

An other take on the notion Subject¹

Dik Bakker
University of Amsterdam

Anna Siewierska
University of Lancaster

TO APPEAR IN: M. Hannay & G. Steen (eds)

The English clause: Usage and structure.

Amsterdam : John Benjamins.

1. Introduction

Subject is both one of the most controversial notions in linguistics and the one most often taken for granted. In traditional linguistics it is typically inherited from Latin grammar, and either seen as the constituent that has Nominative case and agreement on the verb. More fundamentally, it is seen as the most central constituent of the clause, which also plays a role in some more complex syntactic constellations, such as equi NP deletion. In the respective versions of generative theory, Subject is typically treated as a ‘deep’ syntactic position, on which different constituents may land via the application of transformational rules. Relational Grammar (Perlmutter 1982) shares this double deep-to-surface aspect of grammatical relations, i.e. Subject and Object, be it that in this case there are links with semantic and pragmatic aspects of the structure of the clause.

As we will see in more detail below, functional approaches to linguistic theory, and more specifically Functional Grammar (FG), Role and Reference Grammar (RRG) and Relational Typology (RT), take slightly different positions in the area of grammatical relations. What they seem to share, however, is that Subject is seen as a surface syntactic function correlating to but assigned independently from underlying pragmatic and semantic functions and features. Under the guise of notions such as ‘Perspective’ (FG), ‘Privileged Syntactic Argument’ (RRG) and ‘Pivot’ (RT), these theories define Subject indeed as a selfstanding syntactic function assigned by the speaker to the most central constituent of a clause. It is chosen more or less independently from other semantic and pragmatic features of the constituents to which the grammar in question limits the choice. *Mutatis mutandis*, this state of affairs is not completely unlike the position of Subject in formal syntax.

We think that this view of a more or less freely assigned Subject appears to work only for a restricted number of languages, mainly from Western Europe, West Africa and New Guinea. It is almost exclusively in these linguistic areas that languages may be found which have passives with the presence of Agents, be it optional and typically in the form of an oblique. Such a constellation, with all arguments present may in its turn be interpreted as the result of a grammar induced choice between the arguments for the central position. Outside the linguistic areas mentioned passives tend to be obligatorily agentless. However, most of the other morphosyntactic phenomena traditionally related to Subjects are to be found in all areas, and are therefore in need of an explanation as well.

In this contribution, we will propose a less grammaticalized and more dynamic take on the notion Subject, not unlike Givón’s (1997) multifactor approach, which we think may be applied to the vast majority of the languages of the world including those mentioned earlier on. We will go about our job in the following fashion. First, in section 2, we will try to show that even in a grammar of English, a language which is often seen as having grammaticalized Subject to the extreme, there are other factors that determine which of the constituents in a clause is going to be Subject and that it is not just a free choice between the arguments by the

speaker. We will present our observations in terms of one of the leading functional theories to date, Functional Grammar (cf. Dik 1997). Within this theory, Mackenzie (1998) is one of the first contributions drawing attention to the study of sentences in discourse in a principled fashion. Our conclusions, however, are general and might be embedded in any functionally oriented theory. In section 3 we will sketch the outlines of a dynamic, multifunctional framework for the description of Subject-like phenomena which has precisely this aim. Some complex and problematic cases will be discussed in section 4. And in section 5 we will draw some final conclusions.

2. Subject in a grammar of English

In general terms the Subject in an English sentence is the constituent involved in a number of morphosyntactic phenomena. Following Keenan (1976), we divide them into the coding and behavioural properties of Subjects. Under coding properties we find case marking (typically Nominative or Absolutive for Subjects; Accusative and Ergative for Objects) and agreement marking on the verb (typically, the marker varies for Person, Number and Gender features of the Subject constituent). Behavioural properties are syntactic phenomena under control of Subjects. We mention anaphoric binding as expressed in reflexive pronouns; conjunction reduction; equi NP deletion; raising; and the relativized constituent in relative clauses, among others. Finally there is constituent order, which may be both under control of Subjects and employed as a device to mark them. In Indo-European languages it is often the case that all these phenomena, to the extent that they are relevant for a language in the first place, are related to one and the same constituent. English is a case in point. So, in example (1a) *she* has Nominative marking and controls the verb agreement and the empty Subject position in the coordinated clause. Example (1b) shows that the passive construction may reverse the situation, promoting the Patient argument to the Subject position, where it takes over all the coding and behaviour properties of the Agent in (1a).

- (1) a. She_i.NOM approaches_i Jack_k.ACC and Ø_i kisses_i him_k / *Ø_k kisses_k her_i
 b. Jack_k.NOM is_k approached by her_i.ACC and *Ø_i kisses_i him_k / Ø_k kisses_k her_i

In a rendition of the grammar of English which is completely focussed on linguistic competence, it could be stated that Subject may be assigned to Agents and Patients, and also to Recipients, but not to e.g. Instruments or Locatives, as shown in (2a-c).

