition of the special features of subjects which have been noted in the LFG literature, where keyword grammatical relations are used. In HPSG3, it is these *valence features* SUBJ, COMPS and SPR whose values are 'cancelled off' (in a Categorial Grammar-like manner) as a head projects a phrase. A lexical head combines with its complements and subject or specifier (if any) according to the lexically inherited specification, as in (1).[1]

[1]This paper is based on part of a talk given at the Tübingen HPSG workshop in June 1995, and distributed as Manning and Sag (1995). However, it excludes much material presented there, which will now appear in other places (Manning et al. in press, Manning and Sag submitted, Sag and Manning forthcoming). The paper also has an updated analysis – one consistent with that of Manning et al. (in press) – which we believe avoids the flaws attributed to the analysis of Manning and Sag (1995) by Webelhuth (forthcoming). However, because of space limitations, causatives are no longer discussed here and the reader interested in this topic should

Lexical And Constructional Aspects of Linguistic Explanation.
Gert Webelhuth, Jean-Pierre Koenig and Andreas Kathol (eds.).

(1)



When Borsley (1989) suggested dividing the SUBCAT list into multiple valence lists, we believe that he intended that they would replace the SUBCAT list, but this is not in fact what happened. In Pollard and Sag (1994:Ch. 9), the SUBCAT list is kept as an attribute of lexical signs. Its value is the append of the SUBJ, SPR and COMPS lists. As presented there, this move seems more an expediency than a necessity: it allows the HPSG binding theory to be retained unchanged, rather than having to redefine it over the new valence lists. But the SUBCAT list merely summarizes the valence of a lexical sign, without having any independent life of its own. It remains unaffected in the construction of syntactic phrases, except that, in virtue of the various identities between SUBCAT list members and members of valence lists, the SUBCAT list's members become fully specified as the valence list values are identified with actual subjects, complements and specifiers. Once a complete phrase is constructed, the lexical head's SUBCAT list is fully specified, as indicated in (1), and may be used as the locus of binding theory.

This redundancy has been broken in recent work. The *canonical* relationship between the SUBCAT list and the valence lists is still an append relationship, but various other possibilities have been explored.

---

see the other papers just mentioned. We thank the audience at Tübingen, and, particularly, Georgia Green and Stephen Wechsler for helpful comments.

To begin with a simple example, many languages allow arguments to not be realized, in what is sometimes known as "free pro-drop", as in the Japanese sentence (2a):

(2) a. Naoki-ga mi-ta
Naoki-NOM see-PAST
'Naoki saw (it).'

b. $$\begin{bmatrix} \text{SUBJ} & \langle \boxed{1}\text{NP}[nom] \rangle \\ \text{COMPS} & \langle\ \rangle \\ \text{SUBCAT} & \langle \boxed{1}_i,\ \text{NP}[pro]_j \rangle \\ \text{CONT} & {}_{seeing}\begin{bmatrix} \text{SEER} & i \\ \text{SEEN} & j \end{bmatrix} \end{bmatrix}$$

In this sentence, there is only one surface argument. Recent work has argued that the supposition of traces or other empty elements is unnecessary to explain *wanna* 'contraction' (Pullum 1997), and stands in the way of understanding a host of other phenomena thought to legitimate inaudibilia (Sag and Fodor 1994, Sag 1998), and so we would not want to postulate an empty 'pro' element as an independent sign. But nevertheless one wants to capture how in some sense *mi-* 'see' takes two arguments. For example, these two arguments need to appear in a discussion of binding theory when discussing interpretations of this sentence. One can do this by proposing an 'object pro-drop' lexical entry for the verb as shown in (2b), and maintaining the approach that binding theory operates on the SUBCAT list, not the valence lists.

In this model, the SUBCAT list no longer captures surface subcategorization, but is an attribute of only lexical signs, used to explain phenomena such as binding, linking, and 'deep' subcategorization, and hence it has become similar to certain notions of *argument structure*. Thus, in recent work the SUBCAT list has been renamed as ARG-ST, and we will use this name henceforth. But it should be emphasized that the ARG-ST list is a syntactic representation, just like its predecessor the SUBCAT list, and is not to be viewed as a partial semantic representation or some sort of substitute for one.

