Linear orders as coding means

Zygmunt Frajzyngier, Dept. of Linguistic, University of Colorado

Abstract
In traditional and contemporary linguistic theories, word order has been considered a major characteristic of any language. While early studies of inflectional languages considered functions of word order mainly with respect to pragmatics and stylistics, most contemporary studies look at word order as an object itself and, since Greenberg 1963, as a potential predictor of other components of language structure.

The present study looks at word order not as an object in itself, but rather as a formal domain consisting of several coding means in a variety of functional domains. The questions asked are: what kinds of linear means are employed for marking; what functions they encode; and what are the consequences of the fact that a given linear means codes a given function. It has been demonstrated by the previous studies that the word order of major elements is not a predictor of the other elements of language structure. The following linear means were found to exist in various languages, and sometimes in the same language:

(1) Default position of a lexical category, which is not a means of coding any particular function but simply represents the default order of constituents (e.g. clause-final position of verbs in Japanese and Korean, clause-initial position of verbs in some Chadic languages).
(2) Default linear order for a given functional domain: A linear order that, without any other morphological, or prosodic means, codes one and only one function.
(3) Marking by position with respect to a constituent of different category (e.g. post-verbal position in English).
(4) Marking by alternation with respect to the default linear order and to the coding by position.
(5) Marking by relative order of constituents belonging to the same category, whereby the first constituent has a different function from the second constituent, regardless of any constituents that separate them.
(6) Marking by precedence of one constituent with respect to another, when many constituents follow each other.

The study analyzes different requirements and different consequences of these coding means and how they point to cause-effect relationships for a number of hitherto unexplained similarities and differences among languages. A fine-grained analysis of linear order means allows one to explain several differences across languages that were unexplained so far.
1 Aim and scope of the study
In most contemporary works, linear order is studied as an object in itself, with the aim of finding what changes it may undergo and what correlations it has with other aspects of grammar (Siewierska and Bakker 1996, 2009, Dryer 1991, 1992, 1995, 2002). Frajzyngier and Shay 2003 proposed that linear order is one of the formal coding means in language. Other coding means are lexical categories and derivational morphology; phonological means; inflectional morphology; serial verb constructions; and adpositions. The aim of the present study is to demonstrate that linear order, instead of being a single coding means, is a broad formal domain that consists of a number of formal means of coding. The study demonstrates that taking into account various fine-grained linear means allows us to understand connections between the functions coded by linear order and the way other functions are coded in the same language.

The study postulates a fundamental distinction between a linear position for a specific category, which carries no function and is here called ‘default position’, and linear orders that serve as coding means. The study provides evidence for the existence of the following coding means involving linear order:

(1) default linear order for a given functional domain
(2) ‘position’;
(3) ‘alternation to the position and to the default linear order’;
(4) ‘relative order’ and
(5) ‘precedence’.

The first part of the study provides:

(1) the evidence for the existence of these linear coding means;
(2) necessary conditions for their existence; and
(3) consequences of the presence of a given coding means for the form of the utterance.

The rest of the study is organized as follows: Section (2) is a review of the ways linear order is treated in some contemporary theories, both functional and formal, implicitly or explicitly; section (3) provides a brief review of the theoretical assumptions of the present study; section (4) provides evidence for the existence of the default position for single category, a position that is not a coding means; section (5) provides evidence for the default linear order with respect to a given functional domain and describes the implications of the default linear order for the rest of grammatical system; section (6) provides evidence for the existence of position as a coding means; section (7) provides evidence for changes in linear order and extraposition as coding means; section (8) provides evidence for the existence of relative order as a coding means; section (9) provides evidence for precedence as a coding means; and section (10) discusses some of implications and open questions. The study demonstrates that:

(a) a language may have several linear coding means available, and
(b) a single clause may deploy several linear coding means.
A systematic cross-linguistic study of linear order as a coding means has yet to be conducted.

2 Word order in linguistic theory
Siewierska 2006 (1994) is a useful summary of recent studies of word order. In the present section, I treat briefly just a few major approaches and concepts introduced in the study of word order so far, with the aim of contrasting these with the proposed approach.

Greenberg 1963, with whom the contemporary studies of word originated, aimed at finding correlations between word orders of major syntactic categories S, V, and O, and a variety of other grammatical phenomena. Greenberg 1963 did not consider the function of word order, and that made some of his findings so much more intriguing. Over time, studies of a much larger number of languages have confirmed some of his findings, while some of his findings have been found to be incorrect (Dryer 2005a, b, c). Some of Greenberg’s findings received a grammaticalization explanation, e.g., correlation of the position of the verb with whether the language is prepositional or postpositional (Delancey 1994).

Greenberg’s approach of ignoring the function of linear order has been continued by a large number of contemporary scholars. This, however, has not always been the case.

Early language-specific or typological approaches in the 20th century, e.g. Marouzeau 1922-1953, Milewski 1950, 1962, looked at word order as a coding means. For Milewski, linear order had functions similar to those of case marking, an approach that was taken up half a century later by lexical-functional grammar.

Greenberg introduced the notion of a dominant word order, which was crucially important for the type of investigation he conducted and for subsequent studies (see also Tomlin 1986). The word order that was considered to be dominant was the one that occurred most frequently and that was pragmatically unmarked. Mithun 1987 has demonstrated that there are languages in which one cannot talk about the dominant word order.

Dryer’s 2005 studies of word order, like many studies by other linguists, treat word order in the Greenbergian tradition, as an object of the study in itself rather than as a coding means. In this treatment, word order is determined by other factors: ‘Flexible order languages are sometimes described as having “free” word order, though this is misleading, since there are often pragmatic factors governing the choice of word order’ (Dryer 2005a). Dryer, like many others, divides languages into those with rigid word order and those with flexible word order, and for languages with flexible word orders he distinguishes a one or even two dominant word orders. Dryer also recognizes the existence of languages with flexible word order in which none of the orders is dominant, e.g. Dutch and German.