- (2) a. She.REC was sent a Valentine by a one of her pupils
 b. *The hammer.INSTR was opened the door by him
 c. *Home.LOC is stayed by her

Functional Grammar (Dik 1997:267) proposes a language dependent but universally valid Semantic Function Hierarchy for the assignment of Subjects. It takes the following shape:

- (3) Semantic Function Hierarchy (SFH):

Agent > Patient > Recipient > Beneficiary > Instrument > Locative > Temporal

For each language there is a cut off point. To all semantic functions to the left of this point Subject may be assigned; to those on the right it may not. For English the cut off point on the

SFH is between Recipient and Beneficiary. In other languages it is located more to the left (e.g. for Dutch after Patient) or to the right (e.g. for Sundanese after Beneficiary and for Malagasy after Locative). Thus, Subject may – and can – be assigned to any constituent with a semantic function to the left of the cut off point, irrespective of its further semantic and pragmatic properties, or other aspects of the clause in question. The decisive criterion for the choice of Subject, according to FG, is an independently working principle called *Perspective*. In other words, speakers choose their subjects within the relevant limits of the SFH in order to present the contents of the clause from the perspective of the referent in question.

Such a rendition of the Subject assignment rules does not seem to reflect, however, linguistic reality, in other words: it does not comply with linguistic performance, and how speakers actually operate. Compare the acceptability of the sentences in (4a-d).

- (4) a. She bought a new bike.
 b. ???A new bike was bought by her
 c. ?A falling stone hit her
 d. She was hit by a falling stone

Of both pairs (4a/b) and (4c/d), although all versions might be considered well-formed in terms of competence based rules, only the ones with the pronominal Subjects seem acceptable from a performance point of view in the sense that they might be expected to appear in normal discourse rather than as isolated examples in a linguistic discussion of English. In fact, utterances such as (4b) and (4c) are very rare in corpora of spoken English; (4a) and (4d) seem to be the natural, and in fact: the only options. A search through the spoken section of the British National Corpus (BNC; Aston & Burnard 1998) confirms that speakers select their Subjects on the basis of their semantic, pragmatic and discourse properties rather than on the basis of a more or less independent operation such as *Perspective*. Firstly, passives are rare in the spoken language. Moreover, passives with explicit agents are rarer still. Svartvik (1966), in a corpus of written English, found that over 80% of the passives were agentless. For a corpus consisting of a mixture of written and (formal) spoken English, Thompson (1986) found almost exactly the same. The nature of the corpus selection process did not allow us to inspect all candidate sentences for agentless passives. Therefore, we can only more or less impressionistically state that passives with agents are just a fraction of all passives to be found in the spoken corpus, probably considerably less than 15% of them. It may be concluded then that the major reason for speakers of English to opt for the passive is the absence of an agent since it is unknown, irrelevant or obvious, i.e. for discourse pragmatic reasons.² According to Thompson (1986) it is also for discourse pragmatic reasons that speakers of English opt for a passive despite the presence of an agent. The non-agent (i.e. the second argument) will be made Subject when “(it) is more closely related than the agent to either the theme of the paragraph or to a participant in the immediate preceding clause”. Obviously, this may be interpreted in terms of the different types of discourse and sentence topics.

That topicality, and pragmatics in general is an important parameter in the determination of Subjects in English may be gathered from the following data from the BNC. In the section with spoken English we found 236 utterances which contained both a form of *be* and the preposition *by*. Of these, 71 were analyzed as non-passives or in fact agentless passives. The 165 agented passives that remained may be characterized as follows in terms of the form of the two arguments.

Subject	Agent ('by')	number of occurrences
Pronominal	Pronominal	14 (8.5%)
Pronominal	Nominal	102 (61.8%)
Nominal	Pronominal	2 (1.6%)
Nominal	Nominal	47 (28.5%)

Table 1. Distribution of types of passive Subjects and Agents

So, the Subject is a pronoun in over 70% of the cases, the Oblique Agent is pronominal in only 10% of the cases, which is indicative of the overall topicality of the referent in the Subject position. In almost two thirds of the cases, we find the combination of a pronominal Subject and a nominal Agent. When both referents are pronominal all cases but one follow the person hierarchy in (5) below.

(5) Person Hierarchy

$$1^{\text{st}} > 2^{\text{nd}} > 3^{\text{rd}}$$

The only counterexample has a relative pronoun for its 3rd person Subject in which case there is no real choice at all. The two sentences which have a nominal Subject and a pronominal Agent can be found in (6) and (7) below.

(6) and will continue to, to, to, erm so that the, the, the set-up should never have been created by hér. [PS527]

(7) We said that come from the film Cats and he reckoned the copy of Memories he's got is sung by who? [PS0FX]

In both cases the Subject is topical while the Agent is focal, and therefore prefers sentence final position. Finally, of the 47 nominal pairs, the majority have a definite Subject and an indefinite Agent. However, in 8 cases the Subject is indefinite while the Agent is definite. Of these, 2 have a human Subject while the Agent is non-human. They are given in (8) and (9).

(8) On their births a trust a fund of tens of thousands of pound was started for them in their names by their grandmother. This will guarantee all their school fees and the basis of their future. [PS0FG]

(9) And the planet earth is within the reach of gravitational pull and things on the earth are attracted by the moon. [KPAPS000]

In (8) the money is going to be the new topic for the following stretch of discourse. In (9), the earth has been introduced as a topic, and is continued by the related sub-topic in the Subject position of the second clause while *the moon* – semantic function Force, not Agent – is in focus.

