WHAT COMES FIRST, WHAT COMES NEXT: INFORMATION PACKAGING IN WRITTEN AND SPOKEN LANGUAGE

VLADISLAV SMOLKA

ABSTRACT

The paper explores similarities and differences in the strategies of structuring information at sentence level in spoken and written language, respectively. In particular, it is concerned with the position of the rheme in the sentence in the two different modalities of language, and with the application and correlation of the end-focus and the end-weight principles.

The assumption is that while there is a general tendency in both written and spoken language to place the focus in or close to the final position, owing to the limitations imposed by short-term memory capacity (and possibly by other factors), for the sake of easy processibility, it may occasionally be more felicitous in spoken language to place the rhematic element in the initial position or at least close to the beginning of the sentence.

The paper aims to identify differences in the function of selected grammatical structures in written and spoken language, respectively, and to point out circumstances under which initial focus is a convenient alternative to the usual end-focus principle.

Keywords: word order, functional sentence perspective, end-focus, written language, spoken language

1. Introduction

Linearity, one of the essential properties of language, means that components of language structures can only be arranged in a linear sequence. At sentence level, this arrangement may be described as word order, or, more fittingly, as clause constituent order. The ordering of the constituents is not accidental but governed by a set of principles whose relative importance varies across languages and which serve a variety of purposes, such as indicating grammatical structure, semantic relationship between constituents, achieving an intended distribution of information over the sentence, emphasising particular elements, creating cohesion, making sentences easy to understand, etc.

Owing to the typological characteristics of the current English language, the main function of word order is to indicate the grammatical (syntactic) status of an element as

a clause constituent. As a result, positional mobility of clause constituents within sentences is severely constrained; a canonical sequence in a declarative sentence starts with the subject and continues with the verb followed by one or more postverbal constituents, such as the object (or the subject complement) and the adverbial – a sequence known in short as SVOMPT.

The respective principles governing the linear arrangement of constituents may work in agreement or contradict one another. In the latter case, the actual linear arrangement of the sentence results from their relative hierarchy, that is, how strongly they assert themselves in a given communicative situation. Viewed in this perspective, the grammatical principle of standard sequence of clause constituents often works in harmony with the tendency to put the most important information at or close to the end of the sentence, the end-focus principle (Quirk et al., 1985: 1356–1357), and with the tendency to position short, structurally light structures before longer, structurally heavier ones, the end-weight principle (Quirk et al., 1985: 1361–1362). This is especially true in the case of the subject versus the other clause constituents: the subject, the only constituent whose canonical position is preverbal, tends to be shorter and less informative than the postverbal constituents (object, subject complement). However, the subject may occasionally be long or informationally important, yet this does not necessarily mean that it can be removed from the initial position.

Additionally, the principles of end-focus and end-weight are understood as naturally correlated: it seems to be a safe assumption that structurally heavy constituents (relatively longer and syntactically more complex) tend to convey more information than short ones. Moreover, the longer constituents, expressed in full, can be expected to convey new information, while the short ones, expressed by pro-forms, are associated with old information. This makes the structurally heavy elements potentially focal/rhematic. Still, even this correlation does not apply absolutely, as can be seen from the following examples (Huddleston and Pullum, 2002: 1371).

(1) One of his daughters was running a computer store, while the other was still at university, reading law. The one running the computer store earned a lot of money.
(2) There's a toad in the large pool outside the barn.

In (1) the underlined subject of the second sentence is long but thematic (conveying context-dependent information) and remains in the canonical initial position, breaking the principle of end-weight, while conforming both to the grammatical word order and the principle of end-focus. Conversely, the rhematic subject in (2), used within the existential *there is* construction, conveys new, context-independent information and, though itself short, occupies a postverbal position, conforming partly to the principle of end-focus; in this example the linear distribution of clause constituent results from the application of the existential construction.

