Jadwiga Linde-Usiekniewicz
Uniwersytet Warszawski

V

The Encoding Grammar and Syntax

5.1. The Elusive Nature of Syntax

One of the characteristics of syntax that the encoding grammar has to address is its elusiveness coupled with its being a fundamental and intrinsic feature of the language. Even those linguists who would not agree with Chomsky's claim that there is a narrow faculty of language, and it is syntactic, i.e. concerned with linking elements to create larger wholes and can be even further reduced to recursivity [Hauser, Chomsky, Fitch 2002], have to agree that there could be no language without syntax, i.e. linking linguistic signs together. In addition, at least some facets of syntax have to be independent of semantics: otherwise there would be no way of judging some sentences as grammatical but absurd from the semantic point of view, as is the case of the famous Chomskyan

(5.1) *Green colorless ideas sleep furiously.*

In terms of Encoding Grammar, if syntax were to depend absolutely on semantics, and in particular on semantic structure (cf. Chapter III. *The Encoding Grammar And Semantics*) sentences like (5.1.) would never be produced. There would be no means of creating other nonsense texts, such as Jabberwocky . In addition, there would be no room for zeugmas, and semanticists would not be able to use the contradiction tests, because absurd and self-contradictory semantic representations would simply be unencodable. Since all these utterances can be produced it means that syntactic patterns available within a language are general enough to be independent of lexical semantics. More evidence comes from the fact that there is no one-to-one correspondence between semantic relations and syntactic ones, as it was mentioned in *2.2. The Architecture of the Encoding Grammar*, with reference to adjectives and nouns in a modifying construction.

On the other hand, it would be absurd to claim within the same framework that syntax is completely independent of semantics. Nevertheless, there is no contradiction involved between the two claims and the Encoding Grammar actually is able to deal with this double character of syntax in a very traditional way: syntax should be considered fairly independent of lexical meaning, but strongly related to sentence meaning, since it is an element of the encoding means the language possess. It would be wrong however, to say that syntax is completely independent of lexical meaning, since it is considered that valence is a semantic

property of a lexical unit. Nevertheless, when lexical units present themselves to syntax during the encoding procedure, they do not exhibit all their semantic features but only those that are syntactically relevant, i.e. their syntactic slots. Thus it is possible to establish large classes of verbs on the basis of their having the same number of syntactic slots of the same character, the way it has been done by for English verbs by Levin [1993], and was proposed, even earlier, within the MTM model, by Apresjan [1976], among others. On the other hand, lexical units sharing the same lemma (i.e. part of their signifier) are rightly distinguished by their syntactic properties, i.e. the number of syntactic slots and constraints on what can fill those slots. Some examples of that were presented in 3.4.1. *Properties of Lexical Units*, and particularly in examples (3.14) – (3-16) and (3.18) – (3.21).

The relation between meaning and syntax changes if both are seen not as elements of language structure but as different representations of the same utterance being encoded. Here the procedure consists in making the semantic relations, present in the semantic representation, more explicit. Specifically, syntactic relations between elements of an encoded sentence reflect those semantic relations present in that sentence's semantic representation that are not conditioned by language structure, i.e. those that would be considered belonging to *parole*.

The question still remains how the Encoding Grammar would deal with the fact that it is possible to produce nonsense sentences. It would be wrong to claim they are produced through matching some vague, albeit appropriate, signifieds with wrong signifiers, in the way described by Ziff [1967] or in the way it actually happens to people with paraphasia, because the speakers who invent these utterances are perfectly aware of what they are doing. The explanation within the Encoding Grammar would be that in such cases speakers decide to encode meaning in blatant contravention to semantic structure of the language. In the case of (5.1) what is disregarded is the part of the semantic structure common to all languages that possess such lexical units as 'colorless', 'green', 'idea', 'sleep', etc., since the sentence can and has been, successfully rendered in many different languages. By contrast, zeugmas rely on specific items of semantic structure and may not be translatable.

The elusiveness of syntax is not limited to its complex relation to semantics, but resides in the very nature of the syntax itself. This point is better illustrated with contrast between the semantic representation, the syntactic representation and the (surface) morphological representation. While both the meaning of the sentence and its morphology are easily accessible to the language users, the syntactic representation is not. To elaborate this point, however, it is necessary to abandon temporarily the encoding perspective. Thus, language users can be asked what a sentence means and are able to produce equivalent paraphrases (this is what the MTM is all about). In more restrictive terms of discovery procedures permissible under the distributionalist regime, the researcher could ask their informer whether two expressions meant the same or not. Similarly, differences in morphological and phonological representations are observable, even if they are subtle enough. By contrast syntactic relations are not straightforwardly manifest and in order to trace them linguists resort to various procedures, such as reduction, substitution, and others, including several constituency tests mentioned in the previous chapter. It should be noted that these procedures, constituency tests included, are eminently external to the sentences under analysis: they all involve doing something very specific to a given sentence, thus producing a different sentence, and claiming that the initial sentence possess some structural feature by virtue of some feature being present in the resultant one. As the result, syntactic analysis relies more strongly on theoretical constructs than any other kind of linguistic analysis, and the validity of these constructs is subject to methodological and theoretical debate. To give just two outstanding examples: Mel'čuk's [1988; 2009] stand against phrase structure grammar

and the debate in Polish linguistics about subjecthood and the non-Nominative forms, e.g. [Przepiórkowski 2004], [Saloni 2005], [Menantaud 1999] among others.

Yet another example of the elusiveness of syntax is that, although syntactic relations are often expressed through morphology or morphophonology, these markings can be misleading. Three cases will be presented here. The first is based on Frayzyngier, Shay [2003: 116-127]. The authors say that languages have morphophonological means of marking phrase boundaries, and—by the same token—of marking different phrase structures. Their discussion focuses on Mina, a Central Chadic language, in which both lexical and grammatical morphemes have two forms: one with the underlying final vowel retained, and one with the underlying vowel deleted or substituted by shwa. The authors propose to analyze the occurrence of these two forms not as a phonological phenomenon (non-reduced forms being pre-pausal and the reduced forms being non-pausal), but as markers of what they consider the internal structure of the utterance [Frayzyngier, Shay 2003: 117]. However, the examples they discuss as evidence fall into two distinct categories in terms of syntax. In some of them the presence of the unreduced form truly indicates the boundary between constituents, and thus can be rightly claimed to encode syntax. This is the case of distinguishing between attributive syntactic relation between a noun and an adjective, where they form a single noun phrase, and a predicative relation, when even with the absence of copula, the noun and the adjective are two separate constituents.

(5.2a) Ńkwá féš
goat small
'The goat is small'

(5.2b) nkù báétə̀
goat large
'a large goat'

Mina [Frajzyngier, Shay 2003: 124, (original transliteration and glosses)]

In (5.2a) the noun 'goat' retains its underlying vowel, thus indicating a constituent boundary, while in (5.2b) the final vowel is reduced and the noun and the adjective form a single constituent.

Yet, in other examples the situation is not that straightforward and the examples can be analyzed as actually manifesting phonological properties of the sentences and not their syntactic properties. Thus while in (5.3a), taken from [Frajzyngier, Shay 2003: 125 (original transliteration and glosses)], the reduced form ɗə̀ signals that "[t]he object following the verb is part of the phrase to which the verb belongs" [Frajzyngier, Shay 2003: 125]

(5.3a) Kə́ ɗə̀ tìpíɗ̀ tsáy zà
INF gather termites finish AUX
'She finished looking for termites.'

in (5.3b) while the first two elements retain their underlying vowel, the third instance of the same verb does not and signals that the adjunct *wane* can also be considered a part of the verb phrase:

(5.3b) ɗǎ ɗǎ á ɗə̀ wane
fetch fetch 3SG pour a lot
'It rained a lot.'