From the above discussion, and from earlier work by authors such as Thompson (1986) a picture of Subject in English emerges which is much more embedded in the

dynamics of discourse then the purely sentence grammar oriented treatment in standard Functional Grammar and other theories. In the next section we will try and sketch a more encompassing framework for the treatment of Subject in English, and languages in general.

3. Towards an encompassing framework for Subject

The ingredients for our general approach to Subject stem from two domains of linguistic description: the functional and the formal one. Under the functional domain we subsume all relevant semantic and pragmatic features, such as the respective semantic and pragmatic functions, and notions like animacy, definiteness and number. In FG, they pertain to the underlying representation of a sentence, and are typically aspects of referring elements, such as argument and adjunct positions. Although their role may be different for different languages, they are supposed to be drawn from a universal set and to be subject to the respective universal hierarchies that have been proposed in the typological literature, as illustrated in (3) and (5) above by the semantic function and person hierarchies. Under the formal domain we subsume all morphosyntactic phenomena which are controlled by an argument or adjunct, and which were subdivided in coding, behaviour and constituent order in section 1. Also in this case we assume that a maximum set may be determined from which only a subset may be relevant for a specific language. The Subject related morphosyntactic phenomena (MSP) are the explananda whereas the explanation rests on the relevant set of pragmatic and semantic factors (PSF). In short:

$$(10) \text{ PSF}_k \rightarrow \text{MSP}_k$$

where PSF_k and MSP_k are the relevant subsets for language k , and the arrow should be read ‘control’.

We suggest the following interpretation of Subject. Arguably, all languages have MSP which are controlled via one or more arguments or adjuncts. When for a language it can be shown that at least two basic semantic functions, such as Agent and Patient, can each control the relevant set of MSP the control function will be seen not as semantic but syntactic. Subject is the label belonging to this unified syntactic function. We will say that the relevant semantic functions are *neutralized* to Subject. In order for this neutralization to hold in some language, and Subject to be relevant in the first place it should be shown on independent grounds that the semantic functions in question play an independent role elsewhere in the grammar of that language. E.g. they may have different case markings when they are non-subjects, trigger different grammar rules, etcetera.³ Thus, looking at the examples in (11) below from Acehnese, spoken in Indonesia, we would not assume Subject to be relevant for this language.

(11) Acehnese (Austronesian; Durie 1985)

- a. Gopnyan geu-tém jak/*geu-jak.
 3SG 3-want go/*3-go
 ‘He wants to go.’
- b. Gopnyan geu-tém *hët/geu-hët.
 3SG 3-want *fall/3-fall
 ‘He wants to fall.’

In (11a) the subordinate clause has obligatory equi-deletion of its (single) argument, thus rendering the clause infinite. In (11b), however, equi-deletion is forbidden: the argument is

obligatorily present, resulting in a finite subordinate clause. The decision on the expression of the argument hinges crucially on the semantic function of the argument of the embedded verb: Agents get no expression, and Patients are expressed via a clitic. Since this agreement rule is based on individual semantic functions rather than on a neutralized set of them, there seems to be no reason to posit a Subject function for Acehnese. We will call such a language role dominated.⁴

On the other hand, there are languages in which it seems to be pragmatic functions which control some of the MSP under discussion. The example in (12) below, from Mongsen, a dialect of Naga Ao, is a case in point. Topical constituents undergo conjunction reduction, irrespective of their semantic function. Such a language does not seem to be in need of an independent Subject function either.

(12) Mongsen (Tibeto-Birman; Coup, 2003)

t(e)	ai-t(e)-p(e)ti	a	hl(e)p-(e)r			
thus	NRP-be.big	one	wrap-SEQ			
n(e)-ni	m(e)-thu?	t(e)si	a?	pi	t(e)-s(e)la	
2SG.POSS	NEG-reach	until	just	PROX	PROX-untie	
t(e)-(e)r	aj	ok	h(e)m(e)t-(e)r	ai	oj	z(e)k
thus-SEQ	PRT	PRT	hold-SEQ	PRT	PRT	send.PAST

'Thus having wrapped a big one [a bundle], [Lichabai said to Mechatseng]
 "Don't untie [this] before you reach your wife."
 And then, taking hold [of it, i.e. the bundle], [Lichaba] sent [him]'

In this short stretch of discourse, a bundle is introduced with Patient function in the first line. In the second line explicit mention of it is left out again from the Patient position. In the third line, all three topical referents are left out from the respective Agent and Patient positions. Languages such as Mongsen will be called pragmatically dominated.

However, in most languages the state of affairs seems to be less clear cut. The relevant MSP are neither fully role dominated nor fully pragmatically dominated, but controlled by a set of semantic and pragmatic properties of the clause or sentence. In these languages control over MSP is assumed to be syntactic, and Subject to be the function assigned to the argument, or adjunct which exerts this control. The determining semantic and pragmatic factors could be interpreted both as motivating and as constraining factors for Subject assignment in a specific language. We distinguish the following four subdomains for these factors:

- a. Semantic functions
- b. Pragmatic functions
- c. Semantic and formal features of relevant argument and adjunct positions
- d. Other pragmatic, semantic and formal aspects from the context

First, there is the set of semantic functions which can be neutralized under Subject assignment. They may be found on the Semantic Function Hierarchy as given under (3) above. As indicated earlier, we take into consideration individual semantic functions rather than argument positions, as in Functional Grammar. Therefore, the following two examples suffice for a language to have Subject. In example (13a) below control over the equi deleted

argument from the second conjunct is exerted by the Agent intransitive argument of the first conjunct, in (13b) the same is done by a Patient intransitive argument.