Siewierska 1998 looks at variations in word orders in various languages of Europe. Most studies in Siewierska 1998 (ed.) look at word order as an object in itself, along such parameters as fixed versus flexible.

Despite the fact that the universality of the categories S and O has been challenged numerous times (Mithun and Chafe 1999, Frajzyngier and Shay 2003, Frajzyngier and Munkaila 2004), most typological studies of linear order still ignore these facts and talk about the order of S and O.
Functions of word order are usually discussed only in connection with discourse and pragmatics, including topicalization and focusing, as exemplified in earlier studies in Li 1975 and more recently in Kis 1998.

Nichols and Bickel 2005 do not include word order among the coding means in their typology of locus of marking in the clause. ‘In any kind of phrase, overt morphosyntactic marking reflecting the syntactic relations within the phrase may be located on the head of the phrase, on a non-head (i.e. on a dependent), on both, or on neither.’ The evidence that Nichols and Bickel do not consider word order to be an overt coding means is provided by the following statement: ‘The word order makes the syntax clear; but so it does in most of the above examples. Word order is not at issue in this chapter, which deals only with overt markers . . .’.

Lexical-functional grammar treats word order as functionally equivalent to case marking (Bresnan 2001). Frajzyngier and Shay 2003, who treat word order as one means of coding language functions, demonstrate that there is actually no equivalency between linear order and case marking, as case marking in some languages codes a larger number of distinctions than it is possible to code through linear order. Like other studies, they do not differentiate among the various linear coding means.

Generative grammar adopted (implicitly only) linear order, combined with hierarchical configuration, as the main means of the representation of grammatical relations and constituency within the sentence. The parameter of headedness assumed that languages have either the order head-dependent or dependent-head, and that this parameter is responsible for the ordering of elements within a sentence (Chomsky 1995).

A number of contemporary approaches to linear order reflect a tacit assumption that syntax involves putting words together to express meaning, and that therefore linear order is the crucial information for this process, as it presumably tells people in what order to put words. In the minimalist approach, the first operation, ‘Merge’, reflects this assumption.

Viewing linear order as separate from its functions, and taking the accepted notion of dominant order as one of the fundamental properties of language, does not provide a good predictor of language form. Consider two languages, the Koyra Chiini dialect of Songhay (Nilo-Saharan)1 and Polish, both of which are said to have SVO as the basic word order (Dryer 2005, Siewierska 1998):

(1)  
\[
\text{Yer o kar gi}
\]
\[
\text{1PL.S IMPF hit 3PL.O}
\]
‘We hit them.’ (Heath 1998: 128) (s-Subject, o-Object)

If the object is moved before the verb, it follows an auxiliary if there is one, or it must be preceded by the particle na, glossed by Heath as TR:

(2)  
\[
\text{ay na ni kar}
\]
\[
\text{1SG.S TR 2SG.O hit}
\]
‘I hit you.’

---

1 The classification of Songhay remains somewhat controversial cf. Nicolaï 1990.
‘The woman killed the man.’ (Heath 1998: 212)

In Polish, the object can also occur after the verb:

‘Tomorrow, one among your people will put the five-ton vehicle under the unexploded bomb.’ (Sources)

The obvious question is why, in Koyra Chiini, the object in the position before the verb is marked by a preposition and the object in the position after the verb is not. The second question is why the object before the verb in Polish is not marked by a preposition. The most plausible answer to the first question is that the position after the verb in Koyra Chiini is the coding means for the category object. When the noun phrase occurs in another position, it must be marked for the grammatical role by some other means. The answer to the second question is that in Polish, the position after the verb is not a coding means for the category object, and that moving the noun phrase from that position does not require any additional means for the marking of the object.

The conclusion from these facts is that the functions of the SVO orders in Koyra Chiini and Polish are different and that those differences, rather than the formal presence of the SVO structure, are better predictors of formal changes (or lack thereof) than are changes in the linear order.
3 Theoretical assumptions of the present study
Frajzyngier and Shay 2003 demonstrate that a language contains a number of formal coding means that include:

- Lexical categories and subcategories
- Phonological means including prosody, vowel reduction, vowel lengthening
- Derivational and inflectional morphology
- Serial verb constructions
- Adpositions
- Linear order

The function of the formal means, which interact with each other, is to create meaning and to ensure functional transparency. The principle of functional transparency states that the role of every element in the utterance must be transparent to the listener. Transparency relates to the function of the word/morpheme in the utterance rather than to the listener’s need to understand relations in reality or in the cognitive system of the speaker. Functions coded in the language are part of the functional domains grammaticalized in the language, defined as follows:

(a) A domain D is a class of mutually exclusive types of expressions constructed by a set of specific means of coding M(D).
(b) The types of expressions within the domain D have a specific pragmatic or semantic function in common.
(c) A domain D is a class of types of expressions with a certain set of meanings D* such that D* is disjoint with E* for any domain E ≠ D of the same language. (Frajzyngier and Mycielski 1998).

Examples of functional domains are: grammatical relations; semantic relations; locative predication; epistemic modality, with its subdomains of expressing the intended truth, questions about the truth, and denial of truth; deontic modality; indirect object predication; tense; and aspect. Different languages have grammaticalized different functional domains and subdomains. Given common human cognitive abilities and interests, some functional domains have been grammaticalized more often than others.

4 Linear order as a formal domain
In the approach proposed in this study, linear order is not a single coding means but rather a formal domain comprising the following:

- Default position of a category (this is not a coding means);
- Default linear order for a given functional domain;
- Coding by position with respect to a constituent of a different category;
- Coding by alternation to the default linear orders and to the coding by position;
- Coding by relative order of:
  (a) different categories? (whether this is a coding means is an open question)
  (b) identical categories
- Coding by linear precedence.
A language may have several of these linear means, and even a single clause may display two or more of these formal characteristics. In what follows I provide the evidence for the existence of these forms, an illustration of the functions they may have, and the implications of their existence for the grammatical system.