The relative strength of the principle of standard distribution of clause constituents in English means that the position of the rheme (focus) depends to a large extent on the grammatical structure of the sentence. Firbas (1992: 66–69) distinguishes between two essential arrangements of sentences on semantic grounds: the presentation scale and the quality scale (cf. Adam, 2013: 45–46). The presentation scale perspectives the sentence

towards the subject as the most dynamic element (the phenomenon to be presented), whereas the quality scale perspectives the sentence away from the subject (the bearer of the quality) towards the quality itself, which is more dynamic than the subject. Consequently, owing to the position of the subject in English, the vast majority of sentences representing the quality scale are characterised by a post-verbal rheme, while in those representing the presentation scale, the position of the rheme depends on the structural subtype employed: it is postverbal, though not necessarily final, in the frequent *there is* construction, final in sentences employing initial thematic adverbial followed by S-V inversion, while initial rheme is the necessary consequence of application of structures involving rhematic subjects and conforming to the standard linear distribution of clause constituents. It is to be noted that although the *there is* structure represents the most frequent subtype of presentation scale implementation, the range of predicate verbs it can employ is rather limited, while rhematic preverbal subjects readily combine with a much larger set of predicate verbs.

2. Information-packaging strategies

Despite the limited mobility of clause constituents, a change of rheme position can be achieved in English by a number of strategies referred to as information-packaging strategies (Huddleston and Pullum, 2002: 1366):
– simple word-order change (reordering), which is limited in English:
  – preposing: *This one she accepted*.
  – postposing: *I made without delay all the changes wanted*.
  – inversion: *On board were two nurses*.
– realignment (alternative pairing of syntactic functions with semantic constituents in the sentence):
  – existential: *There is a frog in the pool*.
  – extraposition: *It is clear that he's guilty*.
  – left dislocation: *That money I gave her, it must have disappeared*.
  – right dislocation: *They're still here, the people from next door*.
  – cleft: *It was you who broke it*.
  – passive: *The car was taken by Kim*.
To this list may be added one further category, provisionally called reassignment, which, although related to those listed under realignment, goes further away from the default version of the sentence by replacing the original verb with one allowing an alternative assignment of syntactic functions to semantic units. So, for example, the existential construction *There are a large number of photographs in the book* may be reformulated as follows:

(3) The book contains a large number of photographs. (BNC: HHN 295)

Compared to the *there is* construction, the syntactic functions of the nominal constituents have changed: the original adverbial (*in*) *the book* has been changed into the subject, and the notional subject *a large number of photographs* became the object, while

the semantic function of *the book*, the locative, has been preserved, regardless of its syntactic status.

With respect to the principles and operations described above, two comments must be made: while changes in constituent position may be prompted by an effort to arrive at a different linear distribution of information over the sentence, it does not mean that a movement of a constituent automatically leads to a change in its information status. In reality, the word order is the weakest of the indicators of FSP, following the context, semantics, and, in spoken communication, the prosody.

Secondly, owing to the limited mobility of clause constituents in English, the end-focus principle cannot be understood as literally referring to the absolutely final position in the sentence; rather, the end has to be interpreted in the sense of postverbal, near-final position. The truly final position is often occupied by thematic elements following a postverbal rheme. In such cases, preposing (fronting) of such thematic elements, if possible, does not affect the postverbal position of the rheme, but rather makes it more explicit.

(4) I read <u>Dr Zhivago</u> eight years ago, an English edition a friend found in Yugoslavia. (BNC: AK4 718)
(5) Eight years ago I read <u>Dr Zhivago</u>, an English edition a friend found in Yugoslavia.