(13) Dutch

- a. Vader_{AG} sprong en viel
father.AG jump.PAST and Ø.PAT fall.PAST
'Father jumped and fell.'
- b. Vader viel en stond weer op
father.PAT fall.PAST and Ø.AG stand.PAST again up
'Father fell and stood up again.'

Obviously, Subjects are relevant for all languages where neutralization takes place over both arguments of a transitive predicate, typically accompanied by a passive or antipassive construction, as in examples (14) from Dyirbal.

(14) Dyirbal (Australia; Dixon 1994: 13f)

- a. Yabu nguma-nggu bural-n banaga-n^yu
mother.ABS father-ERG see-NONFUT return-NONFUT
'Father saw mother and she returned.'
- b. Nguma bural-nga-n^yu yabu-gu banaga-n^yu
father.ABS see-ANTIPASS-NONFUT mother-DAT return-NONFUT
'Father saw mother and returned.'

In (14a) the Patient argument is zero marked, located in sentence initial position and controls conjunction reduction. In (14b) the same holds for the Agent argument.

The second factor are the pragmatic functions. Also in languages which are not fully pragmatically dominated, they play a central role in the assignment of Subjects. Topics are so central in the organization of sentence structure that Subject is often seen as grammaticalized topic (cf. Givón 1983). It is quite unlikely that Topic will not turn out to be a more or less central element for the assignment of Subjects in the majority of languages, as shown for the English corpus examples in section 2. Focus seems to a less determining factor for Subjecthood crosslinguistically. Although focal constituents, such as Q-words may be fronted in many languages, this is typically either just a marked word order variant, or a special construction such as clefting. The other MSP typically do not apply, as in the case of passivization.

The third factor are the semantic, and arguably also formal, properties of the relevant argument and adjunct positions. In the English corpus examples, we saw that the person hierarchy under (5) was instrumental in promoting 1st and 2nd person referents – pronominal almost by definition⁵ - to Subject position at the cost of 3rd person ones. In the case of two nominal arguments, definiteness and humanness turned out to be decisive factors. These factors play a role in many languages. Sometimes they seem to be grammaticalized to the extent that the alternative is ungrammatical. This is shown in example (15) from Navajo. Another manifestation of the role of such factors are the so-called inverse constructions in languages like Nocte, as shown in example (16).

(15) Navajo (Athapaskan; Witherspoon 1980:5)

- a. At'ééd to yo-odlaq

- girl water ACT-drink.PAST
'The girl drank the water.'
- b. *To at'ééd bo-odlaq
girl water PASS-drink.PAST
'The water was drunk by the girl.'

(16) Nocte (Tibeto-Burman; DeLancey 1981:641)

- a. Nga-ma ate hetho-ang
1SG-ERG 3SG.MASC.ABS teach-1SG
'I will teach him.'
- b. Ate-ma nga-nang hetho-h-ang
3SG.MASC-ERG 1SG-ABS teach-INV-1SG
'He will teach me.'

In Navajo, non-animate can not be Subjects when there is an animate alternative, as in the case of (15b). In (16a) we have agreement with the first person agent. In (16b), where the first person is found in the Patient slot, we nevertheless have agreement with the first person, be it that the verb is marked for non-default Subjecthood.

In this third category we may mention as a factor the nominal versus pronominal status of the referent. We saw that the vast majority of English Subjects in our BNC sample are pronominal. In Cree, the use of the inverse seems to be based on the pronominal vs. nominal form of the arguments of transitives, at least to some extent. Compare examples (17a) and (17b).

(17) Cree (Algonquian; Wolfart 1973⁶)

- a. Ni-sekih-a atim.
I-scare-*DIR* dog
'I scare the dog.'
b. Ni-sekih-iko atim.
I-scare-*INV* dog
'The dog scares me.'

However, we will assume that it is not so much the formal expression of these arguments themselves that play a direct role in the choice of Subject, but their topicality. This factor then leads to pronominal rather than nominal expression of the referent. So it is not the form which triggers the Subject rule. Form and Subject assignment are both determined by deeper pragmatic (and semantic) factors.⁷

The fourth, and final factor instrumental in assigning Subjects is mainly found in languages with a split alignment system, more specifically Accusative-Ergative. The choice for either of these systems is typically based on semantic aspects of the relevant arguments, mainly the prototypical Agent and Patient positions. However, in some of these languages other factors play a role in this decision. This is illustrated by an example from Georgian, in (18) below.

(18) Georgian (Kartvelian; Comrie 1978)

- a. Student-i ceril-s cers.
student-NOM letter-ACC write.PRES

- ‘The student writes the letter.’
 b. Student-ma ceril-i dacera.
 student-ERG letter-ABS write.PAST
 ‘The student wrote the letter.’