Pro-drop is one of a class of cases, together with unbounded dependencies and pronominal affixes, where arguments do not appear on a valence list (Sag and Fodor 1994, Sag and Godard 1994, Miller and Sag 1997, Bouma et al. 1998). Other recent work has focussed on the analysis of data that involves somewhat more interesting dissociations between valency and argument structure than just the valence lists being a subset of the ARG-ST list (Manning 1996, Manning et al. in press, Abeillé et al. to appear). The ability to dissociate argument structure from valence in this way takes HPSG a certain distance from the monolevel, monostratal roots of GPSG and early HPSG. The purpose of this paper is to better motivate the existence of two independent syntactic notions of valency and argument structure and to examine the kinds of dissociations that can occur, with reference to passives, binding and ergative languages. While doing that, we will suggest some argument structure representations that differ from those presented previously, and for which there is interesting empirical support.

5.1 Binding Theory and Passives

The HPSG binding theory is based on hierarchical argument structure rather than constituent structure (as also in Johnson (1977)). As Pollard and Sag (1992, 1994) demonstrate, this approach to binding provides an immediate solution to a variety of dilemmas facing any account of English binding stated in terms of constituency-based notions such as c-command. It maintains three binding principles, analogous to those of GB; they are given informally in (3):[2]

(3) HPSG Binding Theory:

Principle A. A locally a-commanded anaphor must be locally a-bound.
Principle B. A personal pronoun must be locally a-free.
Principle C. A non-pronoun must be a-free.

These principles require an anaphor to be coindexed with a less oblique ARG-ST member, if there *is* such a less oblique coargument. Otherwise, anaphors are free (subject to various discourse and processing considerations) to be bound by appropriate elements in the discourse context.

This binding theory is adequate for English, but crosslinguistic coverage of binding phenomena requires more parametric options (Dalrymple 1993). In many languages, reflexives cannot be bound by just any less oblique (local) NP, but rather their antecedence is restricted to what we might loosely call "subjects". At least to a first approximation this is true of languages such as Japanese, Russian, Inuit, and Sanskrit. Given that the binding theory in HPSG is defined on ARG-ST (an assumption that we will later actively argue for), the natural explanation for such data is to suggest that in these languages, reflexives must be bound by the first element on some ARG-ST list. We will formalize such a notion with the definition and principle in (4), drawn from Manning (1996).

(4) a. An **a-subject** is an entity that is first on some ARG-ST list.
b. A-subject-oriented anaphors must be a-bound by an a-subject.

[2] A-command, a-bound, and a-free are the same notions as o-command, o-bound, and o-free from Pollard and Sag (1994), now defined on ARG-ST, but the new names are meant to evoke the argument structure based theory of binding we employ.

This allows us to explain why *Kaali* is not a possible binder in the Inuit example (5):

(5) Juuna-p Kaali immi-nik uqaluttuup-p-a-a
Juuna-ERG Kaali.ABS self-MOD tell-IND-TR-3SG.3SG
'Juuna$_i$ told Kaali$_j$ about self$_{i/*j}$.'

A second parametrization of the binding theory is that while classical reflexives are clause bounded, many languages allow long distance reflexives. For example, both the Inuit reflexive *immi* and the Japanese reflexive *zibun* can be bound by any a-commanding a-subject. Such long distance anaphors might be said to obey Principle Z (Xue et al. 1994):

(6) Principle Z. A locally a-commanded long distance anaphor must be a-bound.

Now consider the interaction of passive and subject-oriented reflexives. If our theory of passive was that drawn from HPSG1 – a lexical rule that cyclically permuted the SUBCAT, now ARG-ST, list as in (7):[3]

(7)
$$\begin{bmatrix} \textit{active-verb} & \\ \text{ARG-ST} & \langle \boxed{1}_i, \boxed{2}, \ldots \rangle \\ \text{CONT} & \boxed{3} \end{bmatrix} \rightarrow \begin{bmatrix} \textit{passive-verb} & \\ \text{ARG-ST} & \langle \boxed{2}, \ldots \rangle \; (\oplus \langle \text{PP}[\textit{by}]_i \rangle \;) \\ \text{CONT} & \boxed{3} \end{bmatrix}$$

then our prediction is clear: the only possible binder of subject-oriented reflexives, the a-subject, is now the NP that is the subject of the passive ($\boxed{2}$). However, in many languages, this is not in fact the case. Perlmutter (1984) observed this for the case of Russian. While in (8a), the reflexive *sebe* must be bound by the subject, in the passive (8b), the antecedent can be either the surface subject or the agent argument (sometimes known as the logical subject, following Jespersen (1924)).

(8) a. Boris mne rasskazal anekdot o sebe
Boris.NOM me.DAT told joke about self
'Boris$_i$ told me a joke about himself$_i$.'

b. Èta kniga byla kuplena Borisom dlja sebja
this book.NOM was bought Boris.INSTR for self
'This book was bought by Boris$_i$ for himself$_i$.'