The intuitive observance of the end-focus principle by native speakers is probably motivated by processibility concerns: the (near-) final position of an element is a clue for the reader/listener indicating communicative importance. Yet, considering the pragmatic differences in constructing and processing written and spoken language, it is not unreasonable to assume that the strategies of rheme placement might differ in the two modalities. Biber et al. (1999: 1067) give three interrelated principles applying to online production of spontaneous spoken language, all of them affecting the manner in which spoken language is structured: keep talking; limited planning ahead; qualification of what has been said. What, among other things, follows from these general principles is that the structures actually used in spontaneous spoken language tend to be short and potentially imperfect or incomplete, and, consequently, have to be elaborated on and modified retrospectively. According to the authors, the end-weight principle asserts itself more extremely in spoken than in written language, because subjects in spoken language tend to be very short and often consist of a single word (ibid.). Generalisations concerning the application of the end-focus principle in spoken language are much more difficult to make. Nevertheless, Biber et al. suggest that retrospective modification of the message involves tagging on as an afterthought some elements which, in a logically structured and integrated sentence, would have been placed earlier (ibid.). If this is interpreted as reference to the end-focus principle, it is possible to assume that spoken language may occasionally employ initial or near-initial focus. If this is so, the most informative element occurs relatively early in the sentence, while elements carrying less important information follow.

Generalisations about the differences between spoken and written language are made difficult by the fact that, unless the manner of presentation is used as the only criterion, the boundary between the two modalities may be difficult to draw. Neither spoken nor written language are homogeneous systems, but rather constitute a scale with a large

overlapping territory in between. The essential factor here seems to be the amount of planning: a public speech, though presented orally, definitely involves more planning than a spontaneous written memo. Consequently, the choice of structures and the linear organisation of constituents are likely to be affected by planning constraints.

Another consideration of the spoken-written distinction touches upon the signals of importance available in the respective modalities. While written language has to rely on the interplay of three non-prosodic factors (context, semantics, and linearity) (Firbas, 1992: 51), spoken language can use prosody, namely the position of the intonation nucleus, either to confirm the information structure suggested by the non-prosodic factors, or to re-evaluate it. In the latter case, prosody overrides all the other factors combined. What follows from this is that in spoken language there should be less need to use deviations from the usual grammatical linear arrangement of clause constituents as an indicator of rhematicity, since the rhematic item is sufficiently marked by intonation.

This assumption is corroborated by quantitative data (Biber et al., 1999: 909–910) suggesting that, compared with academic prose and fiction, fronting in conversation of a core constituent is three to four times less frequent, and the majority of examples represent object fronting, while predication fronting is extremely rare.

3. A corpus-based study of two information-packaging constructions

To examine the use and frequency of restructuring/realignment operations in authentic samples of written and spoken language, respectively, two constructions have been examined as they occurred in the British National Corpus: the *there is* construction, and the tough movement.

3.1 *There is* construction

The *there is* construction typically serves as a means of introducing into communication a new, context-independent nominal element functioning as the notional subject of the sentence. This subject is almost invariably rhematic, unless the sentence is repeated to create a contrast, representing a case of second instance (Firbas, 1992: 111). The most common subtype of the *there is* construction, the existential-locative construction, includes a thematic adverbial of place, which usually follows the subject, but can also be fronted, particularly when it conveys context-independent information, constituting the most dynamic variant of diatheme (Firbas, 1992: 81). The position of the adverbial does not affect the rhematic status of the subject; however, a sentence with a fronted adverbial corresponds more closely to the end-focus principle, because the notional subject occupies the absolutely final position. The purpose of the research was to compare the proportion of *there is* constructions with fronted adverbials in written and in spoken communication. The hypothesis was that there would be more instances of fronting in written language.