In this language, the choice for either of the alignment subsystems is based on Tense, among other factors⁸ So, in (18a), with present tense, the alignment is Accusative, while in the past tense, illustrated in (18b), the alignment is Ergative. Indirectly, this has a bearing on the choice of Subject. While the relevant criteria for Subject assignment in these examples – case marking and constituent order - conspire in (18a), rendering the Agent Subject, they are in competition in (18b), leaving us the choice of either abandoning the notion here, or developing a criterion to choose. The latter will be our choice, as discussed in section 4 below. We will call a factor like Tense in Georgian contextual, since it does not directly emanate from the referents in the relevant argument positions.

This general sketch of the factors determining Subject assignment may be summarized as in figure 1 below. In this figure we find, on the left hand side the languages for which all MSP are controlled by individual semantic functions. Acehnese was a potential example of such a language. On the right we find the languages where pragmatic functions control the relevant phenomena. Tibeto-Birman languages like Mongsen may provide an example of these. In the middle we find the languages – arguably the vast majority - where the MSP are controlled by a set of factors drawn from the four domains mentioned, i.e. the semantic and pragmatic functions plus the semantic features of the arguments and contextual aspects.

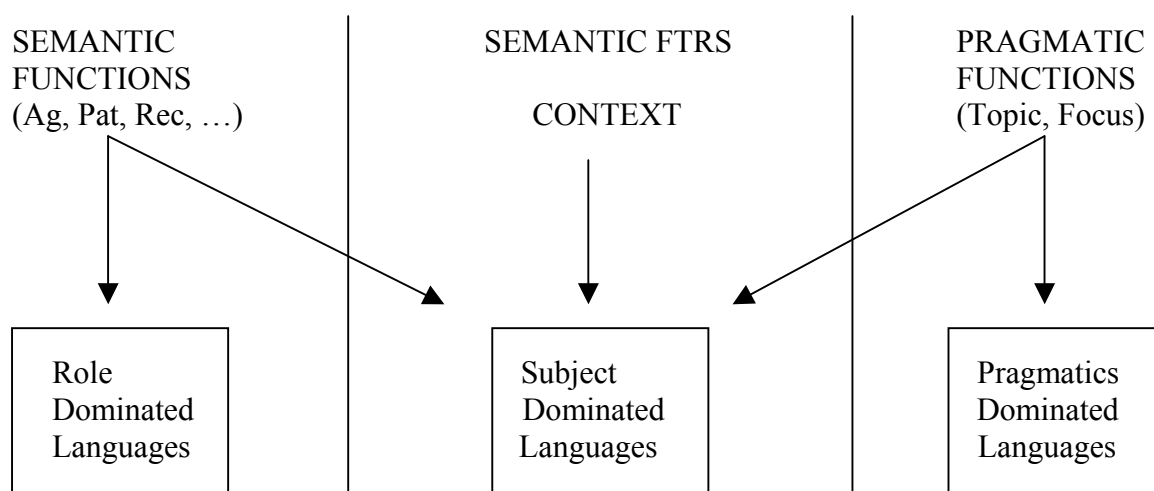


Figure 1. The Subject assignment domain

Figure 1 just represents the domains from which the factors for Subject assignment derive in a static fashion. In practice, these factors should be seen both as the motivating principles for Subject choice and the constraints on it. Within a language, competition between all these principles leads to a specific subset of wellformed constellations, i.e. an instantiation of figure 1 on the basis of concrete values for the respective parameters. Constant competition may lead to diachronic changes in these instantiations.

In the next section we will have a brief look at some more complex cases of Subject assignment, allowing us to fill in some details of the very general picture sketched above.

4. Some complex cases of Subject assignment

Sofar it has been tacitly assumed that the relevant morpho-syntactic phenomena are relevant uniformly for all Subjects in a language, including their feeding into rules such as passivization. However, this is not necessarily the case. There exist a large number of different splits, incongruencies and restrictions in the area of Subject assignment within and across languages. Given the scope of this article, we will not make any attempt at completeness, and discuss only a few of the more well-known ones, but which we think are sufficient to support our general view.

A first example are the so-called split ergatives. As already shown in the Georgian example in (18) above, languages may have several parallel ways of organizing grammatical relations. Not surprisingly, the factors that determine the choice between these systems stem from the same four domains which were mentioned as instrumental for Subject assignment in the first place. So, in Georgian, with Accusative alignment as in (18a), both case marking and constituent order seem to indicate the Agent as the Subject. However, in (18b) case marking would favour the Patient but constituent order would indicate the Agent. Rather than abandoning the Subject notion in such cases, we propose a hierarchy of factors to establish the best candidate. It looks as follows.

(19) MSP hierarchy (first version)

Morphological coding > Constituent Order

So, we suggest that, in case of conflicts case marking and agreement marking are better candidates for establishing Subjecthood than constituent order at the clause level. As shown in Siewierska et al. (1997), as well as in many other places, clause level constituent order is without any doubt the most flexible of the traditional order parameters, and seems therefore to be the least reliably and diachronically stable of the respective domains of Subject diagnostics. Also, opposed to morphological marking, constituent order is not a device that may or may not be developed by languages for a certain purpose. It is trivially always there, and may or may not be employed for certain communicative goals.