Thus for these sentences, i.e. (5.3a, b), the only thing that can be claimed about the contrast between reduced and unreduced forms is that the former mark the verb as forming a

constituent or a phrase with the next element, without specifying the distinction between complement and adjunct.

However, even this claim may be too strong, since in the next pair of sentences from the same source, it would be difficult to postulate two different syntactic patterns [Frajzyngier, Shay 2003: 117; 124]

(5.4a)	Sə-lím-é	**gwáɮá**	mùmbùrkó
	1SG-see-GO	elephant	yesterday
	I saw an elephant yesterday		
(5.4b)	Sə-lím-é	**gwáɮ**	mùmbùrkó
	1SG-see-GO	elephant	yesterday
	I saw an elephant yesterday		

Concerning this pair the authors write that "final vowel retention on the element preceding the adverb codes it as being a separate phrase, and therefore in focus" (as in (5.4a)), while "the non-focused adverb has the preceding vowel deleted" (as in (5.4b)) [Frayzyngier, Shay 2003: 124]. Thus the difference between (5.4a) and (5.4b) encodes different theme-rheme division but does not affect constituency as such.

This is made even more clear when contrasted with Mel'čuk [1988: 28-29] comment about the dependency syntax being incapable of accounting for the semantic ambiguity of

(5.5) *He gave three talks about human rights in Chicago in 1982.*

which the author makes explicit through following glosses:

(5.6a) 'He gave three talks about human rights, and it happened in Chicago in 1982' [there were no other talks on human rights by him].

(5.6b) 'He gave three talks about human rights in Chicago, and it happen in 1982' [his talks on human rights in other places may have occurred at different dates'

Again, this is a semantic distinction, and Mel'čuk is right to link it with the theme-rheme division, but—to formulate it in terms of the Encoding Grammar—this semantic distinction would affect neither the form of the dependency tree, nor, contrary to what Mel'čuk claims, the form of the phrase structure tree, should either be taken as the syntactic representation of this sentence, and should the speaker wish to encode either meaning the way presented in (5.5) and to rely on phonology (pauses and intonation) to encode the theme-rheme division.

Chadic languages are not the only ones that present challenging examples of the elusiveness of syntax. Yet another instance can be found in this example from Kalkatungu [Blake 1983: 145, quoted by Hengeveld, Mackenzie 2008: 287-298 and by Rijkoff 2008: 525]

(5.7)	Cipayi	icayi	yaŋi	ṯukuyu	yauntu
	This.ERG	bite	white.man	dog. ERG	big. ERG
	'This big dog bit/bites the white man'				

Both Hengeveld, Mackenzie [2008] and Rijkoff [2008] follow Blake in analyzing this sentence as having all the elements bearing the ergative marker as not forming a single noun phrase. Rijkoff [2008: 525, original emphasis] also gives other possible permutations of this sentence:

(5.8) **Cipayi ṯukuyu yauntu** yaŋi icayi
Cipayi ṯukuyu icayi yaŋi **yauntu**
Tukuyu cipayi icayi yaŋi **yauntu**
Yauntu cipayi tukuyu icayi yaŋi
yaŋi icayi **cipayi yauntu ṯukuyu**

Both accounts quote Blake's [1983:145] opinion that "there are in fact no noun phrases, but … where an argument is represented by more than one word we have nominals in parallel or in apposition …Each word is a constituent of the Clause" [Blake 1983:145].

Here the ergative marking of all words referring to the dog may be interpreted as semantic only, inasmuch that they mark semantic relations between all the relevant words, without their belonging to a single phrase. This is not so unusual, instances of this can be found cross-linguistically in predicatives agreeing in gender and number, and even case, with either the subject or the object. Agreement in gender and number can be illustrated by Mel'čuk's [2009: 15] Russian examples given as (2.2) in Chapter II and repeated here:

(5.9) *Ja znala **ego** molod+**ym.***
'I (female) knew **him** (when he was) young (masculine singular)'
. *Ja znala **eë** molod+**oj.***
'I (female) knew **her** (when she was) young (feminine singular)'
*Ja znala **ix** molod+**ymi***
'I (female) knew **them** (when they were) young (plural)'

while instances of agreement of case between the predicative adjective and the subject in Polish is discussed in Derwojedowa, Linde-Usiekniewicz [2003: 39-40]. There it is argued that the nominative form of adjective, as in:

(5.10) *Jest inteligentna*
is intelligent-NOM.SG.F
'She is intelligent.'

is the result of agreement between the subject and the predicative, because the default case of predicative is the Instrumental, which occurs with nouns as predicatives, as in:

(5.10) *Jest nauczycielką*
is teacher-INSTR.SG.F
'She is intelligent.'

and, more importantly, in subjectless phrases with nominal predicate, as in:

(5.11)	*Być*	*piękną*	*to*	*marzenie*	*każdej nastolatki*
	be	beautiful	COP	dream	every teenager
		INSTR.SG.F			GEN.SG.F GEN.SG.F

'To be beautiful is a dream of every [female] teenager.'

Thus, if the multiple ergative forms in Kalkatungu do not mark belonging to single clause, but the semantic relation only, this language could indeed be considered a "flat language" (as

Blake [1983] calls it, i.e. non-configurational one.[1] However, at first glance all these permutations could be interpreted just as well as resulting from extremely free order in which modifiers can be separated from their heads, as seems the case of Latin poetry, as illustrated by one example from Vergil (Eclogue 1), quoted by Mel'čuk [1988: 36, original glosses]:

(5.12) *Tityre, tu patulae recubans sub tegmine fagi, silvestrem tenui*
Tityrus, you branchy resting under cover beech rural thin
musam meditaris avena
music play reed-pipe
'You, Tityrus, resting under the cover of branchy beech, play rural tunes on a reed-pipe

Yet another example of morphologically unambiguous sentences, nevertheless giving rise to syntactic ambiguity, can be found in Polish, among sentences with secondary predicates. Thus in

(5.13) *Jan pamięta ciotki młode*
John-NOM remembers aunt-ACC.PL young-ACC.PL
'John remembers [his] aunts [as] young [persons]'

taken from Derwojedowa, Linde-Usiekniewicz [2004:40] both the noun *ciotki* and the adjective *młode* are in the accusative case, which could lead to their interpretation as belonging to a single noun phrase with the adjective-noun order reversed to encode prominent rhemacity of the adjective. This is not however the case, since (5.13) bears no rhematic intonation on the adjective, and it can be interpreted as an instance of the secondary predicative (and not attribute) agreeing in case with the object.

5.2. Structure vs. Representation and Deep vs. Surface Syntax

In Chapter II, *The Encoding Perspective Embodied: The Encoding Grammar* a distinction was made between representation and structure in all layers or modules of the Encoding Grammar. This distinction has been elaborated on for semantics (*3.2 The Semantic Representation*; *3.3. The Semantic Structure*), and for the theme-rheme division (*4.5. The Theme-Rheme Division Defined Within the Encoding Grammar*). It has been argued that while the representation of the theme-rheme division is a part of the semantic representation of the utterance, the structure of such division is not limited to a single module, but its different elements, i.e. means of encoding some sentence elements as themes or rhemes, can be found in structures corresponding to different modules (in semantic structure, in syntax, in morphology and in phonology). Some of the instances of elements of this structure have been presented in the previous section: it has been argued that the distinction between the meanings presented as (5.6a) and (5.6b) would be encoded on (5.5) by appropriate intonation and that the difference between (5.4a) and (5.4b) concerned the presence of a theme-rheme division between the verb and the adverb in (5.4a) and the absence of such division in (5.4.b). Also in the previous section the notion of syntactic representation of the utterance was invoked, in contrast to the syntactic structure.