According to Adam (2013: 62–69), the existential construction represents the most frequent subtype of presentation scale implementation (65%), which is more than the other three types combined: rhematic subject in the preverbal position (25%); front-

ed adverbial followed by S-V inversion (8%); and locative thematic subject followed by rhematic object (2%). Adam's quantitative data are based on two types of texts, fictional narrative and biblical texts, neither of which can be considered truly spoken language. Additionally, the frequency of the respective implementation subtypes varies even in the two above-mentioned text types, with the proportions indicated representing the overall distribution in both. The prevalence of the *there is* construction may be explained by the fact that it is mostly used with the verb *be*, whose existential semantics makes it ideally suited for the purpose of presentation, while other verbs are of marginal importance. Conversely, the grammatically canonical presentation construction employing a rhematic subject in preverbal position usually must employ a verb where the meaning of existence or appearance is an extension of other semantic features, and in most cases cannot be used with the predicate *be* as an alternative to the *there is* construction, unless accompanied with the subject-verb inversion:

(6) All of a sudden, there was a picturesque castle on the hill. (Adam, 2013: 59)
(7) All of a sudden, a picturesque castle appeared on the hill.
(8) All of a sudden, on the hill was a picturesque castle.
(9) ?All of a sudden, a picturesque castle was on the hill.

Although Adam does not include (9) among the model presentation sentences, it would be marginally acceptable because the adverbial *all of a sudden* renders the structure dynamic, corresponding in meaning to the semantics of the verb *appear*. Without the initial adverbial, and therefore interpreted statically, the sentence would be infelicitous, while the version employing the verb *appear* is still perfectly acceptable:

(10) *A picturesque castle was on the hill.
(11) A picturesque castle appeared on the hill.

In the course of the present research, the *there is* construction proved to be a frequent structural pattern in BNC overall, with 245,828 hits in 3,922 different texts amounting to a total of 98,313,429 words; i.e. a frequency of 2,500.45 instances per million words. The frequency was significantly higher in spoken (4,038.96 instances per million words) than in written language (2,318.26 instances per million words).

Three hundred examples of the *there is* construction were retrieved from the written and the oral part of the BNC, respectively, and the relative proportions of three subtypes were established: with fronted adverbial, with non-fronted adverbial, and without an adverbial explicitly present. The results are given in the Table below.

Table 1. Distribution of fronted/non-fronted adverbials in spoken and written texts

there is	total		fronted Adv		non-fronted Adv		no Adv	
	absolute	per cent	absolute	per cent	absolute	per cent	absolute	per cent
written	300	100%	59	19.7%	49	16.3%	192	64%
spoken	300	100%	12	3%	75	25%	213	71%

One surprising finding is the generally high proportion of *there is* constructions without an explicit adverbial. This suggests either a high frequency of the purely existential

subtype, or, possibly, cases of ellipsis of the adverbial, which is explicitly present in the preceding context, or at least implied by it. On the other hand, the difference in the frequency of structures without an adverbial is not strikingly different in spoken and written material. What is sharply different, however, is the proportion of adverbial fronting, which is negligible in spoken language, while amounting to roughly one fifth of the examples in written language. The disproportion would be even more prominent were it not for the fact that a single short section of spoken language contained three examples of fronted adverbials used within a repeated pattern.

(12)   Then within the teams, <u>in the Policy Team there were</u> three Policy Development Officers, <u>in the Neighbourhood Development Team there were</u> two Development Officers and <u>in the Community Development Team there are</u> – one, two – there are – well there are above these six officers that you could i-, that you could identify, but there are other people that relate to them, but then it gets a bit complicated so we keep it at that. (D95 428)

Another characteristic of many examples retrieved from the spoken part of the corpus is vagueness of reference, with the adverb *there* used repeatedly both as the formal subject and a referential expression pointing to some spatial setting within the same sentence.

(13)   All I do know is that we're very conscious a great deal of money's been spent by the Council and by the Trust to try and make the place outside more inviting those tubs been put <u>there</u> er they been planted etc there's a lot of litter <u>there</u> we've put litter bins <u>there</u> <u>there</u>'s taxi rank <u>there</u> <u>there</u>'s lot's of things been put <u>there</u> I think the problem is it isn't the people who do that... (D91 758)

Moreover, some of the adverbials in the initial position are adverbials of indefinite frequency or conjuncts, rather than fronted adverbials of place, which may follow in the postverbal position in the same sentence.