5 Default position for a lexical or grammatical category
When a lexical category can occupy only one position in a clause, such a position is considered to be the default position. Because it is predictable, this position is not a coding means for grammatical relations, though it may be a coding means for the domain of modality.

In many languages, the verb is assigned to a specific position. In some languages, the default position is occupied not by a lexical category but rather by a functional category, e.g. the predicate, which can be either verbal or nominal. The present study uses the default position of the verb as illustration.

Languages with default position of the verb include:

- Verb-final languages: Japanese, Korean, Amharic, Tigrinya, all Cushitic and Omotic languages; and
- Verb-initial languages: Many Celtic languages, some Central Chadic languages, and Kabyle (Berber), if one considers subject pronouns to be prefixes rather than separate lexical items.

The position before the verb in verb-final languages, and the position after the verb in verb-initial languages, could potentially be a coding means for grammatical relations. However, this is not the case in the languages examined for this study. In these languages, the elements preceding the clause-final verb and the elements following the clause-initial verb may occur in different orders. Since the position after the verb in verb-initial languages and the position before the verb in verb-final languages can be occupied by any argument or by an adjunct, the position before the verb is not a coding means for grammatical relations or for semantic relations between the arguments and the verb. Here is an example from Japanese. While the verb is usually in clause-final position, the order of constituents before the verb is fairly flexible:

(6) a. Taroo ga Hanako ni sono hon o yatta
    ‘Taro gave that book to Hanako.’

In Kafa (Omotic), which is verb-final, the position before the verb is not a coding means for any grammatical relation. This position can be occupied by the subject, the object, or an adverb (transcription and glosses preliminary, from Z.F. notes):
Here is an example of the position before the verb occupied by an adverb:

(8)  kètó  tôki  hâgiyètè  
     house  together  build (IMPF)-PL  
     ‘they build the house together’

A lexical or grammatical category other than the verb or the predicate can also have a fixed position in the utterance. One such category is complementizers, which in some languages must be clause-initial or clause-final. Interrogative particles, whether for polar questions or for content questions, sometimes occur in clause-initial position or in clause-final position. Similarly, negative particles occur in clause-initial or clause-final position in many languages. In some languages negative particles occur in the position after the verb (Wandala, Frajzyngier in press, Dryer 2010 for other Central African languages). The fixed position of these grammatical markers, like the fixed position of the verb, is not a coding means. Moreover, these markers do not constitute a reference point for other linear order coding means.

6 Default word order with respect to a given functional domain

The default word order with respect to a given functional domain (henceforth called ‘default (linear) order’) is a morphologically and syntactically unmarked configuration that carries a specific function in a given domain and that is the least-marked function within that domain. For example, a linear order may carry the unmarked function in the domain of modality, in the domain of information structure, in the domain of pragmatics, or in the domain of reference (Mandarin, Li and Thompson 1989; similar claims have been made for Russian). It is theoretically possible that the default linear order carries a function within the domain of grammatical or semantic relations, but data indicating that such is indeed the case are not immediately available.

Changes to the default linear order can code other functions within the same functional domain. Whether changes in the default linear order can code functions in different functional domains remains to be investigated. The existence of the default linear order, and its function, can be discovered through a study of the forms and functions resulting from alternations to the default order. Here is the evidence for the existence of default word order, and its properties, in English.

The hypothesis to be argued here is as follows: With non-interrogative prosodic characteristics, the order NP V codes the indicative modality, i.e., a modality in which the speaker conveys some proposition as true. For the notion of intended truth in indicative sentences, see Frajzyngier 1985 and Frajzyngier 1987. Here are a few examples from the British National Corpus (http://www.natcorp.ox.ac.uk/):
123 They went to hear the Sibelius Violin Concerto superlatively played by Isaac Stern.

212 I went to a group called ‘Overeaters Anonymous’.

The evidence for the proposed hypothesis is provided by the fact that all other subdomains within epistemic modality, viz. questions about the truth; denial of the truth; and hypothetical modality, including doubt and hedging about the truth; require changes to the default configuration as well as certain other changes, e.g. changes in the prosodic pattern or the deployment of additional lexical or grammatical morphemes.

Questions about truth require a change in the subject-auxiliary order and, for most predicates, an insertion of an auxiliary (all examples from British National Corpus):

1854 Did number-crunchers come to lunch

3570 ‘Don’t you do what you can to help your friends?’ she demanded with a show of spirit, and when he didn’t answer she went on, ‘Or are all your relationships based on profit?’

1194 By the way, Hans, did you hear any more from Ostkreuz?’

319 Had they been jealous, thirty years ago in Oxford, of the young Phoebe’s special status?

Contradicting somebody’s assumption about the truth, and the coding of the hypothetical modality, both require use of an auxiliary and inversion of the subject-auxiliary linear order:

681 Neither do I believe that a painter could have painted it — had he seen what I saw.

2257 Had he spotted it he would probably have opted for 18...

Similarly, the denial of truth requires the insertion of auxiliary. Although the relative linear order does not change, the structure itself is changed through the insertion of the auxiliary:

137 I didn’t swamp the family with knitwear, I just concentrated on improving the quality and finish of the garments so that instead of being embarrassed by wearing ‘Mum’s efforts’ they were proud to wear them.

2205 The Kha-Khan did not appear to be offended.
Change from epistemic to deontic modality also requires changes in the formal means, e.g. insertion of an auxiliary:

1834 In the end the man became so nervous that I **had** to hold his arm and literally steer him through the crowd to the right spot.