Perlmutter argued from these data that the passive must have a complex representation of some sort. In particular, Perlmutter used these examples to argue within Relational Grammar (RG) that both the logical subject and the surface subject of a passive must be a 1 at some level: the logical subject is the initial 1, while the surface subject is the

[3] We use $\oplus$ to indicate list concatenation or append, and round brackets to indicate optionality.

final 1.

In essence we accept this argument, and suggest that we want a representation for passives (at least in languages like Russian) where both the surface subject and the logical subject qualify as a-subjects. However, we argue that such an analysis does not require multiple strata of grammatical relations, as in RG, but can more restrictively be captured by suggesting that passive lexemes possess a nested argument structure, which has the immediate result that passive clauses have two a-subjects.[4] Indeed, below we will present arguments from ergative languages that such an alternative analysis in terms of argument structure rather than grammatical relations is not only possible but necessary.

Various ways have been suggested within HPSG for licensing derived types, such as the passive lexemes that we are dealing with here. While any of them could be used to produce a similar analysis to the one presented here, we will develop our account in terms of a theory of derivational types, which specify a declarative relationship between a SOURCE stem and a RESULT stem (which is morphologically 'derived' from it). Such an approach is closely related to what Copestake (1992) proposes (see also Meurers (1995)). It has the advantages of allowing inheritance within the hierarchical lexicon of HPSG to extend over both stem and word types and derivational types (as in the approach of Riehemann (1993)) while preserving the locality of information and lexical integrity of words within the syntax that is well-captured within the lexical rules approach. Thus we will suppose that the universal characterization of passive is as in (9):[5]

(9)
$$\begin{bmatrix} \textit{passive-drv} & \\ \text{RESULT} & \begin{bmatrix} \textit{pass-v-lxm} & \\ \text{ARG-ST} & \langle \boxed{2}_j, \langle \boxed{1}, \text{PRO}_j \rangle \oplus \boxed{3} \rangle \\ \text{CONT} & \boxed{4} \end{bmatrix} \\ \text{SOURCE} & \begin{bmatrix} \textit{trans-v-lxm} & \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \oplus \boxed{3} \\ \text{CONT} & \boxed{4} \end{bmatrix} \end{bmatrix}$$

Such a derivational type is to be read as saying that basic and other derived lexemes of the SOURCE type license additional lexemes of the RESULT type. We propose that the passive lexeme's ARG-ST value is a

[4] For a similar argument, cf. Grimshaw (1990:167–173).

[5] This passive is intrinsically promotional; some have argued that the universal rule of passive should only mention subject demotion, to account for passive-like structures where nothing is promoted, such as in Lithuanian, but we would provide a different (though related) type for such cases.

list consisting of the second (undergoer) argument of the source lexeme followed by a list that is the same as the ARG-ST value of the source, except that the second element has been replaced by a PRO placeholder. These placeholder elements in ARG-ST lists are used to mark positions coindexed with an element higher in the ARG-ST, and are needed for binding, as we will see below. The passive's ARG-ST value is thus a 'nested' list (a list that contains another list as a member), a fact that will play a crucial role in our account of constraints on binding.

The passive verb of (8b) will then be:

(10)
$$\left[\begin{array}{ll} \text{ARG-ST} & \langle\ \boxed{2}\text{NP}[nom]_j,\ \langle\boxed{1}\text{NP}[instr]_i,\ \text{PRO}_j,\ \boxed{5}\text{PP}_k\rangle\ \rangle \\ \text{CONT} & {}_{buying}\left[\begin{array}{ll} \text{BUYER} & i \\ \text{BOUGHT} & j \\ \text{BENEFICIARY} & k \end{array}\right] \end{array}\right]_{pass\text{-}v\text{-}lxm}$$

In (10), the reflexive beneficiary $\boxed{5}$ is inside the nested ARG-ST list, and therefore it is a-bound by two a-subjects. This means that if the beneficiary $\boxed{5}$ is a long distance a-subject-oriented anaphor, then Principle Z and the a-subject principle can be satisfied by $\boxed{5}$ being coindexed with either $\boxed{1}$ or $\boxed{2}$, both of which are a-commanders and a-subjects. This is exactly the result we want to explain the Russian data above. Note that our theory predicts that the surface subject is another possible binder of the anaphor in (8b), but this is being ruled out due to its being an inanimate NP. Similar data that supports this analysis occurs in many languages; (11) shows a passive from the syntactically ergative language West Greenlandic Inuit.[6] Examples from Sanskrit, Hindi, and Japanese are discussed by Manning (1996:57,124–127), and Shibatani (1988).