It has been also mentioned in Chapter II, particularly in *2.2. The Architecture of the Encoding Grammar*, that a two-level syntax needs to be postulated, i.e. a deep-syntactic

[1] Configurational and non-configurational languages will be discussed further on.

submodule and a surface-syntactic submodule. From that it follows that the Encoding Grammar needs to distinguish both representations and structures for each submodule (as presented in Figure 1.). In its two-level approach to syntax the Encoding Grammar follows the MTM model, and accepts the general rationale given by Melčuk [2004b: 248]:

> DSyntAs [=deep-syntactic actans] constitute an interface between SemAs [=semantic actants], determined mainly by semantic considerations (plus expressibility), and SSyntAs [=surface-syntactic actants], determined exclusively by syntactic considerations (distribution, word order, structural words, agreement and government, morphological forms, control of gerunds/reflexives, and the like, relativization, etc.). DSyntAs are called in to strike a balance and to find a compromise between these two types of considerations so as to facilitate the description of the correspondence Sem-Structures ⇔SSynt-Structures.

However, it should be kept in mind that within the Encoding Grammar the relation between the meaning and the surface syntax is unidirectional, and this would affect the way the deep syntax and the surface syntax are portrayed. The differences between the Encoding Grammar and the MTM concerning syntax will be dealt in the subsequent sections.

Interestingly, even linguists who work with the MTM, express doubts as to the necessity of the Deep Syntactic Structure within the model. Kahane [2009] argues that speakers do not select semantemes but signifying units, i.e. "deep linguistic signs whose combinations produces simultaneously a SemR [=semantic representation] and a SSyntR [=surface-syntactic representation". Thus the original deep syntactic structure becomes a derivation structure, i.e. "the structure recording how the rules of grammar had combined to derive a sentence"[ibidem].

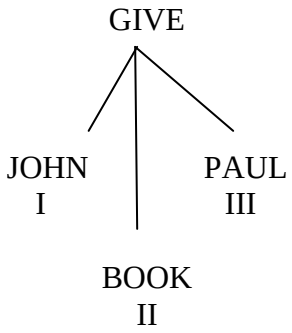
This solution, which seems to a degree to resemble the Minimalist Program, is not acceptable within the Encoding Grammar. There are two kinds of arguments for rejecting it. On the theoretical grounds, the Semantic Representation (which substitutes the MTM Semantic Structure in the approach proposed here) is not built of semantemes, i.e. signifieds of linguistic signs. Additionally, there are some features of syntactic representations of sentences that cannot be arrived at without the intervention of the deep-syntactic module and the deep-syntactic representation.

The first instance concerns the grammatical passive voice as opposed to lexical diatese. To illustrate this point, let us assume that the semantic representation to be encoded corresponds to the fact that John gave Paul a book, and the language in which it is going to be encoded is English. In the semantic representation we have all the necessary elements, including John's initial ownership of the book and the transfer of said ownership to Paul. The semantic representation of this sentence would also contain the information about semantic roles of each of the participant. (It is immaterial at this stage if these roles are identified as Source for 'John', Goal for 'Paul' and Theme for 'a book', accordingly to Fillmore [1968] or in terms or as Agent for 'John', Counteragent for Paul and Object for 'a book', in the MTM tradition.)

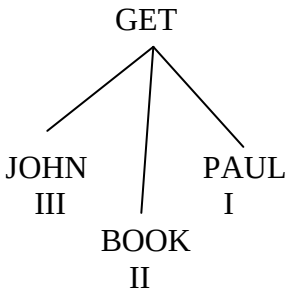
This meaning can obviously be encoded with the use of the verb *to give* and with the use of the verb *to get*, and the choice lies with the speaker. Thus first stage of encoding (by the deep syntactic module) would produce two different deep syntactic structures: one (5.14a) with the giver as the "deep subject"[2], indicated by I, 'book' as the object, indicated by II, and the receiver as the third participant, indicated by III; and the second (5.14b) with the receiver as the deep subject, the book as the deep object and the giver as the third participant.

[2] Deep-syntactic actants will be defined in the following section.

(5.14) a.



b.



While (5.14b) corresponds to

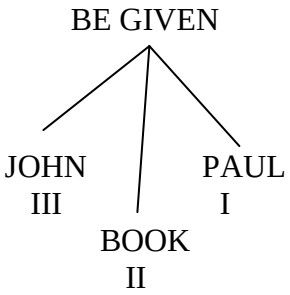
(5.15) *Paul got a book from John*.

(5.14a) obviously may produce several different surface representations:

(5.16a) *John gave a book to Paul.*
(5.16b) *John gave Paul a book.*
(5.16c) *Paul was given a book by John.*

among which the difference between (5.16a, b) on the one hand and (5.16c) on the other is crucial to the reasoning presented here. If it were not for the distinction between the deep syntactic representation and the surface syntactic representation, we would need a separate deep syntactic representation for (5.16c), i.e.

(5.17)



which would imply that the syntactic structure is built up not of lexemes but of specific wordforms, and the signifieds would have to be matched not to lexemes but to wordforms. In addition, there would be no way of accounting for the difference between lexical diatese, as illustrated by the choice of either the verb *to give* or *to get*, and the grammatical diatese, i.e. the voice. In many languages, including Polish and Romance, there is a difference in markedness between corresponding sentences with lexical and grammatical diateses, which further supports the thesis that should more operations be necessary to produce the eventual surface representation, more marked it becomes.

Another argument in support of both deep and surface syntactic representations can be made out of co-referential pronouns, already discussed in *2.2. The Architecture of the Encoding Grammar*. As it has been argued there, since their use relies on syntactic, i.e. surface subjecthood, as in the case of Polish reflexive possessive *swój*, or on the linear order

of the sentence (compare (2.15) and (2.16), they need to appear in the surface representation only.

Yet another evidence for the validity of the distinction between the deep-syntactic and surface-syntactic representations comes from the existence of cleft and pseudo-cleft sentences and the fact that in spite of their apparent copular structure, they are not identificational sentences[3] nor predicational copular sentences. While in ordinary identificational sentences the identity is a part of the semantic representation and the deep-syntactic representation, and in predicational sentences some property is attributed to some objects, cleft and pseudo-cleft sentences do not encode identity, but the theme-rheme division and except for expressing that division, both their semantic and deep syntactic representations are the same as these of their un-cleft counterparts. This distinction is borne out by the fact that that only pseudo-clefts exhibit "connectivity" [Iatridou, Varlacosta: 1988: 4], while identificational and predicational sentences do not[4]. The difference can be illustrated by their original pair of examples (original glosses):

(5.18a) *What John is is important to himself.*
[=John is important to himself]
(5.18b) *What John is is important to him.*
[=some property which John has is of importance to him]

A contrast between identity sentences and pseudo-cleft sentences is illustrated by the following Spanish pair, taken from [Moreno Cabrera 1999: 4291-4292]:

(5.19a) *La que viene es mi mujer.*
'Who is coming is my wife.'
Intended meaning: my wife is coming
(5.19b) *La que viene ha sido mi mujer.*
'lit.'who is coming has been my wife; the one that is coming has been my wife'.
Intended meaning: my ex-wife is coming

Moreover, in pseudo-clefts and clefts, if the cleft element corresponds to the object (or other complement) of its uncleft counterpart, it exhibits some object properties, i.e. case marking, as in Spanish (with the preposition *a* marking a personal direct object:

(5.20) *A quien vi es a Juan.*
'whom I saw was Juan.'