While these facts are commonly known, the reasons for their existence have not been explained in terms of functional motivation. The obligatory marking of modalities other than the indicative indicates that the NP V order in English codes the indicative modality. This explanation is not trivial, as the same order in other languages does not necessarily carry the meaning of the speaker wanting to convey the proposition as true. In Polish, which in the present analysis does not have a default linear order, asking questions about the truth does not involve changes in the linear order of subject and verb (all examples from Korpus Języka Polskiego Wydawnictwa Naukowego PWN (http://korpus.pwn.pl), with Z.F.’s glosses and translation):

(9) **Grecy i Trojańczycy kupowali je od** Greeks CONJ Troys buy:FREQ:3PL:M:PAST 3PL:ACC from
    **Fenicjan pracęjących w kopalniach w Bitymi i** Phoenicians:GEN working:PL:GEN in mines:LOC in Bythmia and
    **Tracji** Thrace

‘Greeks and Trojans were buying them from Phoenicians working in the mines in Bythmia and Thrace.’

(10) **Kiedy on kupował, oni sprzedawali.** when 3SG:M buy:3SG:M:PAST 3PL:M sell:3PL:M:PAST

‘When he was buying they were selling.’

Polar questions can be coded by the interrogative particle **czy** in clause-initial position, without any additional changes to the structure of the clause:

(11) **Czy Lulek sam odczuwał cierpienia** whether Lulek alone feel:3SG:M:PAST suffering:PL
    **uciętżonych?** oppressed:PL:GEN

‘Did Lulek himself feel the suffering of the oppressed?’

The notion of default linear order differs from the notion of dominant word order, as found in traditional typological studies, in one fundamental way. Recall that frequency of occurrence is the main criterion in the choice of the dominant word order. The default order is a coding means within some specific functional domain, while the notion of the dominant order does not include the function. Dryer 2005 accepts that a language may have several dominant orders. A given functional domain can have only one default order if it uses linear order as a coding means. The notion of default linear order allows one to investigate phenomena separate from those associated with the notion of dominant word
order. For example, changes to the default word order must trigger compensatory changes; otherwise, the function coded by the default order will be compromised. Studies concerning the dominant word order do not postulate or envision compensatory changes, since no function is associated with the dominant order.

The existence of a default word order provides an opportunity to generate a variety of linear coding means, since every departure from the default word order creates new formal means for coding a function.

7 The coding by position
The term ‘position’ in the proposed approach refers only to the position preceding or following a certain reference point. The position itself is a coding means. Its closest equivalent in contemporary theory is ‘adjacency’. The importance of position has long been recognized in linguistic theory and in individual language descriptions. In generative grammar, under the term ‘adjacency’, it refers to the place in the underlying representation of one element next to another. Heath 2005 makes use of the notion of adjacency to describe the conditions in which certain morphophonological processes can take place in Tamashke.

Coding by position crucially requires a reference point. The reference point must meet the following criteria: It must be phonologically or morphologically distinct from all other constituents in the utterance; there must be only one reference point per utterance; and nothing can occur in the position between the reference point and the element marked by the position. Since the verb in many languages has a variety of affixes attached to it, it is readily identifiable and often serves as reference point. However, even if a description of linear order involves reference to the verb, this does not necessarily mean that the position with respect to the verb is a coding means. The position with respect to the verb is not a coding means if it can be occupied by a large number of lexical categories coding a variety of functions. One does not know whether position is a coding means in a given language unless the writers of individual language descriptions have studied the question and explicitly say so. Otherwise, linguists must do the research on their own.

I provide evidence for the existence of coding by position by describing the coding of direct and indirect objects in English, where the position after a transitive verb is a coding means for the direct object. Nothing else can be inserted into this position. Why the verb is the reference point is not difficult to understand: The verb has inflectional marking for tense, it has the third-person singular agreement marker, and it is preceded by a subject pronoun or auxiliary verb(s).

Compare an ungrammatical utterance with the prepositional phrase following the verb:

*I did get from him a postcard

The reason the clause above is ungrammatical is because the grammatical role of the noun phrase ‘a postcard’ is not transparent. And the reason it is not transparent is because the coding means, the position after the verb, is occupied by a prepositional phrase.
Although prepositional phrases can occupy the position after an intransitive verb, they cannot occupy the position that is a coding means for the category object:

1. be `cause I’m going to Madrid
2. on the `teenth

No adverb can be inserted in the position after a transitive verb:

1. but [[@] `really I’ve got about. thr`ee w`eeks]
2. **less than th/at**

*but I’ve got really about three weeks*

If the principle of functional transparency is satisfied and there is no direct object in the clause, the adverb may follow the verb:

1. ^so . I shall have . r/oughly#
2. from the twenty-ninth of June .
3. to the eighth of July .
4. on which I can . [@] I can spend the wh`ole of th/at time*

The predicate of the equational clause in English is not the object, as the position after the copula can be occupied by an adverb:

1. *this *this is awfully k/ind#
2. *be`cause*
3. *the :whole bloody (!!th=ing#)*
4. he *he *he is* :really `God al:m\ighty#
5. he *knows \everything# - -

In contrast with the position after the verb, the position before the verb is not a coding means for grammatical relations, as it can be occupied by elements other than the subject:

1. actually *got it for y/ou#)
2. I *really `wouldn’t*

and - *finally I rather burst \out#*
1. ^I said :look h`ere#

and *very /often#`
1. you *get a :student :who :probably \does understand the p/assage#
1.2.1 ((then)) *this im*mediately brings in
1.2.1 the **funds** *of*

814 A> he [@] *sort of* looks at _one#

<104 A>*. ((because they *[o:I]*)* they *also ((:know who the :head of this))*)

1.2.1 [:@:m] I *don’t think they’ve* . they *ever in :fact embodied anything :quite as - :quite as far-reaching#*

The fact that an adverb can occur in the position between the subject and the verb is made more interesting by the fact that this is a position where adverbs cannot be inserted in French, as briefly discussed below.