(11) Naja Tobiasi-mit uqaluttuun-niqar-p-u-q taa-ssu-ma
Naja.ABS Tobias-ABL tell-PASS-IND-INTR-3SG [DEM-SG-ERG
itigartis-sima-ga-a-ni
turn.down-PRF-PRT.TR-3SG-4SG]
'Naja$_j$ was told by Tobias$_i$ that he$_k$ had turned self$_{i/j}$ down.'

Thus the data from passives that we have examined argues for three things: (i) that there must be a new more articulated argument structure for passives along the lines that we have proposed; (ii) that passive must be stated so as to realign argument structure, not valence lists; and (iii) that binding possibilities are sensitive to this argument structure, and not to surface phrase structure or surface valence patterns.

[6] As well as lexical reflexives, Inuit has a reflexive pronominal agreement marker, here glossed as '4th person', its traditional name. See Manning (1996) for justification of the syntactic ergativity of Inuit.

5.2 Syntactically ergative constructions

In cases of dissociations between argument structure and surface valency, the HPSG architecture predicts that binding possibilities and related phenomena should depend solely on the argument structure and be independent of valency. This prediction is startlingly confirmed by the behavior of syntactically ergative and Western Austronesian languages. This is examined in more detail in Manning (1996), Wechsler and Arka (to appear), and Wechsler (this volume), but will be illustrated briefly here, with an eye to the development of an HPSG analysis.

Western Austronesian languages allow various relationships between argument structure and valence list configuration, mediated by so-called voice morphology. The best known case of this is Tagalog (Schachter 1976, Kroeger 1993), but here we will present some evidence from Toba Batak (Schachter 1984), which has a more rigid configurational surface structure than Tagalog, and hence demonstrates some points more clearly. In particular, it clearly shows the independence of binding from surface structure command relationships. Toba Batak has a distinction between active voice (*mang-*) and objective voice (*di-*) forms of verbs:

(12) a. Mang-ida si Ria si Torus
AV-see PM Ria PM Torus
'Torus sees/saw Ria.'

b. Di-ida si Torus si Ria
OV-see PM Torus PM Ria
'Torus sees/saw Ria.'

The active voice (12a) has the logical subject of the clause in the clause final subject position, while the objective voice (12b), which tends to be used in unmarked contexts, has what we might term the Undergoer (Foley and Van Valin 1984) in subject position. Schachter (1984) provides evidence that both arguments in both voices in (12) are core roles (as opposed to obliques and adjuncts); see also the more extensive arguments in Kroeger (1993) for Tagalog. Thus the correct analysis is not to view one of (12a) or (12b) as a passive or antipassive (as has often been done in the generative literature), but rather as both exhibiting different relationships between argument structure and surface valence. There is strong evidence that a verb and the following NP of a transitive clause form a constituent, which we will call a VP, regardless of the verbal voice chosen. These VPs can be coordinated regardless of their voice:

(13) a. Man-uhor baoang jala mang-olompa mangga halak an
[AV-buy onions] and [AV-cook mangoes] man
'The man buys onions and cooks mangoes.'

b. Di-tuhor si Ore jala di-lompa si Ruli mangga
[OV-buy PM Ore] and [OV-cook PM Ruli] mangoes
'Ore buys and Ruli cooks mangoes.'

Other arguments are that an adverb cannot appear in the middle of the VP between the verb and the NP, though adverbs can generally occur between other major constituents, and that the pitch accent of a sentence occurs on the last stressed syllable of this VP, in both voices, including on the verb of an intransitive sentence (Emmorey 1984). Thus the first NP of transitive clauses will be analyzed as being on the COMPS list and will combine with the verb as a head-complement phrase.

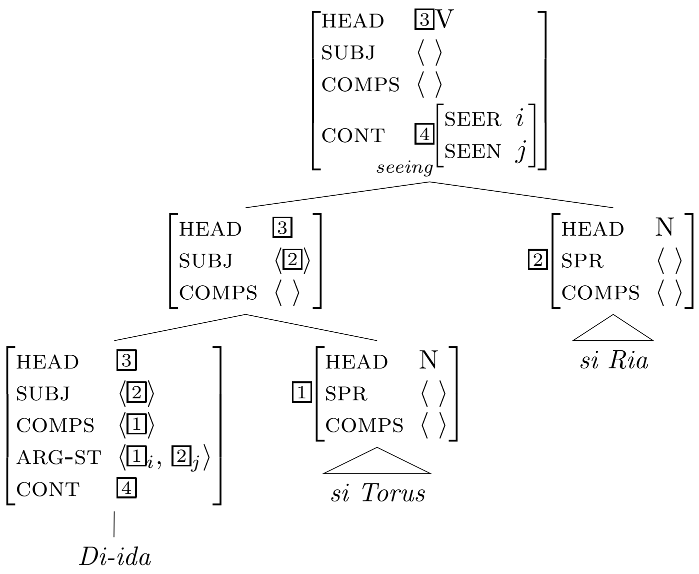
Conversely, the final NP in the examples above will be analyzed as a VP-external subject. This NP behaves similarly to the *ang*-marked NP in Tagalog. It may optionally be fronted before the verb in questions or as a topic, while the VP-internal NP may not be. Further, as in Tagalog, relativization is restricted to this NP, and following the Keenan-Comrie (1977) hierarchy, if only one NP can be relativized on, then that NP is the subject. Moreover it is this VP-external subject NP that must be the controllee, regardless of the verbal voice:

(14) a. Mang-elek si Bill si John [man-uhor biang __]
AV-persuade PM Bill PM John AV-buy dog
'John is persuading Bill to buy a dog.'

b. Mang-elek si Bill si John [di-pareso doktor __]
AV-persuade PM Bill PM John OV-examine doctor
'John is persuading Bill to have a doctor examine him.'

This suggests an analysis of (12b) as in (15), and lexical entries for the verbs in (12a) and (12b) as in (16a) and (16b), respectively.

(15)



(16) a. $\begin{bmatrix} \textsc{phon} & \langle \text{mang-ida} \rangle \\ \textsc{subj} & \langle \boxed{1} \rangle \\ \textsc{comps} & \langle \boxed{2} \rangle \\ \textsc{arg-st} & \langle \boxed{1}\text{NP}_i,\ \boxed{2}\text{NP}_j \rangle \\ \textsc{cont} & {}_{\textit{seeing}}\begin{bmatrix} \textsc{seer} & i \\ \textsc{seen} & j \end{bmatrix} \end{bmatrix}$ b. $\begin{bmatrix} \textsc{phon} & \langle \text{di-ida} \rangle \\ \textsc{subj} & \langle \boxed{2} \rangle \\ \textsc{comps} & \langle \boxed{1} \rangle \\ \textsc{arg-st} & \langle \boxed{1}\text{NP}_i,\ \boxed{2}\text{NP}_j \rangle \\ \textsc{cont} & {}_{\textit{seeing}}\begin{bmatrix} \textsc{seer} & i \\ \textsc{seen} & j \end{bmatrix} \end{bmatrix}$

However, despite this clear evidence for phrase structure and grammatical relations, reflexive binding is insensitive to this structure. Reflexivization shows that an a-subject can bind a non-a-subject (and not vice versa) regardless of the verbal voice of the sentence (Sugamoto 1984):

(17) a.	[Mang-ida	diri-na]	si	John
	AV-saw	self-his	PM	John
	'John$_i$ saw himself$_i$.'			

b.	*[Mang-ida	si	John]	diri-na
	AV-saw	PM	John	self-his
	*'Himself$_i$ saw John$_i$.'			

(18) a.	*[Di-ida	diri-na]	si	John
	OV-saw	self-his	PM	John
	*'Himself$_i$ saw John$_i$.'			

b.	[Di-ida	si	John]	diri-na
	OV-saw	PM	John	self-his
	'John$_i$ saw himself$_i$.'			

To account for these reflexivization patterns using a surface structure based notion of command would mean suggesting that the phrase structures of the sentences in (17) and (18) are radically different. But all available evidence indicates that the phrase structure is the same despite the changing verbal voice. On the other hand, these facts just fall out of the HPSG theory of binding. For instance, although the NP *si John* does not c-command the reflexive in (18b), it nevertheless a-commands the reflexive – the structure of this example is identical to (15). Thus these data provide startling support for defining binding theory on a level of argument structure that is distinct from both surface phrase structure and valence lists.[7]

5.3 Generating different linking patterns

Finally we will discuss briefly how the various different patterns of mapping between argument structure and the valence lists can be licensed. We can seek to explain both the commonality of types like intransitive verbs and transitive verbs across all languages and the systematic link-

[7] See further Manning (1996) and Wechsler and Arka (to appear) for arguments and evidence that these binding facts cannot be explained simply by reference to a thematic hierarchy, as is argued for in Schachter (1984) and Sugamoto (1984). See also Davis (1996) and Davis and Koenig (1996) for general evidence against the use of thematic hierarchies.