Several other facts supporting the claim that within the Encoding Grammar the procedure for producing cleft and pseudo-cleft sentences has to be a part of surface syntax, and not of deep syntax, can be mentioned[5]. One is that many languages need to mark the cleft elements in

[3] Unless one follows rigorously the framework proposed by Bogusławski [1977]. In his account, however, they would not be about ordinary identity, but about identity of concepts.
[4] Iatridou, Varlacosta [1988: 2] consider every sentence with "a free relative in on of the copular positions and a phrase in the other copular position modifying that free relative" a pseudo-cleft sentence, and distinguish specificational pseudo-clefts and predicational pseudo-clefts. In the terminology adopted here only specificational pseudo-clefts are considered pseudo-clefts, while predicational ones are considered straightforward copular sentences.
[5] Within different frameworks cleft and pseudo-cleft sentences are analyzed in various ways, however discussing these approaches would fall beyond the scope of this book.

cleft sentences by a pro-form, e.g. English, and French, among others, but the pro-forms are not necessary in ordinary copular sentences. This is of particular importance because if both cleft-sentences and pseudo-cleft sentences were in fact identificational sentences, a special explanation would be necessary as to why in one configuration with respect to the copula the pro-form is necessary and in the other it is not. Compare:

(5.21a) *My brother is the director here.*
(5.21a *The director here is my brother.*
(5.22a) *It is my brother who is the director here.*
(5.22b) *Who is the director here is my brother.*

Additionally, a special explanation would be necessary to account for the fact that some languages have only cleft-sentences, e.g. French, German, and no pseudo-clefts, since in identity sentences in these languages there is no corresponding restriction on which element can appear in the pre-copular position and which cannot.

In consequence, since the Encoding Grammar distinguishes between deep syntax and surface syntax, two different syntactic representations are postulated: the deep-syntactic one, somehow similar to what was proposed in (5.14) and the surface-syntactic one. For reasons explained in the two following section it is assumed that the deep-syntactic representation should be made explicit in terms of dependencies, while the surface-syntactic representation, at least for some languages, should be presented as a phrase structure. Correspondingly, it is postulated that the syntactic structure of a language is two-leveled: there is a deep-syntactic structure and a surface-syntactic structure.

For syntax the distinction between representation and structure is slightly different that it was claimed for semantics. While the semantic representation was fairly independent of the semantic structure and the Encoding Grammar task was just to overcome restrictions of the semantic structure by alternate encoding, this relation does not hold for syntax. Syntactic representation on either level cannot contain anything that is not present in the corresponding syntactic structure, because the sentence would be ungrammatical. In addition, the difference between the semantic structure, even expressed syntactically, and the syntactic structure of a language lies in the fact that the former only states what can be encoded within a particular language and what cannot, while the latter deals mainly with how it is encoded. Contrary to semantics, syntax does not work with signifieds only but with bilateral signs, although abstract ones: in particular the deep-syntax is concerned mostly with lexemes and abstract grammatical morphemes, while surface syntax deals with wordforms.

The main difference between the deep-syntactic structure and the surface syntactic structure lies in their relative universality. Thus the approach presented here broadly follows Mel'čuk [2004b: 249], who states about the deep syntactic structure (in his terminology) that "it must be cross-linguistically universal in that it uses a universal inventory of DSynt-relations" and that:

> .
> The DSyntS must be abstract enough to be valid for most or even all languages. Its formalism, used to reflect the syntactic organization of sentences, must be sufficient to represent whatever syntactic constructions the researcher may encounter.
> Therefore, the DSyntS should represent syntactic constructions of language L in the most general way, so that particular syntactic properties of an L's real sentence do not appear in its DSyntS. [Mel'čuk 2004b: 249],

However, it would be argued that languages do differ in the way semantic relations are converted to deep-syntactic relations, e.g. if the syntactic analysis of Kalkalungu as a

language without complex noun phrases presented above is right, this language would lack syntactic relations between nouns and adjectives and demonstratives.

By contrast, the surface-syntactic structures of different languages vary considerably, and are strongly language-specific.

5.3. Deep Syntax

5.3.1. Deep-Syntactic Actants

In its general approach to deep syntax the Encoding Grammar follows closely the MTM model, as presented by Mel'čuk [1988; 2004a, b; 2009]. However, several differences needed to be introduced to account for the fact that the Encoding Grammar is unidirectional, while the MTM actually matches the meanings with appropriate texts in a two-directional way. Thus the Encoding Grammar cannot rely on surface-syntactic properties of actants to define deep-syntactic actants (marked with Roman numerals in (5.14) and (5.16)), the way it is done for MTM [Mel'čuk 2004b: 254]:

> In accordance with their intermediate character, DSyntA-slots(L) are numbered as a function of two different set of properties: the relationships with the surface-syntactic actants and the relationships with semantic actants.
>
> *Syntactic properties of DSyntA-slots*. Roughly speaking, DSyntAslots(L) are numbered in the order of decreasing syntactic obliqueness with respect to L. The degree of obliqueness of a DSyntA is determined from that of the corresponding SSyntA. The latter reflects the hierarchy of SSyntAs, which is established through the analysis of their observable behavior […]

On the other hand, the approach presented here cannot rely on numbering of semantic actants the way it is done in [Mel'čuk 2004a: 43], because within the Encoding Grammar the semantic representation contains semantic roles and not numbered semantic actants. Within the MTM model numbered semantic actants are related to lexical units [Mel'čuk 2004a], while in the Encoding Grammar specific semantic roles are converted into more general deep-syntactic actants by the deep-syntactic module and can therefore be manifest in the syntactic representation, and not the semantic one. As it is widely known that the semantic roles do not correspond directly to surface syntactic elements (subjects, objects, etc.) and thus an intermediate level is necessary in which different semantic roles can be converted into deep-syntactic ones. The Encoding Grammar cannot postulate converting the semantic roles directly into deep-syntactic actants for two reasons. One comes from the well-known fact that the number of semantic roles is much greater than that of possible syntactic actants. Yet there is no consistent many-to-one way in which large groups of semantic roles correspond to deep-syntax ones. Thus depending on the lexical meaning of the verb, its deep subject can be agentive, experiencer, patient, etc., while for other verbs the same semantic roles appear as other deep-syntactic element. Examples abound, both cross-linguistically [cf. Haspelmath 2001: dative vs. nominative encoding of the experiencer] and within a single language. The latter case can be illustrated by the existence of conversives, but not exclusively. Another instance of different semantic roles being expressed by the same syntactic element can be found among many polysemous verbs, as in the following text-book example:

(5.23) *Mary baked the cake.*

contrasted with

(5.24) *Mary baked the potatoes.*

and with

(5.25) *The potatoes have to bake for more than 30 minutes.*

which appear in [Ruppenhofer *et al.* 2006: 8]. This phenomenon is not however restricted to polysemy and converses, but can be illustrated by simply contrasting sentences with the same surface syntactic pattern, but with different lexical verbs, e.g. by contrasting (5.23) with

(5.26) *Mary ate the cake.*

The second reason is related to conversives alone. As it is generally known, conversives are defined by the fact that they differ in terms of syntactic configuration of their semantic actants, compare *give* vs. *get* and Polish *uderzyć* 'to hit' vs. *oberwać* 'to be hit'. Thus within the Encoding Grammar the lexical units need to be described not only in terms of semantic roles of their actants but also in terms of their deep-syntactic roles. This is further borne out by the fact that some lexical units display alternate ways of converting their semantic actants into deep-syntactic ones [Mel'čuk 2004a: 42, original examples with syntactic markings removed]:

(5.27) *supply food to the peasants // supply peasants with food*
load hay on the truck // load the truck with hay
spray paint on the wall//spray the wall with paint

Thus the semantic roles of actants need to be distinguished from their deep-syntactic roles and the deep-syntactic roles need to be defined independently of semantic ones. Yet, as it has been said above, within the Encoding Grammar or within any encoding oriented perspective, deep-syntactic roles cannot be defined with reference to corresponding surface elements without the definitions being circular. Such circularity would obtain if it were said that deep-syntactic actant I, i.e. "deep subject" is the element of the deep-syntactic structure that is encoded as surface subject if no additional processing, e.g. passivization" occurs, and at the same time it were said that "deep subjects" are encoded as surface subjects unless some additional processing takes place.