**English ditransitives**

An interesting piece of evidence for coding by position is provided by English ditransitive constructions. These are constructions in which a verb is followed by a noun phrase referring to a person who receives an object and then by a noun phrase referring to the object that was moved. This object is coded by the position following the second noun phrase. The obligatory structure is V NP NP, i.e., the position after the verb must be occupied by one noun phrase and the position after that noun phrase must be occupied by another noun phrase:

1.1.0 <50 B> *you give them the :lot {*you see*} # *.*

Compare the ungrammaticality of an utterance in which the afterthought clause ‘you see’ is placed in any of the coding positions, i.e., after the verb or between the two noun phrases:

*you give you see them the lot

*you give them you see the lot

One also cannot insert an adverb in the position after the verb or between the indirect object and direct object:

*you give happily them the lot

*you give them happily the lot

The constraints on the occurrence of an afterthought phrase or an adverb after the verb or after the first post-verbal argument are characteristics of a specific language.
Since the position after the transitive verb in English is the coding means for the object function of the noun phrase, placing an adverb after the verb would in effect destroy the function of the verb as a reference point. Since the position after the indirect object is the coding means for the category object, placing anything else in that position makes the role of the ensuing noun phrase opaque, and therefore makes the utterance ungrammatical.

In a language that does not code grammatical relations by position, inserting an adverb or an afterthought clause after the verb and before the indirect object, or after the indirect object but before the direct object, is perfectly grammatical. This is the case in Polish, where the role of every element is marked by case inflection on the noun:

(12)  a. *Dajesz* rozumiesz *im* wszystko
      Dajesz  understand:2SG:PRES 3PL:DAT everything:ACC
      give:2SG:PRES

b. *Dajesz* im *rozumiesz* wszystko
     Dajesz  understand:2SG:PRES everything:ACC
     give:2SG:PRES 3PL:DAT

‘You give them everything, you see’

Koyra Chiini also appears to code the grammatical role of the object by the position following the verb, as a noun phrase in a position other than that following the verb must be marked for the object role by the preposition *na* (see examples 1-3, above). If the noun phrase follows the auxiliary, it does not have to be marked by the preposition *na*.

In Tswana (Niger-Congo), Creissels (2002), the position after the verb appears to be restricted to the argument determined by the role of the extension, e.g. agent for cusative extension, benefactive for applicative extension. If there are no extensions, the position after a transitive verb can be occupied by the object only. Nothing can be inserted between the verb and the noun phrase.

8 Alternations to the default linear coding and to the coding by position

Modifications to the default linear order, and alternations to the default position for a lexical category, are coding means used for a variety of functions. In English and German, subject-auxiliary or subject-verb inversion is one of the means of coding of interrogative functions.

2_6_0 <14 a> *he* has applied this time has he

3_1_1 <449 B> ^why has he m\ade it do then^ -

4_7_0 <633 b> ^how \old is she^#

In some languages, putting a noun phrase in clause-initial position is one of the means involved in topicalization. There are additional means to code the grammatical role of the topicalized noun, such the use of a pronoun or the use of the noun itself in the default coding position:

4_7_0 <751 c> well ^Harry Matthews’s ‘dad^#

|4_7_0 <752 c> I ^wouldn’t feel :sorry for :h\im^#
In some SVO languages, putting the subject in clause-final position is a means to code focus on the subject, as in the following example from Pero (West Chadic):

(13)  lí-kò-n  mái  bátùurè
put-Compl chief European
‘It was the white man who appointed the chief.’

Compare the nominal subject in the pragmatically neutral clause:

(14)  púcci  yòw  gwándùm  má-tù-lù  máy-cù
long ago Gwandum COND-FUT-put chief-3PL
‘Long ago, when Gwandum wanted to select their chief . . .’

Topicalization of a noun phrase in Pero involves fronting:

(15)  mánnà  mól-kò  àà,  péemù  kàt-tàwù  cig-i-m
spouse brother-2M no, 2REP PROH-touch body-DEF-NEG
‘Your brother’s wife, oh no, you are not going to touch her.’

(16)  kàdà  tìá-mù  ci-cún-kò  nin-cínù-i  tà-cáá-dànì
skin goat-DEM REL-skin SUB-3PL-REL FUT-take-
‘The skin of the goat which they have removed they will take it.’ (Pero, West Chadic, Frajzyngier 1989)

In Japanese, where the verb is usually in clause-final position, the object can occur in post-verbal position. When this is the case, the object must be marked by the accusative case marker o (Ogihara 2009):

(17)  watashi-ga  make-ta  wake  hanashi-o
I-NOM turn-PST indeed topic-ACC
‘Indeed, I changed the topic.’ (callhome train ja_1048: 421.91 424.83), Ogihara 2009)
Changes to the default linear order can be used for functions other than focus, topicalization, and information structure. For example, in Lele (East Chadic), differences in linear order code the difference between the third-person plural subject pronoun and the unspecified human subject pronoun. The third-person plural pronoun subject is marked by the pronoun ge following the verb:

(18) \textit{kiré dàgè ná ásièè è gé kâsigú}  
Kire 3PL ASSC Asue go 3PL market  
‘Kire and Asue went to the market.’ (Frajzyngier 2001)

The unspecified human subject is coded by the form \textit{ge} preceding the verb:

(19) \textit{ge yàábé-ŋ na kib-rè hamli nēy}  
HUM tell DAT-1SG HYP mouth-2F light very much  
‘I was told that you do not keep secrets . . .’

(20) \textit{dày haŋ ba na ge bé-y kùrbáló}  
3M DEM COM HYP HUM give-3M chief  
tamà-y ba na-y wèl ná undo  
wife-3M COM HYP-3M sleep COM 3F  
hà kur na wèl na \textit{ge digr-iy ná}  
till place HYP dayHYP HUM kill-3M COM  
kolo tumadu-y ná dārī-niy dē  
reason death-3M COM anger-1SG NEG  
‘The first one said that if he is given the chief’s wife to sleep with till the daybreak, then he can be killed, because death does not worry him.’