ing differences between syntactically ergative and accusative languages through the use of multiple inheritance within a hierarchical lexicon (Pollard and Sag 1987, Riehemann 1993). That is, in a syntactically accusative language, a transitive verb will say both that it is transitive, and that it obeys an accusative linking pattern, and so on. A partial presentation of some necessary types is presented in (19). Note in particular that types like *intrans-v-lxm* and *trans-v-lxm* only specify the ARG-ST list of their type, and say nothing about the valence lists.[8]

(19) a. *verb-lxm*: $\begin{bmatrix} \text{CAT} & \text{V} \\ \text{SPR} & \langle\ \rangle \end{bmatrix}$

b. *subj-v-lxm*: *verb-lxm* $\wedge \begin{bmatrix} \text{SUBJ} & \langle\ [\ ]\ \rangle \end{bmatrix}$

c. *intrans-v-lxm* : *subj-v-lxm* $\wedge \begin{bmatrix} \text{ARG-ST} & \langle \text{NP}[\textit{core}] \rangle \oplus \textit{list}(\textit{obl-np}) \end{bmatrix}$

d. *trans-v-lxm* : *subj-v-lxm* $\wedge \begin{bmatrix} \text{ARG-ST} & \langle \text{NP}[\textit{core}], \text{NP}[\textit{core}], \ldots \rangle \end{bmatrix}$

e. *acc-canon-lxm*: $\begin{bmatrix} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & \mathsf{compression}(\boxed{2}) \\ \text{ARG-ST} & \boxed{1} \oplus \boxed{2} \end{bmatrix}$

f. *erg-canon-lxm*: *erg-canon-intrans-lxm* $\vee$ *erg-canon-trans-lxm*

g. *erg-canon-intrans-lxm:* $\begin{bmatrix} \textit{intrans-v-lxm} & \\ \text{SUBJ} & \boxed{1} \\ \text{COMPS} & \mathsf{compression}(\boxed{2}) \\ \text{ARG-ST} & \boxed{1} \oplus \boxed{2} \end{bmatrix}$

h. *erg-canon-trans-lxm:* $\begin{bmatrix} \textit{trans-v-lxm} & \\ \text{SUBJ} & \boxed{1} \\ \text{COMPS} & \mathsf{compression}(\langle \boxed{4} \rangle \oplus \boxed{2}) \\ \text{ARG-ST} & \langle \boxed{4} \rangle \oplus \boxed{1} \oplus \boxed{2} \end{bmatrix}$

Before, we suggested that a *canonical* lexeme is one where the valence lists 'add up' to the ARG-ST. We wish to maintain this intuition, but also to allow for recent work on the treatment of causatives and light verbs, and on the handling of syntactically ergative and Western Austronesian languages. In (19e–f), we therefore introduce a generalization

[8]We assume a division among the arguments of a verb into *core* and *oblique* arguments (Manning 1996). In most languages, all verbs have a subject, and so the language would make all verbs *subj-v-lxm*, but we allow for subjectless verbs in the initial verb type. The disjunction in (19f) appears necessary. In syntactically ergative languages, with intransitive verbs, the first argument on the ARG-ST list becomes the subject, whereas with transitive verbs, it is the second argument on the ARG-ST list that becomes the subject.

of this notion, whereby the SUBJ and COMPS lists are allowed to add up in certain constrained ways to a list that is the compression of the argument structure, that is, what it 'flattens out' to once we promote the members of its embedded lists to be on a par with the other list members, eliminating embedded PROs in the process. See Manning et al. (in press) for further discussion and exemplification.[9]

Not all languages consistently maintain the relationship whereby the ARG-ST list is the append of the SUBJ, SPR, and COMPS lists, *in that order*. Rather, in Western Austronesian languages, another ordering is possible, and indeed is unmarked. In this pattern, it is the second core argument of the ARG-ST of a transitive verb that becomes the SUBJ. In syntactically ergative languages, the unmarked relationship in Philippine languages is the only relationship possible for expressing transitive verbs (Dixon 1994, Manning 1996). These languages, and the Western Austronesian languages, provide strong support for two independent syntactic levels, realized in HPSG3 by the valence lists and ARG-ST, and provide crucial evidence for the argument structure based theory of binding that HPSG provides.

The relation between argument structure and the valence lists is handled by separate types (19e–f), which cross-classify with the arity or polyadicity types. A verb in a particular language will then inherit its subcategorization type, and one of the types in (19e–f). For an accusative language like English, a transitive verb would have a type like (20a), for a syntactically ergative language like Inuit, there would be a transitive verb type like (20b), while a Western Austronesian language like Toba Batak would allow both these constructions via a transitive verb type like (20c). The type in (20c) (along with verb-particular information) will then license the two Toba Batak signs that were shown in (16).