(14)   <u>Sometimes</u> there's a good write up beforehand sometimes there's nothing. (D91 636)
(15)   <u>Then</u> there's a lady over here who comes or used to come on a regular basis a couple of times a month she comes now a couple of times a year. (D91 616)

Conversely, examples representing written language often include lengthy fronted adverbials, sometimes two at a time.

(16)   <u>In other political states during the twentieth century</u>, there has been strong support for the view that art should serve a social purpose. (A04 307)

The findings suggest that adverbial fronting may serve different purposes in written and spoken language. As shown in the Table above, in written language fronted adverbials slightly outnumbered non-fronted ones (19.7% to 16.3%). Although the use of some initial adverbials, especially short ones, serves the purpose of textual connection, creating

cohesion of the text, it is still reasonable to assume that, in written language, a proportion of adverbial fronting in *there is* constructions is used intentionally to boost the rhematic status of the notional subject by position (end-focus) in addition to its rhematicity already resulting from the *there is* construction. On the other hand, the conspicuous scarcity of fronted adverbials in the *there is* construction in spoken language may be due to an unconscious tendency to present the most important information early in the sentence (near-initial focus). Although the *there is* construction is traditionally interpreted as a means of shifting a rhematic subject into the postverbal position, i.e. removing it from the beginning of the sentence, strictly from the point of spoken language it can be perceived as a short signal of communicative importance of the subject: *there is* tends to be pronounced as a single unstressed syllable with a reduced vowel, or, in the case of *there are/were*, as two unstressed syllables: in both cases a unit of extremely low prosodic prominence. In some respects, it is reminiscent of the traditional English rhythmic pattern used in poetry, the iambic metre (Crystal, 1997: 74; Halliday, 1989: 49–50), as well as the well-known English tendency to start sentences with rhythmically light words (Mathesius, 1975: 159).

3.2 Tough movement

Another syntactic pattern explored in BNC from a quantitative perspective was the tough movement.

(17)   ... and my word he was difficult to follow! (KC0 7824)

Unlike the *there is* construction, where the subject is almost invariably rhematic, the tough movement is less transparent with respect to FSP: the rheme may be either in the main clause (typically carried by the postverbal adjective) or in the subordinate clause (the infinitive), while the subject of the construction is typically thematic; the construction is also known as the object-to-subject raising. In fact, the thematicity of the original object of the infinitive that has become the subject is one of the main reasons for the use of the tough movement; English sentences tend to start with informationally light, context dependent elements, and such elements naturally occupy the initial position if they are encoded as subjects. The tough movement may be interpreted as an alternative to structures involving a subject *that*-clause, either in the usual extraposition or, much more rarely, in the initial position.

(18)   ... it was difficult to follow him!
(19)   ... to follow him was difficult!

However, while the non-extraposed infinitive subject clause in (19) is unambiguous in terms of FSP, the rheme being the subject complement *difficult*, the extraposed variant in (18) is open to both interpretations: the rheme is either the adjective *difficult* or the infinitive *to follow*, depending largely on the contextual factor. The FSP interpretation of tough movement follows the same principles: the component conveying context-independent information will be likely to constitute the rheme. Analysis of tough movement

constructions in the corpus revealed a prevalence of rhematic infinitives over rhematic subject complements. In this respect, the tough movement may be considered as a structure implementing the end-focus principle, especially in comparison with the extraposed subject *that*-clause.

Accordingly, the BNC was searched for structures containing the following pattern: *be + hard/easy/difficult/impossible + infinitive*. The assumption was that these adjectives were among the most frequent ones occurring in tough movement (Biber et al., 1999: 728). The search yielded a combination of *it*-extraposition, tough movement, and possibly some other accidental structures. Consequently, the examples retrieved from both the written and the spoken component of BNC were analysed manually and the number of actual tough movement constructions was established in each. The number of both of the *it*-extraposition and the tough movement combined, and the actual number of tough movement constructions was then recalculated in relation to the whole written and spoken parts of the corpus, with the following results.