One way in which the circularity can be avoided is to define deep-syntactic actants in terms of division into themes and rhemes. This has been proposed by Holvoet [Holvoet, Semėnienė 2005; Holvoet 2009a, b], who adopted for this purpose Bogusławski's [1977] approach with its multiple, hierarchical theme-rheme division and accordingly defined deep subject as the first order theme, the deep object as the second order theme etc. In its original form this solution cannot be used, since the Encoding Grammar approach to the theme-rheme division cannot follow Bogusławski [1977] in its understanding of themes, rhemes and the division of sentences along this dichotomy, as it has been argued in the previous chapter. However, it can be adapted, substituting the notion of themes of different order by the abstract notion of likelihood to become a theme. The resulting set of tentative definitions is presented below:

(5.28) 1st deep-syntactic argument = „deep subject"
- most likely to become the theme,
- most likely to be separated from the rest by the TR

boundary

2nd deep-syntactic argument = „deep direct object"
- second most likely to become the theme,

3rd deep-syntactic argument = „deep indirect object"
- third most likely to become the theme,

4th deep syntactic argument = "deep oblique object"
- fourth most likely to become the theme,

Several comments are necessary. First of all, it should be noted that although the Encoding Grammar defines themes and rhemes in complementary terms, the proposed definitions do not contain reference to rhemes. One reason is that here we are dealing with structure and not with representation and it assumed that the structure of the language deals with themes and rhemes in different ways. The second reason is that the rhemacity is reserved for distinguishing deep-syntactic circumstantials (or adjuncts) from deep-syntactic arguments: circumstantials are defined within the Encoding Grammar as those elements that are as likely to become themes as they are likely to become rhemes. Thus the verb complements that correspond to semantic actants tend to be encoded as arguments and tend to belong together with the verb either to the rheme or to the theme of the sentences. This is of particular importance when the question whether an element is an argument or a circumstantial concerns an obligatory semantic participant which can remain unexpressed [Mel'čuk 2004a: 12]. Thus, the question test attributed to Jarmila Panevová[6], which shows that an adjunct can be unknown while a semantic argument cannot, as in:

(5.29) *A: John has arrived.*
B : Where has he arrived ?
*A : *I have no idea.*

(actually taken from [Recanati 2007]) does not work for some arguments which correspond to optional semantic slots [Mel'čuk 2004a: 31-32]. Thus in the case of *to die of cancer* [ibidem] (5.29) would provide evidence for its being an adjunct or a circumstantial and not an argument

(5.30) *A: John died.*
B : Of what ?
A : I have no idea.

However, by virtue of its unlikelihood of becoming a theme, contrasted by its likelihood of becoming a rheme, as evidenced by the fact that it would tend to stay in place as a part of the verb phrase and not be separated from it, unless by a very marked procedure, e.g. by clefting in languages that allow it, it can be argued that it is actually an argument.

Secondly, from the way the deep-syntactic arguments are defined it does not necessarily follow that deep subjects will emerge as themes. Actual division into themes and rhemes is the property of the semantic representation of a sentence, while deep-syntactic actants are property of semantic and syntactic structure of the language. However, in cases in which the distribution of thematic and rhematic elements in the semantic representation matches the likelihood of becoming thematic or rhematic elements the encoding is straightforward and the resulting surface representation is not a marked one. By contrast,

[6] I have been unable to trace the exact reference.

should there be a mismatch between the likelihood of becoming a theme related to the deep-syntactic actants of the verb and the thematic character to be encoded, a marked surface representation is produced. As it has already been mentioned, the degree of markedness of the surface structure is intuitive, and corresponds to the complexity of processing necessary to achieve a given surface structure, thus passive sentences are more marked than active ones.

Thirdly, it should be noted that for the first argument not only themacity, but also a theme-rheme boundary has been invoked. This part of the proposal is the most tentative and its role has to do with the fact that on the one hand the Encoding Grammar wants its definition of arguments do be as cross-linguistically universal as possible, in agreement with [Mel'čuk 2004b: 249] quoted above, and on the other hand it needs to account in some ways for the fact that at least in some languages, there is one argument that tends to be separated from the rest of the sentence or at least from its deep-syntactic governor, i.e. the verb, by the theme-rheme boundary and that element is in normal course of events the surface subject. Though it is possible to think that this feature of the first argument is restricted to so called configurational languages only, or it is a feature of a deep-syntactic structure of both configurational and non-configurational languages that is not further elaborated in encoding by the latter ones.

Fourthly, it should be noted that although the deep-syntactic relations should be as universal as possible, the way semantic relations can be encoded as deep-syntactic relations vary from language to language. Thus a language may encode the Experiencer as the deep-surface indirect object, as in Polish, or as the possessor, i.e. without the appropriate deep-syntactic slot, as in the English gloss

(5.31) *Połamią mu nogi*
They will break to-him legs
'They will break his legs'.

In Polish the possessive configuration is impossible:

(5.32) **Połamią jego nogi.*

if the sentence refers to a living human being and/or the legs are still attached to the body[7], while in English one may come across instances of double encoding, i.e.

(5.33) *They will break his legs for him.*

Although the proposal presented here seems to match best such languages that first of all tend to separate the Subject from the Verb Phrase and tend to encode themes either by intonation or by syntactic devices that would result in fronting the theme, interestingly, some indirect evidence in favor of it comes from languages that mark either themes or rhemes by morphological means. One of such languages is Tariana, described in [Aikhenvald 2003], which has a suffix *–nuku/-naku* that marks topicalised non subjects (as opposed to ordinary non-subject) and a suffix *-nhe/-ne* for rhematic subjects [Aikhenvald 2003: 139ff]. Thus, within the framework proposed here it is possible to claim that *–nuku/-naku* indicates that a deep non-subject (i.e. an argument unlikely to be a theme) is nevertheless encoded as a theme, while *-nhe/-ne* signals that deep subject (i.e. the argument most likely to become a theme) is encoded as a rheme. Similar case can be made for all instances in which there is a special

[7] (5.32) is not anomalous if the legs have been amputated or when talking about a dead body. Cf. *Połamali nogi zamordowanej na kawalki i wrzucili do zsypu* 'They broke the murder victim's legs into pieces and threw them into the waste disposal chute'.

morphological marking, either direct or indirect, on a topicalized deep object [Nikolaeva 2001; Dalrymple, Nikolaeva 2011].

5.3.2. Configurationality and Non-configurationality in Deep Syntax

The issue of configurational vs. non-configurational languages has been of relatively little importance to dependency analyses. It is no surprising, since the dependency trees, with several elements being co-dependent on the same head, seem even better suited to describe languages in which elements depending on the verbal governor can appear in any possible order [Mel'čuk 2009: 88] with respect to the verb, without violating the projectivity [Mel'čuk 1988: 35ff, Osborne 2003: 241-2]. However, the situation changes when relations between nouns and adjectives (or adjective-like elements) are taken into account. Consider the Kalkatungu sentence (5.7), repeated here as (5.34)

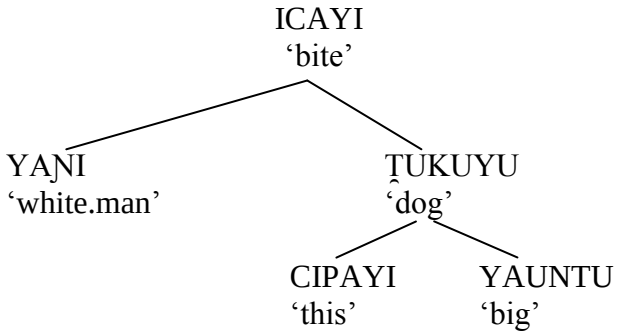
(5.34)

Cipayi	icayi	yaɲi	ʈukuyu	yauntu
This.ERG	bite	white.man	dog. ERG	big. ERG

'This big dog bit/bites the white man'

If it is assumed that Kalkatungu is an extremely free order language, and the alternative orderings of elements (as presented in (5.8)) simply represent surface discontinuities or surface non-projectivity, a deep syntactic representation of this sentence could be:

(5.35)



However, if the analysis of this example follows Blake [1983: 145], the way both Hengeveld, Mackenzie [2008: 287-298] and Rijkoff [2008: 525] do, it would mean that in dependency syntax there is no dependency (and no syntactic relation) between *ʈukuyu*, *cipayi* and *yauntu.* Rijkhoff [2008: 525], following Blake considers these, and similar cases, as "nominals in parallel or appositions". In terms of dependency syntax this could be at first glance interpreted as something parallel to Mel'čuk's [2009: 55] appositive adverbial relation or a modificative adverbial relation, postulated by him as one of the surface syntactic relation in English and illustrated by

(5.36) Modificative-adverbial:

[*As always*] ***elegant,*** ←**mod-adverb-**[*Alan*] *walked* [*away*].