(20) a. *eng-trans-v-lxm*: *trans-v-lxm* $\wedge$ *acc-canon-lxm*

b. *inuit-trans-v-lxm*: *trans-v-lxm* $\wedge$ *erg-canon-lxm*

c. *toba-trans-v-lxm*: *trans-v-lxm* $\wedge$
($acc\text{-}canon\text{-}lxm \vee erg\text{-}canon\text{-}lxm$)

---

[9] compression can be defined as follows ('$\leftarrow$' designates 'only if'):

(i) $\mathsf{compression}(\langle\ \rangle) = \langle\ \rangle$.

(ii) $\mathsf{compression}(\langle PRO|Y\rangle) = Z \;\leftarrow\; \mathsf{compression}(Y) = Z$.

(iii) $\mathsf{compression}(\langle X|Y\rangle) = \langle X|Z\rangle \;\leftarrow\; X$ is a *synsem*, $\mathsf{compression}(Y) = Z$.

(iv) $\mathsf{compression}(\langle X|Y\rangle) = Z \;\leftarrow\; X$ is a *list*, $\mathsf{compression}(X) = X'$,
$\mathsf{compression}(Y) = Y'$, $\mathsf{append}(X', Y') = Z$.

5.4 Conclusion

We have argued that HPSG must draw a fundamental distinction between argument structure and the valence features which Borsley proposed, which distinguish grammatical relations. We have examined, unfortunately superficially, data from a range of languages to try to show that one can use this ARGUMENT-STRUCTURE list to considerable linguistic advantage. This in turn seems to alter the character of HPSG, by providing an important second kind of organization on the dependents of lexical heads. In particular, we discussed how theories of grammar that define binding on surface phrase structure configurations or surface valence lists are unable to satisfactorily account for binding patterns seen in Austronesian and ergative languages, or the binding patterns of 'subject-oriented' reflexives when they occur with passive or causative verbs. Following the reasoning laid out in slightly different terms in Manning (1996), we have argued that it is possible to give a universal characterization of binding in terms of our notion of argument structure that generalizes nicely over accusative and ergative languages, and that correctly predicts binding patterns with passive – and causative – verbs.

In the process of developing this account, we have been led to a number of more specific proposals about the nature of passives and argument realization or linking. A perspicuous way of formulating these proposals seems to be in terms of a small set of universally available types and constraints associated with them (also universal, we might hope). Although the ideas sketched here are preliminary, we hope that they can serve as a basis for subsequent HPSG research that will try to distill generalizations from seemingly diverse cross-linguistic patterns like these and to organize them into a tight system of universally available types and simple constraints. The recognition of argument structure as an independent dimension of grammatical organization seems to be an important first step to take in the realization of this goal.

References

Abeillé, Anne, Danièle Godard, and Ivan A. Sag. to appear. Two Kinds of Composition in French Complex Predicates. In *Complex Predicates in nonderivational syntax*, ed. Andreas Kathol, Erhard Hinrichs, and Tsuneko Nakasawa. New York: Academic Press.

Borsley, Robert. 1989. Phrase-Structure Grammar and the *Barriers* Conception of Clause Structure. *Linguistics* 27:843–863.

Bouma, Gosse, Robert Malouf, and Ivan A. Sag. 1998. Satisfying Constraints on Extraction and Adjunction. MS, Groningen University and Stanford University.

Copestake, Ann. 1992. The Representation of Lexical Semantic Information.

Dalrymple, Mary. 1993. *The Syntax of Anaphoric Binding*. Stanford, CA: CSLI.

Davis, Antony. 1996. *Linking and the Hierarchical Lexicon*. Doctoral dissertation, Stanford.

Davis, Antony, and Jean-Pierre Koenig. 1996. Lexical Semantics and Linking Constraints in the Hierarchical Lexicon. Paper presented at the 3rd International Conference on HPSG, Marseille.

Dixon, R. M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Emmorey, Karen. 1984. The Intonation System of Toba Batak. In *Studies in the Structure of Toba Batak*, ed. Paul Schachter. 37–58. UCLA Occasional Papers in Linguistics, Number 5.

Foley, William A., and Robert D. Van Valin, Jr. 1984. *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.

Grimshaw, Jane. 1990. *Argument Structure*. Cambridge, MA: MIT Press.

Jespersen, Otto. 1924. *Philosophy of Grammar*. London: George Allen and Unwin.

Johnson, David Edward. 1977. On Relational Constraints on Grammars. In *Syntax and Semantics Volume 8: Grammatical Relations*, ed. Peter Cole and Jerrold M. Sadock. New York: Academic Press.