As a second case, we will have a brief look at inverse languages, such as Nocte, as introduced above. Taking examples (16a/b) for our only data, and looking at the relevant morphosyntactic aspects, we have both the arguments for Subject candidates in (16a). The Agent has verb agreement and is fronted, the Patient is unmarked. The same holds for (16b). However, the Agent is in the first position, but agreement and zero marking now go to the Patient. Disregarding syntactic control structures for the sake of argument, we would consider the Agent for Subject in (16a) and the Patient in (16b). Not only do these constituents have two out of three features in their respective sentences, they also control the strongest of the morphological features, i.e. verb agreement, which has a syntactic long distance aspect which is absent in zero marking. Thus, Nocte follows the same pattern as Georgian.

A third case, and which is extensively studied in Van Valin & LaPolla (1997:356f), is Icelandic. Consider the following examples.

(20) Icelandic (Indo-European; Zaenen et al 1985)

- a. Ég hjálpa-ð-i þeim.
1SG.NOM help-PAST-1SG 3PL.DAT
'I helped them.'
- b. þeim va-r hjálp-að af mér.
3PL.DAT be.PAST-3SG help-PASTPRT by 1SG.DAT

- ‘They were helped by me.’
- c. Þeim líka-r matur-in-n og borð-a mikið.
 3PL.DAT like-3SG.PRES food-DEF-SG.NOM and eat-3PL.PRES much
 ‘They like food and eat much.’

Icelandic has the usual Subject features, as shown in (20a). Subjects take the (unmarked) Nominative case; its Person and Number are marked on the verb. Icelandic has a regular passive, which promotes the Patient-like argument to Subject position, and demotes the Agent, to the extent that it is present in the first place, to an Oblique. In both cases, the Subject controls conjunction reduction. In the case of transitives with a dative second argument, which should probably be interpreted as a Recipient rather than a Patient, passivization has a rather different set of side effects however, as shown in (20b). Indeed, in such cases, the Dative argument is fronted, and the Agent gets Oblique status. However, the potential new Subject does not neutralize its case to Nominative, and does not copy its Person and Number features onto the verb. The latter takes 3rd Person marking, the form typically found in so-called impersonal passives across languages. Interestingly, however, conjunction reduction is controlled by this dative argument, as shown in (20c). From our perspective, this is enough to consider the Dative argument a Subject, be it a weaker version than the one in the Active case.⁹ The marking of the verb is taken care of by a default rule. Dative marking takes place by the same rule that takes care of it in active contexts. It is simply not neutralized as in the case of Accusative arguments being assigned Subject.

We think there are good reasons to consider ‘long distance’ syntactic control such as conjunction reduction and equi deletion a firmer proof of Subjecthood than more local coding phenomena. In the light of this then, we extend our hierarchy of (19) above to the form in (19’).

(19’) MSP hierarchy (extended)

Syntactic Control > Morphological coding > Constituent Order

We could say then that in Icelandic, Subjects do not neutralize the case marking of non-Patients nor do they mark them on the verb. However, they will be fronted and also have control over conjunction reduction. In other words, they feed only partially into passivization, and will therefore still be seen as Subjects, be it weaker ones.

Things are, however, more complex than this. This could be shown on the basis of some more examples from Dyirbal. Consider the following.

(21) Dyirbal (Australian; Dixon 1994:12f)

- a. N^yurra ngana-na bura-n.
 2PL.NOM 1PL.ACC see-NONFUT
 ‘You saw us.’
- b. Nguma yabu-nggu bura-n.
 father.ABS mother-ERG see-NONFUT
 ‘Father saw mother.’
- c. N^yurra ngana-na bura-n banaga-n^yu.
 2PL.NOM 1PL.ACC see-NONFUT return-NONFUT
 ‘You saw us and we returned.’
- d. Ngana bural-nga-n^yu n^yurra-ngu banaga-n^yu.
 1PL.NOM see-ANTIPASS-NONFUT 2SG.DAT return-NONFUT

‘We saw you and we returned.’

When looking only at (21a) and (21b), Dyirbal manifests itself as a split ergative language, with the (pro)nominal state of the relevant arguments as the determining factor to choose between the subsystems, much in the way of Cree in (17) above. However, in the light of more complex constructions, we may have to reconsider this view. Although there may be no doubt about Subjecthood in (21a), in the case of conjoined clauses as in (21c), conjunction reduction turns out to be controlled by the Patient rather than the Agent. Under application of the Antipassive, control over conjunction reduction would go to the Agent, as in (21d). Note that this is effected by the demotion of the default Subject, the Patient, to a postverbal position, where it is marked by the Dative rather than the Accusative. We could say that default Subject assignment in Dyirbal is to the Patient in all cases.¹⁰ In the case of two non-topical – and thus nominally expressed – arguments the case marking follows the Ergative alignment, with an unmarked Patient which has control over conjunction reduction. The antipassive shifts both to the Agent, which is then the Subject. In the case of two topical – and therefore pronominal – arguments the default assignment of Subject is again to the Patient. However, case marking follows a semantic pattern, more or less like the case of Icelandic dative predicates in (20). The effect of the Antipassive leaves the marking of the Agent unaffected, but it assumes the Subject function now. The Patient gets the oblique Dative case. Thus, we have Subject effective at the highest of the three domains, and case marking being determined partially by semantic functions, i.e. in the case of topical arguments. The order of the arguments may be fully determined on the basis of the semantic functions. The Agent is always in the first position, independent of its Subjecthood, and the Patient is preverbal when a Subject and postverbal when not.