One of the important questions in linguistic theory is why some languages have coding means that other languages do not; for example, why interrogative clauses in English and German can be marked by auxiliary-subject or verb-subject inversion, while in Polish or Moroccan Arabic they cannot. We can now provide an answer to this question. The relevant hypothesis would have the form: If the order of two different categories codes a function within a given functional domain, a change to this order is available for the coding of a different function within the same domain. This hypothesis explains the use of subject-auxiliary inversion in English, and possibly in German, to code functions other than the default function, viz. the speaker’s will to convey the proposition as true (indicative modality) (Frajzyngier 1985).

In English, the inversion of the default order would result in the position after the verb being occupied by a noun. With transitive verbs, this position codes the object. Such an inversion would thus violate a fundamental coding means in English. Insertion of an auxiliary and inversion of subject-auxiliary order permits one to change the modality of the clause while preserving the coding means for the category object. As shown in section 6, subject-auxiliary inversion is one of the means deployed to code other modalities, including interrogative, and hypothetical.

In German, the position of the verb in clause-second position is fairly fixed. Given that the inversion of the verb and the subject is exploited for the coding of interrogative
modality, it is possible that the verb-second position in German is actually a coding means for indicative modality. Since the category object in German is marked by inflection, the position after the verb is not a coding means for the category object, although it can be occupied by an object. Hence, the motivation that is proposed to be responsible for the insertion of auxiliaries in English does not obtain in German.

9 Relative order as a coding means

The notion of relative order refers to the order of two elements that belong to the same category and have the same inflectional marking and the same prosodic characteristics. Coding through relative order implies that the first element has one function and the second element has another. If the order of the elements were to be inverted, their functions would also be inverted. Here is the evidence for the existence of relative order as a coding means.

Pragmatically neutral equational clauses in the present tense in Russian consist of two noun phrases without any copula. The functions of subject and predicate are marked by the relative order of the noun phrases. The first noun phrase is the subject and the second noun phrase is the predicate:

(21a)

Ведь ‘digito’—‘па́лец’, ‘stigma’—‘кра́пина’—т. е. ‘кра́пчатые па́льцы’
Ved’ digito palets stigma krapina t.e. krapchatye pal’tsy
but ‘digito’ finger ‘stigma’ spot i.e. ‘spotted fingers’
‘but digito is ‘finger’ and stigma is ‘spot’, hence ‘spotted fingers’ ([Д. В. Семенов. Сенсационный кактус // «Первое сентября», No. 14, 2004], Russian National Corpus)

One can reverse the order of the noun phrases, but that reverses the subject-predicate relationship in the clause:

(21b)

Ведь ‘па́лец’—‘digito’, ‘кра́пина’—‘stigma’—т. е. ‘кра́пчатые па́льцы’
Ved’ palets digito krapina stigma t.e. krapchatye pal’tsy
but ‘finger’ digito ‘spot’ stigma i.e. ‘spotted fingers’
‘but finger is ‘digito’ and spot is ‘stigma’, hence ‘spotted fingers’

In equational clauses in Hebrew in which two nouns have the same prosodic marking (though there is a break between them), the functions of subject and predicate are marked by relative order, the first noun being the subject and the second noun being the predicate:

(22a) Moshe ha-sofer
Moshe DEF-writer
‘Moshe is the writer’
In Hdi (Central Chadic), which is a predicate-initial language, the relative order NP NP indicates that the first NP is the predicate and the second NP is the subject:

(23) klùgà wú kághá klùgà
  plate what thing 2SG plate
  “You are a plate of what, Plate?”

(24) klùgà dáf-i ká-à
  plate food-1SG COMP-3SG
  “I am a plate of food,” he said

(25) mìnd-à ráyá mbìtsà
  man-GEN hunt Mbitsa
  ‘Mbitsa is a hunter’ (Frajzyngier with Shay 2002)

In all Cushitic and Omotic languages, the verb in a pragmatically neutral clause occurs in clause-final position. In some Cushitic and Omotic languages, the order alone of two NPs before the verb indicates that the first NP is the subject and the second NP is the object. This is the case in Kafa. If the object is indefinite it does not have the accusative case marking (own notes on Kafa, transcription and analysis preliminary):

(26) kàfá ikkèánâmùbfíó bùnóó bi fsmègbàttòcè
  Kafa one masculine child:M one feminine coffee
  bi-fsmè-gàttòcè
  3M-like-COND
  ‘if the boy likes coffee in Kafa . . .’

(27) ábicí i hàmmí ànàmì hàfíó
  how 3M:SG going male child:M
  ‘how a boy goes and asks’

màcyà hàfíó bè éc-yí bètò gátà
  girl child:F 3SG: ask-CAUS be how
  ‘how a boy asks the girl’ . . . .

Even languages that have case marking deploy relative order as a coding means when the case distinctions are neutralized. In Polish, nouns belonging to the neutral class do not distinguish between the nominative and accusative case. The relative order of two neutral or two masculine non-human nouns indicates which one is the subject and which is the object: The first noun in the order is the subject and the second noun is the object. Such a pair of nouns can be placed anywhere with respect to the verb provided that their relative
order is kept constant. Here is a constructed example to illustrate the principle. In all three sentences, the relationship between the participants is the same:

(28) a. *cielę przestraszyło dziecko*
    calf:N frightened:SG:N child:N

b. *cielę dziecko przestraszyło*
    calf:N child:N frightened:SG:N

c. *przestraszyło cielę dziecko*
    frightened:SG:N calf:N child:N

‘A calf frightened a/the child.’

The preposition *z* can be followed by a noun in the genitive case. The meaning of such a construction is that of a locative complement, ‘from X’. The noun of the locative complement can be modified by another noun, which must also be in genitive case. There is thus a sequence of two nouns in the genitive case. Which is the complement of the verb and which is the modifier of the complement is indicated solely by the relative order of the two nouns in the genitive case:

(29a) *wygnął go z raju nieświadomości*
    chase:3M:SG 3SG:M:ACC from paradise:GEN ignorance:GEN
    i wynalazł grzech
    CONJ invent:3M:SG sin:ACC

‘He chased him out from the paradise of innocence and invented sin.’