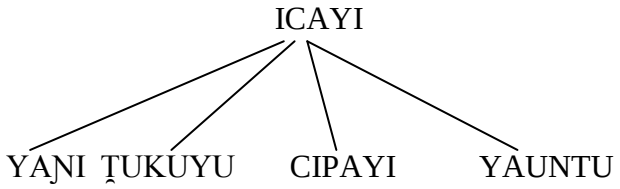
Appositive-adverbial:

[*An old*] ***man***, ←**appos-adverb-** [*Alan*] *works* [*less*].

but adapted to the deep-syntax analysis. However, this account would still have the noun as head, since both the modificative and the appositive element depend on a noun, thus the

discontinuity would not disappear. Since this proposal is not satisfactory, another needs to be made, the one that would have all nominal phrases depending on the verb.

(5.37)

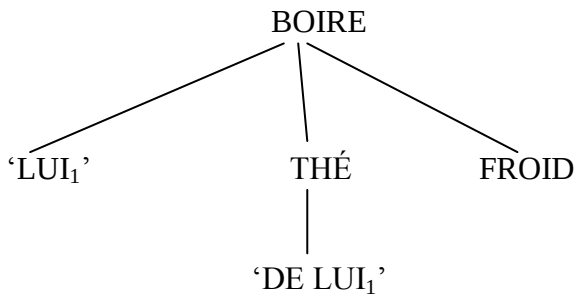


Similar analyses could be made for deep-syntactic representation containing secondary predicatives, as in (5.9), which is what Mel'čuk seems to suggest (and actually states for a similar French constructions [Mel'cuk 1988: 122; 2009: 60]:

(5.38a) ***Elle*** *semblait* ***fatiguée***
'She seemed tired'
(5.38b) ***Elle*** *est rentrée* ***heureuse***
'She returned happy'
(5.38c) *Il buvait son* ***thé froid****/sa* ***tisane froide***
'He drank his tea cold/his herbal tea cold'
[Mel'čuk 2009: 60, original emphasis and glosses]

Thus for (5.38c) (to illustrate just one instance) the dependency representation would be

(5.39)



with deep-syntactic representation of pronouns

Similar case can be made for (5.13), repeated here as (5.40)

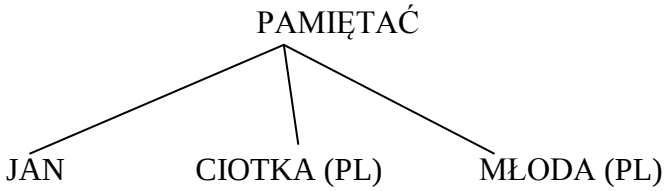
(5.40) *Jan pamięta ciotki młode*
John-NOM remembers aunt-ACC.PL young-ACC.PL
'John remembers [his] aunts [as] young [persons]'

in particular, because the predicative does not have to agree with the direct object as to case, as can be seen in (5.41) (structurally parallel to (5.9)):

(5.41) *Jan pamięta ciotki młodymi*
John-NOM remembers aunt-ACC.PL young-INSTR.PL
'John remembers [his] aunts [as] young [persons]'

with the deep-syntactic representation of both (5.40) and (5.41) being:

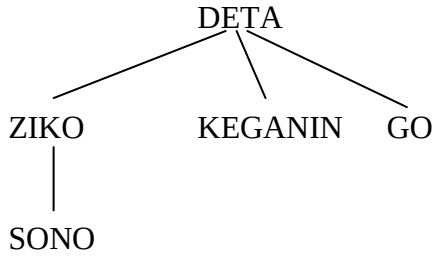
(5.42)



Rijkhoff [2008: 525] notices the similarity of (5.7) to Japanese numeral syntax. Indeed, Japanese numeral phrases do not depend syntactically (in terms of surface syntax) on the noun they quantify [Mel'čuk 2009: 59 (original example)], but depend on the verb, and the classifier marks the semantic relation between the noun and the numeral:

(5.43a) *Sono ziko +de keganin +ga* ***go+nin deta***,
this accident injured.people five emerged
LOC SUBJ CLASS
'In this accident, five people were injured.'

(5.43b)



What is in fact proposed here is that secondary predicative constructions in Polish, Russian, French and possibly other languages, and Japanese numeral phrases as well, should be considered as having not only the adjective or numeral phrase depending on the verb in the surface syntactic representation, but in the deep-syntactic representation as well.

The difference between these languages and Kalkatungu, should it indeed be a flat language, is that the secondary predicative constructions and numeral constructions are restricted lexically, and what is more important, both the languages with secondary predicatives and Japanese possess ordinary modifying constructions, while in Kalkatungu, under the analysis proposed along the lines presented in (5.37), such constructions would be absent. From this it would follow that in order to account for the alleged Kalkatungu type of non-configurationality the Encoding Grammar needs to be able to say that while it agrees with Mel'čuk's claim that deep-syntactic relations need to be relatively cross-linguistically universal, there is some room for typological variation.

5.4. Surface Syntax

As could be seen throughout the present book, the Encoding Grammar tries to follow the MTM model as closely as possible, and it deviates from it only in cases where MTM relies strongly on being a bi-directional model. The second instance of such deviation is that the Encoding Grammar is rigorously trying to present the information structure as an integral part of sentence representation at each level. This also is related to the exclusively encoding character of the approach. The notion of surface syntax as the level intermediate between the deep syntax and morphology is one of the solutions taken from the MTM, however, as to the nature of the surface syntax the Encoding Grammar and the MTM have to part ways. As it has

been said in Chapter II (*2.3. The Encoding Grammar And The Meaning⇔Text Model* and note 18) the Encoding Grammar needs a hybrid model of syntax, where at least for some languages the surface-syntax representation has to be posited as having a form of a phrase-structure tree, instead of a dependency tree.

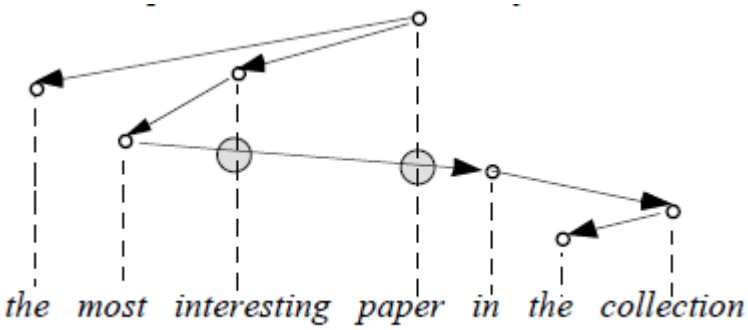
Mel'čuk [1988; 2009] argues for superiority of dependency syntax over the phrase-structure syntax and his arguments are valid. Dependency syntax matches better the semantic representation of a sentence, and it naturally models subordination between elements.