A final example we would like to discuss is from another Australian language, Yidin^y. According to Dixon (1977), this language has ergative alignment for nominal NP’s, and accusative alignment for pronominal ones. However, other than the case of Dyirbal in (21), both the Agent and the Patient may control conjunction reduction. This is exemplified in (22) below.

(22) Yidin^y (Dixon 1977:390)

Ngayu	bama	bandāR	wandāN.
1SG.NOM	person.ACC	follow	fall
‘I followed the person and I/he fell.’			

This utterance unites the unmarked arguments from the two alignment systems, i.e. both Subject candidates. Indeed, each of them may be interpreted as the controller of the reduced Subject position in the second conjunct. Obviously, there will hardly ever be real ambiguity for hearers in these cases: the most topical of the two referents will be the natural controller. In a linguistic model restricted to the sentence level, such cases remain unsolved.

We are convinced that this does not exhaust by far the possible constellations relevant for Subject assignment in the languages of the world. However, our framework might be a start for rethinking the treatment of Subjects in a functionally oriented theory.

5. Conclusions

In this contribution, we have tried to sketch a general framework for the assignment of Subjects within languages. We started out with spoken data from the British National Corpus. We showed that, in the case of passives with agents, the choice of Subjects was never a ‘free’ choice of the speaker. In all attested cases the choice seemed to be based on pragmatic, and to

a lesser extent, semantic factors of the relevant arguments. Extending our observations to other languages, we summarized the four domains from which these factors stem. However, a static, synchronic picture is an abstraction which may describe but not explain what Subjects in fact are. Thus, we think that choice of Subject is part of the dynamic, performance aspect of a language, based on a complex set of pragmatic and semantic factors rather than a more or less compulsory but speaker controlled choice among two or more candidates. Interestingly, since the rules for the selection of Subjects seem to operate mainly stochastically rather than in a yes/no fashion, they work in two ways vis à vis competence. On the one hand, they create constraints on what would logically be possible within the limits set by the grammar in terms of the Semantic Function Hierarchy and other devices. In that sense they work as a filtering device, be it a functional rather than a formal filter. Competition between the respective forces together with other properties of the respective grammars lead to different outcomes for different languages. Native speaker observations concerning the limits to Subject assignment in some language which are based on individual sentences-out-of-context are typically interpreted as what competence has to say about Subject. On the other hand, these forces in competition tempt speakers to cross the boundaries of competence, and to produce utterances which strictly would be considered unwellformed. Examples from Dutch of such transgressions are found in (23) and (24) below.

(23) De reiziger-s word-en verzocht uit te stappen.
 DEF passenger-PL AUX.PASS-PL request-PASTPRT get.off
 ‘Passengers are requested to get off the train.’

(24) Die broek pas ik niet!
 DEM pair.of.pants fit.1SG 1SG not
 ‘Those pants do not fit me.’

In (23), Subject is assigned to the Recipient, which is normally outside the bounds of the Dutch SFH. However, (23) is frequently heard spoken by conductors on trains. Given the topicality of passengers in such a context, and the contrast between the concreteness of the referent of the Recipient and the abstractness, and total unfitness for Subjecthood of the Patient argument, this form has become a fully acceptable utterance in such contexts. It is corrected only by school teachers and their likes. In (24), the same features which are found to be active in Subject assignment, such as Person and Animacy seem to be in the process of bringing about a reinterpretation of the predicate scheme of the verb *passen* ‘to fit’. In its standard dictionary reading, the first argument would be the Patient (here: the pair of pants) and the second one the Experiencer (here: the first person). However, younger speakers of Dutch will typically reverse the roles, and interpret the Experiencer as the first argument, and therefore make it the default Subject, while the original meaning is maintained.¹¹ Without making performance factors central, there would be no way apart from acquisition errors to explain these, and many other diachronic changes. And without taking performance, and discourse into consideration it is probably impossible to explain and describe what speakers do when linguists say that they assign Subject.

Abbreviations

1	first person
2	second person
3	third person
ABS	absolute

ACC	accusative
ACT	active
AG	agent
ACC	accusative
ANTIPASS	antipassive
AUX	auxiliary
BEN	beneficiary
DEF	definite
DEM	demonstrative
DIR	direct
ERG	ergative
FEM	feminine
INSTR	instrument
INV	inverse
LOC	locative
MASC	masculine
NEG	negative
NOM	nominative
NONFUT	nonfuture
NRP	\$\$\$
PASS	passive
PAST	past tense
PASTPRT	past participle
PAT	patient
POSS	possessive
PRES	present
PROX	proximate
PRT	particle
REC	recipient
SEQ	continuation marker
SG	singular

References

- Aston, G. and L. Burnard (1998). *The BNC Handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.
- Comrie, B. (1978). 'Ergativity'. In W.P. Lehmann (ed.) *Syntactic typology: studies in the phenomenology of language*. Austin: University of Texas Press.
- Coup, A.R. (2003).
- DeLancey, S. (1981). \$\$\$
- Dik, S.C. (1997). *The Theory of Functional Grammar, part I*. Berlin: Mouton de Gruyter.
- Dikker & xxx (200x), \$\$\$
- Dixon, R.M.W. (1977). *A grammar of Yidin*. Cambridge: Cambridge University Press.
- Dixon, R.M.W. (1994). *Ergativity*. Cambridge: Cambridge University Press.
- Durie, M. (1985). *A grammar of Acehnese*. Dordrecht: Foris.
- Givón, T. (1983). 'Topic continuity in discourse: an introduction'. In T. Givón (ed.) *Topic Continuity in Discourse: Quantitative Cross-Language Studies*. Amsterdam: John Benjamins.