(Sources)

Reversing the order of the two nouns will reverse the semantic relationship:

(29b) *wygnął go z nieświadomości raju*
    chase:3M:SG 3SG:M:ACC from ignorance:GEN paradise:GEN
    i wynalazł grzech
    CONJ invent:3M:SG sin:ACC

‘he chased him out from the ignorance of paradise and invented sin’

In Kabyle (Berber, Afroasiatic), a noun can occur in two forms: absolute (free) and annexed. The verb in Kabyle codes the person, gender and number of the subject through a system of affixes. In pragmatically neutral clauses, the verbal piece, with subject affixes, occurs in clause-initial position. The role of the noun in the annexed state is determined only if the noun is within the prosodic group of the verb. In such case, it is the subject. The role of the noun in the absolute state within the prosodic unit of the verb has to be interpreted as the object, because the absolute state cannot be bound by a preceding pronoun. The roles of the nouns in the annexed and the absolute states outside of the prosodic group of the verb are undetermined. (All Kabyle examples, with glosses and translations, courtesy of Amina Mettouchi. The analysis incorporates the findings in Mettouchi and Frązyngier (submitted):)
As long as the state marking is retained, the order of the two noun phrases can be changed without changing the truth value of the clause. Such a change in word order is possible with heavy subjects, e.g. subjects modified by a clause:

(30b) ye-čča a-ksum we-qcic
3M. SG-eat.PFV ABS.SG-meat ANN.M-child
‘The boy ate meat.’ (Kabyle)

Here is a natural discourse example:

(31) ufa-nt t-xeddem-ed t-i-γrif-in
find.PFV-3F.PL 3F.SG-do.PFV-PROX F-ABS.PL-pancakes-F.PL
-t-mēttut-nni n babatsent F-(ANN)woman-ANAPH GEN their_father
‘They found their stepmother cooking pancakes.’

If the verb has an object suffix and there is a nominal object in the clause, the object must occur in the annexed state. Such an object is outside of the prosodic group of the verb (marked by /)The structure V-OBJ.Suffix NP[ANNEX]/NP[ANNEX] indicates that the first noun is to be interpreted as the subject and the second noun as the object.

Changing the relative order of the two nouns would change the grammatical relations of the nouns:

(32) te-wwet=it te-qcic-t / we-qjun-nni
SBJ3F.SG-hit.PFV=OBJ3M.SG F.ANN\-child-F.SG ANN.M-dog-ANAPH
‘The girl hit it, the dog’

(33) yeddem-it weqic / ughanim
SBJ3MSG.grab(PFV)-OBJ3MSG boy.ANN. reed.ANN
‘the boy grabbed it, the reed’ (Nait-Zerrad 2001: 62)

Relative order alone allows the listener to interpret what is the subject and what is the object in Modern (and Biblical) Hebrew and in those Central Chadic languages which are verb-initial. In Modern Hebrew, which is usually considered to be an SVO language, the two noun phrases can be separated not only by the verb but also by adverbs and prepositional phrases:

(34) asinu po lup
we_did here loop
10 Linear precedence
In a language with extensive inflectional marking of case on the noun and of subject and possibly object on the verb, linear order is available to code a large number of functions unrelated to the grammatical or semantic relations between the predicate and noun phrases or between the noun phrases. The constituents of the clause can be put in a large number of configurations and still preserve the same relationships between the verb and the noun phrases or between the noun phrases. In Polish, the object (in brackets) can precede the verb and the subject (also in brackets) can follow the verb:

(35) tylko że [krzyż walecznych] zostawił [twój
except COMP cross valor:PL:GEN left:3SG:M 2MSG:POSS
wuj Staś] dopiero w dwudziestym czwartym.
uncle Stas only in twenty:INSTR four:INSTR
‘Except that, with respect to the Cross of Valor, your uncle Stas left it only in nineteen twenty-four.’ (Sources)

All possible configuration of the noun phrases enclosed in brackets are allowed. The prepositional phrase dopiero w dwudziestym czwartym ‘only in nineteen twenty-four’ can also occur in any order with respect to the verb and the noun phrases.

In a language with obligatory case marking on the noun, the number of linear orders allowed in a clause is N! (N factorial), where N consists of all case-marked elements and a predicate. In Polish, all possible arrangements of S, V, and O are allowed, provided that case marking and marking on the verb allow one to identify the roles of noun phrases. A clause consisting of three noun phrases and a verb can thus have up to 24 linear word orders (4! = 1 x 2 x 3 x 4), while a clause with four nouns and a verb may have up to 5! (= 120) linear orders:

(36) Dal miejsce wola malpielew,
because 3SG:M amuse:3F:PAST
‘The lion gave the place of the ox to the monkey, because it [the monkey] amused him.’ (Niemcewicz, Bajki)

Any linear order in the first line of the sentence above will convey the same relationship in reality with respect to the protagonists, viz. the ox, the monkey, and the lion. The phrase miejsce wola ‘the place of the ox’ can have the order of elements reversed, but the two elements cannot be separated by another element.

A similar situation obtains in many Australian languages and in other Slavic languages that have not lost case marking. The question, then, is what are the functions of these 24 or 120 linear arrangements, how to discover them, and how to describe them.
Studies that look at the dominant or the basic word order all claim that Polish is an SVO language. And indeed, it is entirely possible that statistically SVO is the most frequent order in Polish. But one can look at the SVO order in Polish (and Russian) as just one of the variants of many possible orders and try to discover the function of this order just as one would discover the function of any other order.