He also gives some arguments in favor of dependency syntax being a better tool for defining projectivity and describing both projective and non-projective constructions [Mel'čuk 2009: 85-87] , but he also [Mel'čuk 2009: 88] argues that

> the D-approach is much less rigid (than the C-approach) and has the inherent ability to accommodate easily what is known as 'non-configurationality' and long-range dependencies. The perturbations introduced into the word order of a sentence by its Communicative Structure—Frontings, Extractions, Postponings, etc. plus all sorts of 'displacements' in such languages as German or Russian—can wreak havoc on a C-structure, since even the closest-knit phrases can be torn apart and permuted. On the other hand, D-structure, without linearity and contiguity, is totally insensitive to such permutations: they happen in the linearized DMorphS of the sentence and do not at all affect the SSyntS. The reason is obvious—a strict and complete separation of hierarchical (=syntactic) and linear links in the D-approach. As a result, the D-approach does not know problems in representing discontinuities, which, in the literal sense of the word, simply do not exist in a D-structure. [Mel'čuk 2009: 88]

Thus one is supposed to understand that non-projective structures produced by superlative constructions, i.e.

(5.44)



[Mel'cuk 2009: 86]

is syntactically discontinuous (i.e. the discontinuity is due to lack of projectivity), while for sentences with the order changed for communicative reasons, this order is not manifest on the syntactic level, but on the morphological one. Then two questions should be ask: one, of minor character: why the Serbo-Croatian example [Mel'čuk 2009: 87, original glosses]

(5.54) *Verski mu je učitelj odvratio,*
lit. 'Of-faith to-him has [actually, 'is'] the-teacher answered'
'The teacher of faith has answered to him'.

illustrates syntactic non-projectivity and not just different morphological representation, and one more general: At what stage are the reflexive pronouns inserted? They are supposed to appear in their surface form (i.e. as reflexives and not as ordinary nominal marked for co-referentiality) on the surface-syntactic level (or, in Mel'čuk terms to have nominal mapped

into them). Nevertheless, since they are affected by linear order, they should appear (or acquire their surface form) on the deep-morphological level only. To illustrate this the example (2.20b) from Chapter II repeated here as (5.55) will be used:

(5.55) *Jego*$_1$ *żona*$_2$ *powiedziała Janowi*$_3$*, że* Ø$_2$ *chce rozwodu.*
His$_1$ wife$_2$ told John$_3$ that [she]$_2$ wanted a divorce

If the linear order is modified, and *Janowi* is displaced to the initial position, *jego* becomes co-referential with *Jan:*

(5.56) *Janowi*$_1$ *jego*$_1$ *żona*$_2$ *powiedziała, że* Ø$_2$ *chce rozwodu.*
His$_1$ wife$_2$ told John$_1$ that [she]$_2$ wanted a divorce

While this conundrum can be most likely resolved within the MTM model in which all levels of sentence representation are 'available' at all times, this solution cannot be used within the Encoding Grammar. Similarly, it can be easily shown that non-projectivity of (5.44) is of different nature than the non-projectivity of sentences in permuted order, of the kind presented in (5.12) and possibly also present in (5.54): while (5.44) is non-projective already within the deep-syntax [Linde-Usiekniewicz 2010], in (5.12) the non-projectivity disappears once the dependency relations are established independently of the actual order.

Mel'čuk's argumentation focus mostly on superiority of dependency syntax over the phrase-structure or constituency syntax, and less on a straightforward comparison of the two. Comparisons are mostly invoked as evidence for the theoretical and typological advantages of dependency over constituency. By contrast, in papers on dependency studies by linguists associated more closely with generative linguistics, as is the case of Osborne [2003; 2005]; Gross, Osborne [2009]; Osborne, Putnam, Gross [2011], explicit mention can be found of valid differences between the two approaches not addressed within the MTM argumentation.

Thus in [Osborne, Putnam, Gross 2011] several important observations can be found: One is that the dependency relation is "a strict mother-daughter relation", whereas constituency is a part-whole (=meronomic relation)" [Osborne, Putnam, Gross 2011: 320]. In consequence the dependency tree indicates clearly which element is the head, while constituency tree does not [Osborne, Putnam, Gross 2011: 321]. Both Mel'čuk [2009] and Osborne, Putnam, Gross [2011] see this feature as an advantage of the dependency analysis. The authors [2011: 325] also argue, with reason, that dependency trees present configurations as endocentric, while constituency analysis does not.

A comparison between the way the dependency syntax and the constituency syntax is presented both within the MTM model and outside it raises yet another question: i.e. should the dependency trees reflect the actual order of the sentence elements. Within MTM, in the surface syntax both kinds of trees appear, while in all works by Osborne (quoted above) there is a tendency to present dependency trees as mirroring the actual sequence of elements. Additionally, in Osborne (2003: 241) a distinction is made between pre-dependants and post-dependants.

The point whether dependency trees should reflect actual order is not moot. If they do, dependency syntax is no longer "insensitive to permutations", as Mel'čuk [2009: 88] puts it. If they do not, then their usefulness for typology may be actually lesser than Mel'čuk claims. He rightly illustrates the advantages of dependency descriptions as being capable of identifying consistently centrifugal and centripetal languages, though even for them, it fails to account for the relative order of dependants: In his original examples

(5.57a) Japanese
Itiban takai siraga+de+no sensei+wa kono omosirokunai hon+o kai+ta
lit. 'very tall gray-haired professor this boring book wrote'

(5.57b) Welsh:
Ysgrifennodd athro tal iawn a gwallt llwyd ganddo y llyfr undonnog hwm
lit. 'wrote professor tall very and hair gray to-him the book boring this'

not only all dependants precede the verbal head in the Japanese case, and follow the verbal head in Welsh, but also in both language they do it in the same order, i.e. with the subject preceding the object. While it can be argued that in both cases the ordering of dependants follows the ranking (i.e. numbering) of deep-syntactic actants, in case of languages that are neither consistently centripetal nor consistently centrifugal and would have one dependant preceding the verb and all the other follow it, as is the case of SVO languages and OVS languages, the ranking of deep-syntactic actants may naturally account for the SVO order, but by the same token it cannot account for OVS order.

Yet another question remains, i.e. if argument ranking actually reflects some hierarchy between the co-dependants or a hierarchy of their closeness to the head [cf. Mel'čuk 2009: 73-74]. For the languages that are centripetal in their arrangement of verbal dependants, e.g. Japanese, it seems that the actual ordering reflects the hierarchy of closeness, while the centripetal languages, e.g. Welsh, seems to reflect the hierarchy of actants and not that of their closeness to the head.

What the observations made above suggest is that dependency syntax has some minor weaker points with respect to constituency syntax. Even though it provides answers to possible critiques, these answers may not be optimal, and even less so, within the Encoding Grammar, especially, if it is remembered that since the Encoding Grammar does not separate the theme-rheme division either from semantics, or from syntax, the Surface Syntactic representation needs to be affected by the permutations of elements that encode the theme-rheme divisions. Since the theme-rheme division is binary and symmetric, there needs to be some syntactic representation that would share this binary and symmetric character and here is where the constituent structure comes to play. It provides natural cutting points for thematic and rhematic parts of utterance, since the most natural and straightforward division is between the subject and the verb phrase. In addition, while dependency syntax adequately accounts for fronting, i.e putting the thematic (or rhematic) element first, it does not account for what happens with the remainder of the sentence. The case in hand is provided by Polish. Although it is considered a relatively free-order language, fronting of objects (or even of non-objects which are not adjuncts (see above), results in re-arrangements of remaining sentence elements. Thus if we take the often quoth example

(5.58) *Kolumb odkrył Amerykę*
'Columbus discovered America'

with the object fronted, we see that not only the object is fronted, but the subject is put after the verb, as in (5.59),