- Givón, T. (1997). 'Grammatical Relations: An Introduction'. In T. Givón (ed.) *Grammatical Relations*. Amsterdam: John Benjamins, 1-84.
- Hetzron et al (1980). \$\$\$
- Keenan, E. (1976). 'Towards a universal definition of "Subject"'. In C. Li (ed.) *Subject and Topic*. New York: Academic Press, 61-77.
- Kibrik, A.E. (1997). 'Beyond Subject and Object: toward a comprehensive relational typology'. *Linguistic Typology* 1-3, 279-346.
- Mackenzie, J.L. (1998). The basis of syntax in the holophrase. In: M. Hannay and A.M. Bolkestein (eds) *Functional Grammar and Verbal Interaction*. Amsterdam: John Benjamins, 267-296.
- Perlmutter, D.M. (1982). 'Syntactic representation, syntactic levels, and the notion of Subject. In P.I. Jaconbson and J.K. Pullum (eds) *The Nature of Syntactic Representation*. Dordrecht: Reidel, 283-340.
- Siewierska, A. (1998). 'Passive-to-Ergative versus Inverse-to-Ergative'. In A. Siewierska & J.J. Song (eds) *Case, Typology and Grammar*. Amsterdam: John Benjamins, 229-246.
- Siewierska, A. (2004). *Person*. Cambridge: Cambridge University Press.
- Siewierska, A. and D. Bakker (2004). 'Three takes on grammatical relations: a view from the languages of Europe and North and Central Asia'. *Proceedings International Symposium on the Typology of Argument Structure and Grammatical Relations*. Kazan May 2004.
- Siewierska, A. & D. Bakker (forthc). 'The Agreement Cross-Reference Continuum: Person marking in FG'. In K. Hengeveld & C. de Groot (eds) \$\$\$\$\$\$.
- Siewierska, A., J. Rijkhoff & D. Bakker (1997). 'Appendix - 12 word order variables in the languages of Europe'. In A. Siewierska (ed.) *Constituent Order in the Languages of Europe*. Berlin: Mouton de Gruyter, 783-812.
- Svartvik, J. (1966). *On Voice in the English Verb*. The Hague: Mouton.
- Thompson, S.A. (1986). 'The passive in English: a discourse perspective'. In R. Channon and L. Shockey (eds) *Festschrift for Ilse Lehiste*. Dordrecht: Foris.
- Van Valin, R.D. and R. LaPolla (1997). *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.
- Witherspoon G. (1980). 'Language in culture and culture in language'. *International Journal of American Linguistics* 46, 1-13.
- Wolfart, H.C. (1973). 'Plains Cree: A grammatical study'. In *Transactions of the American Philological Society* n.s. 63, pt. 5.
- Zaenen, A., J. Maling and H. Thráinsson (1985). 'Case and grammatical functions; The Icelandic passive'. *Natural Language and Linguistic Theory* 3, 441-483.

¹ These ideas were presented first at the International Symposium on the Typology of Argument Structure and Grammatical Relations in Kazan, may 2004. The authors wish to thank Nick Smith from the University of Lancaster for providing them with the corpus data discussed in section 2. We also wish to thank the editors and an anonymous reviewer for suggestions which lead to important changes in the original version.

² Givón (1979:59) assumes that missing agents are always recoverable by the hearer, and therefore left out by the speaker, at least in terms of the type of referent.

³ Here, we follow Dikker & \$\$\$ (200\$), who argue for establishing the set of relevant semantic functions in a language on language internal grounds rather than assuming that the set of semantic functions is universal and given a priori for any language. A test for relevance of a tuple of semantic functions is particularly crucial for functions which typically share the same argument position, and are therefore mutually exclusive, such as Agent and Force.

⁴ RRG (Van Valin 1996:274f) would call Acehnese a language with a semantic rather than a syntactic pivot.

⁵ In some languages the form referring to Speaker and Addressee may be of a nominal nature. E.g. in Thai speakers use the word *phǒm* 'hair' in self-reference when speaking to a superior, and *tua* 'body' for the Addressee when this is someone closely related, e.g. a spouse (cf. Siewierska 2004:228). It remains to be seen whether these behave like pronouns or like nouns in the corresponding languages.

⁶ As quoted in Siewierska (1998:241).

⁷ This is fully in compliance with our view on the expression of arguments in a functional theoretical framework. Cf. Siewierska & Bakker (forthc).

⁸ See Dixon (1994:104f) for a number of complex conditions on Ergativity.

⁹ Such examples from Icelandic were already discussed by Hetzron et al (1980).

¹⁰ In that sense Dyirbal is a 'deep ergative' language.

¹¹ Older speakers would interpret (11) as a refusal to try and put on the pants, since there is a second meaning of the same verb i.e. 'see whether it fits' with a more prototypical Agent-Patient argument structure.