One way to describe the functions of multiple linear orders for the same constituents is by studying the precedence of one constituent over the other. These individual precedences are coding means for a variety of functions. Most of these functions remain to be discovered. The usual, and sometimes poorly defined, suspects, viz. topicalization and focus, old and new information, theme and rhyme, do not provide a sufficiently rich system of distinctions to account for the dozens of potential word orders available for the same constituents. The precedence of one element over another, and of the subsequent element over the one following it, offers a rich system of coding means, and therefore the possibility of expressing a multitude of functions. The following diagram illustrates a sequence of coding through precedence for a five-constituent utterance, where the constituents are noun phrases and the verb is in clause-final position. A similar diagram can be constructed for other lexical or phrasal categories and for verbs in different positions. In the diagram below, NP1 has precedence over NP2, NP2 has precedence over NP3, and NP4 has precedence over NP5:

```
NP1 NP2 NP3 NP4 NP5 V
  |   |   |   |   |   |
  1  2  3  4
```

Precedence differs from relative order in that a change of places in the coding by precedence does not result in a reversal of the function of elements involved. Precedence differs from default configuration in that the order of all elements can be changed without changing the functions encoded.

It is likely that the choice of some precedences may be dictated by rhythmic preferences for the utterance, similar to the ones that exist in poetry and songs. Some precedences may reflect the order in which constituents have been mentioned in the previous discourse, and some may reflect the fact that a constituent has not been mentioned previously. It is also entirely possible that the precedence of one element over another is a coding means for a specific function rather than an outcome of other constraints on the grammar. Unfortunately, not much is known about the functions of precedence of elements, apart from clause-initial precedence. As an example, here is an analysis of two clauses in which precedence plays an important role. As it turns out, the roles of the precedence in the two examples are quite distinct.

In example 36, above, the verb *dał* (‘gave:3M:SG) is new information, unexpected. The noun *miejsce* ‘place’ makes its first appearance in the poem here, but it refers to the position of the ox at the royal court. Ox’s administration was mentioned earlier. The noun *małpa* ‘monkey’ is mentioned for the first time. The noun *lew* ‘lion’ is also mentioned for the first time, but it refers to the ruler mentioned in the preceding lines. It appears, therefore, that the linear order in this particular line starts with a new predicate whose use contrasts with the predicate of the preceding line, and the order of noun phrases is from
better known, or easily recoverable, to less known. Hence, each noun phrase is marked by precedence with respect to the ensuing noun. The typical parameter of new vs. old information is not sufficient to account for the functions of precedence.

In the following example, all arguments and adjuncts precede the verb. In this clause, all noun phrases and the adjunct phrase appear to be in contrastive focus with respect to some unstated elements in reality. The precedence of each noun phrase appears to correlate with its importance in speaker’s discourse. Hence, precedence here is a coding means for a function other than new vs. old information:

(37) ty mnie, na polskiej ziemi,
    2SG:NOM 1SG:ACC on Polish:ADJ:F soil
    polskiego słowa zabronisz
    Polish:ADJ:M:GEN word:GEN forbid:FUT;2SG
    ‘Are you going to forbid me the use of a Polish word on Polish soil?’

A more precise rendering of this example would be:

‘You [of all people] are going to forbid me [and not somebody else] the use of a Polish [and not some other] word on Polish [and not some other] soil?’

The importance of linear precedence is that it significantly increases the number of formal means available in the language and consequently the number of functions that can be coded.

11 Conclusions

The study has demonstrated that the formal domain of linear order consists of several coding means, some of which are:

• Default coding within the given functional domain;
• Position with respect to some category (in the languages discussed, this is limited to the position with respect to the verb);
• Alternation to the default order and to the coding by position;
• Relative order of two identical categories with the same inflectional marking;
• Precedence of one element over another in a sequence consisting of several items consisting of the same lexical category with the same inflectional marking;

It is possible that further studies will discover other linear coding means.

The study has also shown that linear means code a variety of functions within different functional domains:

• A language may have several linear coding means available in the same or different functional domains (e.g., English in the domain of modality and in the
coding of grammatical relations; Polish in the domain of grammatical relations and in a variety of pragmatic functions).

- A single clause may deploy several linear coding means.
- Two identical linear orders may have different functions across languages.

Taking linear order as a coding means and correlating it with other coding means available in the language allows us to explain a number of similarities and differences across languages, in particular:

- Why a language may have a fixed order for a specific category and flexible order for other categories. This is the case with the fixed position of the verb in Japanese, Korean, Cushitic and Omotic languages;
- Why a fixed position for a given lexical category is not a coding means and does not create a positional coding means (verbs in verb-initial and verb-final languages, e.g. Hdi, Celtic languages, Japanese, Korean, Cushitic, and Omotic languages);
- Why coding by precedence is available when the language has extensive system of inflectional marking on nouns and verbs;
- Why changes in the default linear order in a specific domain trigger compensational changes (Koyra Chiini, Cushitic languages)

The findings of this paper have serious implications for the study of linear order change. That issue will be undertaken in another study.

References
Dryer, Matthew S. 2005a. Order of Subject, Object and Verb. In Haspelmath, et. al.,
Chapter 81.
Dryer, Matthew S. 2005c. Relationship between the order of object and verb and the order of relative clause and noun. In Haspelmath, et. al., Chapter 96.


Sources: Sources to the Polish Frequentative Dictionary. Warsaw.


*The work on this study was supported by Pays de la Loire Chaire Régional de Chercheur Étranger. Dennis Creissels, Shlomo Izre’el, Il-il Malibert, Amina Mettouchi, Saeko Ogihara, Graziano Savá, and Erin Shay, shared with me their knowledge, their data, their insights, and their time. Comments from Claude Hagège and Lionel Galand, were helpful for the write-up of the final version of this paper. To Zygmunt Saloni I am grateful for the access to the Source to the Polish Frequentative Dictionary. Without the knowledge of the speakers of Gidar, Pero, Hdi, Lele, and Kafa (Shiferaw Hajito), this study would not have been possible. Erin Shay, in addition to substantial comments, offered most important editorial assistance. None of the people from whom I have benefited so much is any way responsible for the errors of fact or interpretation.*