(5.59) *Amerykę odkrył Kolumb*

while allowing the subject to remain pre-verbally would result in a highly marked configuration with two-level division into themes and rhemes, and a prominent rheme signaled by appropriate intonation:

(5.60) *Amerykę ↑ Kolumb ↓odkrył*

by which the speaker would correct their audience's assumption (most likely previously expressed) that America was discovered by somebody else. The same would happen with

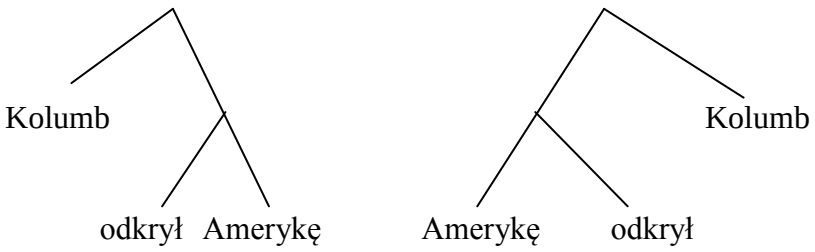
(5.61a) *Jan pojechał do Krakowa*
'Jan went to Krakow'
(5.61b) *Do Krakowa pojechał Jan*
(5.61c) *Do Krakowa ↑Jan↓ pojechał*

It should be noted that this phenomenon cannot be accounted for by claiming that the verb has to occupy the second position, as in German, because if the sentence is more complex, the subject-verb sequence does not get re-arranged:

(5.62) *Do Krakowa Jan pojechał na koniu*
'To Krakow Jan went on horseback'

This phenomenon cannot be adequately explained in terms of dependency syntax, while it is naturally accounted for if constituency is invoked. The kind of fronting observed in (5.59) and in (5.61b) is less marked, because it respects the constituents of (5.58) and (5.61a) respectively and simply implies re-ordering of two major (first-order constituents), so the two resulting constituency trees are but mirror images of each other and some original constituents conserve their adjacency

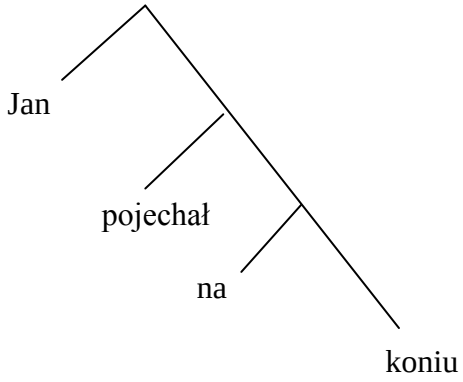
(5.63)



By contrast, (5.59) and (5.61c) are not mirror-image representation of the same constituent configuration and involve permutations that affect the constituency, since original constituents would no longer be adjacent

An interesting phenomenon can be observed in the case of (5.62), which involve neither prominent theme-rheme divisions nor prominent themes or rhemes, and divides simply into a theme 'to Krakow' and the rheme 'John went on horseback'. Although one element of the verb phrase has been taken out of it and fronted, the remainder conserves the adjacency of its constituents:

(5.64)

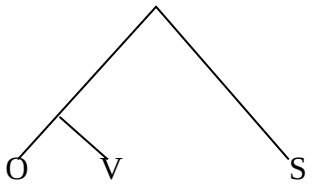


Thus, while the Encoding Grammar agrees with Mel'čuk as to the relative universality of dependency relations and their cross-linguistic applicability, it has to take a different stand when the distinction between deep and surface syntax is concerned. What is proposed here is that while deep-syntactic representation should be modeled along dependency lines, the surface-syntactic relations can be modeled as dependency relations or as constituency relations, depending on the language. Thus for the languages for which surface-syntactic relations are best described in terms of dependencies only, it would be appropriate also to have their surface-syntactic structure described in terms of possible dependencies, and, by the same token, all the surface- syntactic representations modeled within dependency syntax. By contrast, those languages which need to rely on constituency relations (even to account for different degree of markedness of various linearizations obtained when encoding the theme-rheme division), would have both their surface-syntactic structure and surface-syntactic representations modeled in terms of constituency relations. Thus languages would, not surprisingly, show stronger resemblance and more universal character at the deep-syntactic level and a lesser one on the surface-syntactic level, which is a very common-sense claim. In addition, having constituency-based surface syntax does not necessarily imply a rigid word order. Just the contrary, the constituency or actually the degree to which constituency can be violated by the word order would account for different degrees of markedness of different linearizations. This is of particular importance when the encoding the division into themes and rhemes (which is partly achieved through syntax) is taken into account. To the typological distinction between subject-driven and topic-driven languages (which prominently mark either subject or topic), and focus driven languages (i.e. those that mark rhemacity) one could add yet another type, i.e. theme-rheme boundary driven languages: those in which surface syntax reflects prominently the theme-rheme boundary, overshadowing other syntactic features. The theme-rheme boundary driven languages can be theoretically envisioned as falling into two different subtypes: one in which the boundary is imposed on dependency relations only, and the second in which the boundary is imposed on constituency relations and results in their being overcome (or violated) in linearization. While Polish, and maybe many other free order languages, may represent the second type, it is possible that Turkish and Hungarian, considered non-configurational languages, would represent the theme-rheme boundary driven type without constituency, however, all examples of free order provided by typological and other studies, e.g quoted in [Hagenveld Mackenzie 2009] have to be studied as to the degree of markedness of the word order they exhibit.

The idea of mirror-image constituency can be also of interest when talking about word order in general. Thus if it is possible to speak of constituency relations without specifying on which side the head constituent should appear with respect to the subordinate constituent, it

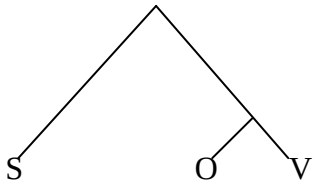
would be possible to hypothetize that not only SVO languages may exhibit this type of relations. OVS languages, or at least maybe some of them, could be described in terms of mirror-image constituency affecting both first order and second order constituents,

(5.65)

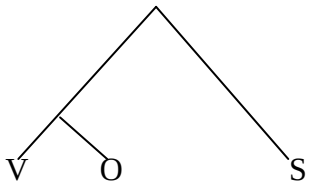


while SOV languages and VOS languages could be described by claiming a mixture of preposed and post-posed heads for different levels of constituency,

(5.66a)



(5.66b)



provided language internal evidence is found for the Subject being the external constituent.

Finally, it should be noted that within the model proposed here the rigidness of the constituent relations, considered a disadvantage in universal language description, becomes an advantage within the Encoding Grammar. First of all, it is better suited to describe syntactic structure, which is by definition a set of restrictions and not a set of possibilities. Thus, although the Encoding Grammar rejects some of the theoretical framework of the MTM model in its approach to syntax it cannot, at the end of the day, but agree with the following statement of Mel'čuk:.

> "English is a very exotic in that it uses constituency almost as its only expressive device in syntax, i.e. as the only device for encoding syntactic structure in actual sentences. In other words, constituency (marked by word order and prosody) is in English the principal observable phenomenon used to indicate on the surface, albeit indirectly, the underlying syntactic relations. Let mi emphasize: constituency is a MANIFESTATION of syntactic structure, not syntactic structure itself. But thinking in or even simply working mostly with English lures the researcher into mistaking this idiosyncratic surface trait of a particular language (i.e. relying mostly on word order and prosody to mark syntactic relations) for a universal mechanism of syntactic representation. It is then, to be expected that in languages that do not use constituency as the main surface expressive device, the constituency, or phrase-structure formalism will perform poorly as a means to represent syntactic structure
>
> Notice that dependencies, on the contrary, are not surface observable phenomenon at all and therefore they are not language specific. Being abstract relation, they possess enough generality to accommodate any type of syntactic organization" [Mel'čuk 1988: 4-5]