Keenan, Edward L., and Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8:63–99.

Kroeger, Paul. 1993. *Phrase Structure and Grammatical Relations in Tagalog*. Stanford, CA: CSLI Publications.

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

Manning, Christopher D., and Ivan A. Sag. 1995. Dissociations between Argument Structure and Grammatical Relations. Tübingen HPSG workshop, http://www.sultry.arts.usyd.edu.au/cmanning/papers/tuebingen.ps.

Manning, Christopher D., and Ivan A. Sag. submitted. Argument Structure in HPSG: Modeling valency alternations and binding. *Nordic Journal of Linguistics*.

Manning, Christopher David, Ivan A. Sag, and Masayo Iida. in press. The Lexical Integrity of Japanese Causatives. In *Readings in Modern Phrase Structure Grammar*, ed. Robert Levine and Georgia Green. Cambridge: Cambridge University Press.

Meurers, Detmar. 1995. Towards a Semantics for Lexical Rules as used in HPSG. Paper presented at the Conference on Formal Grammar, Barcelona, Spain. Revised version available at http://www.sfs.nphil.uni-tuebingen.de/∼dm/.

Miller, Philip H., and Ivan A. Sag. 1997. French Clitic Movement without Clitics or Movement. *Natural Language and Linguistic Theory* 15:573–639.

Perlmutter, David M. 1984. The Inadequacy of Some Monostratal Theories of Passive. In *Studies in Relational Grammar*, ed. David M. Perlmutter and Carol G. Rosen. 3–37. Chicago, IL: University of Chicago Press.

Pollard, Carl, and Ivan A. Sag. 1987. *Information-Based Syntax and Semantics*. Stanford, CA: Center for the Study of Language and Information.

Pollard, Carl, and Ivan A. Sag. 1992. Anaphors in English and the Scope of Binding Theory. *Linguistic Inquiry* 23:261–303.

Pollard, Carl, and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, IL: University of Chicago Press.

Pullum, Geoffrey K. 1997. The Morpholexical Nature of English *to*-Contraction. *Language* 73:79–102.

Riehemann, Susanne. 1993. Word Formation in Lexical Type Hierarchies: A Case Study of *bar*-Adjectives in German. Master's thesis, Tübingen.

Sag, Ivan A. 1998. Explaining the Conjunct Constraint. Paper presented at the annual meetings of the Linguistic Society of America. New York.

Sag, Ivan A., and Janet D. Fodor. 1994. Extraction Without Traces. In *West Coast Conference on Formal Linguistics*, 365–384. Stanford Linguistics Association.

Sag, Ivan A., and Danièle Godard. 1994. Extraction of *De*-Phrases from the French NP. In *Proceedings of the North Eastern Linguistics Society, 24*, ed. Mercè Gonzàlez, 519–541.

Sag, Ivan A., and Christopher D. Manning. forthcoming. Lexicalizing quantifier scope: A response to Pollard and Yoo. MS, Stanford University and The University of Sydney.

Schachter, Paul. 1976. The Subject in Philippine Languages: Topic, Actor, Actor-Topic or None of the Above. In *Subject and Topic*, ed. Charles N. Li. 491–518. New York: Academic Press.

Schachter, Paul. 1984. Semantic-role-based Syntax in Toba Batak. In *Studies in the Structure of Toba Batak*, ed. Paul Schachter. 122–149. UCLA Occasional Papers in Linguistics, Number 5.

Shibatani, Masayoshi. 1988. Voice in Philippine languages. In *Passive and Voice*, ed. Masayoshi Shibatani. 85–142. Amsterdam: John Benjamins.

Sugamoto, Nobuko. 1984. Reflexives in Toba Batak. In *Studies in the Structure of Toba Batak*, ed. Paul Schachter. 150–171. UCLA Occasional Papers in Linguistics, Number 5.

Webelhuth, Gert. forthcoming. Causatives and the Nature of Argument Structure. In *Complex Predicates in non derivational syntax*, ed. Andreas Kathol, Erhard Hinrichs, and Tsuneko Nakazawa. New York: Academic Press.

Wechsler, Stephen. this volume. HPSG, GB, and the Balinese Bind. Stanford, CA: CSLI Publications.

Wechsler, Stephen, and I Wayan Arka. to appear. Syntactic Ergativity in Balinese: an Argument Structure Based Theory. *Natural Language and Linguistic Theory*.

Xue, Ping, Carl Pollard, and Ivan A. Sag. 1994. A New Perspective on Chinese *ziji*. In *Proceedings of the West Coast Confererence on Formal Linguistics*, 432–447. Stanford, CA. CSLI Publications.