Table 2. Frequency of tough movement constructions in spoken and written texts

	Number of words	Number of *it*-extraposition + tough movement	per million words	Number of tough movement alone	per million words
written	87,903,571	5,615	63.88	2,455	27.92
spoken	10,409,858	265	25.46	116	11.14

Admittedly, owing to the method used to retrieve the examples, which included search for lexical items, the results are necessarily incomplete. Still, the differences in the relative frequency of tough movement in spoken and written language are prominent enough to suggest systematic differences. If tough movement is considered as a means of implementing the end-focus or near-end focus, corpus findings provide a clear indication that written language employs it to a much higher degree than spoken language.

4. Conclusions

Using the BNC as a source of data for the analysis proved to be a double-edged sword: on the one hand, it provided a large sample of authentic material, increasing the statistical validity of findings, on the other hand, the sheer volume of raw data turned out to be a considerable processing challenge. Clearly, general differences in information structuring in written and spoken language cannot be reduced to differences identified in the use of the two syntactic constructions explored. However, as both represent fairly frequent structural patterns, and as the differences in their distribution are sufficiently prominent, it is reasonable to assume that the findings do corroborate the initial hypothesis that spoken language uses fewer word-order modifications to conform to the end-focus principle than written language. Linear arrangement of clause constituents with the rheme at or close to the end may therefore be considered a product of planned production, which is characteristic of written language. Conversely, spoken language follows the end-focus principle less consistently, partly because of lack of planning time, partly because it uses

prosody as a signal of communicative importance of a given element, regardless of its position in the sentence.

Further research will be needed before the tendencies identified can be considered conclusive. Such research might focus on the distribution of other FSP-significant syntactic structures in the two modalities of language and should address issues like the differences in the length and complexity of clauses (clausal units) in written and spoken language, exploring the assumption that longer structures yield more potential for word-order and FSP modifications than shorter ones.

REFERENCES

Adam, M. (2013) *Presentation Sentences* (*Syntax, Semantics and FSP*). Brno: Masaryk University.
Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan (1999) *Longman Grammar of Spoken and Written English*. Harlow: Longman.
Crystal, D. (1997) *The Cambridge Encyclopaedia of Language*, 2nd ed. Cambridge: Cambridge University Press.
Firbas, J. (1992) *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: Cambridge University Press.
Halliday, M. A. K. (1989) *Spoken and Written Language,* 2nd ed. Oxford: Oxford University Press.
Huddleston, R., G. K. Pullum et al. (2002) *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
Quirk, R., S. Greenbaum, G. Leech and J. Svartvik (1985) *A Comprehensive Grammar of the English Language*. London: Longman.
Mathesius, V. (1975) *A Functional Analysis of Present-Day English on a General Linguistic Basis.* Prague: Academia.

Sources
British National Corpus (BYU-BNC). Available at: http://corpus.byu.edu/ [2016-08-29].

CO DŘÍVE A CO POTOM: INFORMAČNÍ STRUKTURA PSANÉHO A MLUVENÉHO JAZYKA

Resumé

Základní funkcí slovosledu v současné angličtině je signalizovat syntaktické funkce jednotlivých konstituentů pomocí jejich standardního lineárního uspořádání. Tím je ovšem významně omezena variabilita slovosledu. Ostatní principy uspořádání větných členů se proto mohou uplatnit pouze v omezené míře, případně jinými prostředky než pouhou změnou slovosledu v rámci stejné syntaktické konstrukce, jak je to obvyklé v češtině.

Vedle gramatického principu slovosledu v angličtině fungují dva další obecné principy uspořádání, tzv. end-weight a end-focus, tedy tendence řadit jednotlivé konstituenty od kratších a strukturně jednodušších k delším a strukturně složitějším, a od informačně méně zatížených (tematických) k informačně nejdůležitějším (rematickým).

Cílem výzkumu založeného na analýze autentických příkladů z British National Corpus bylo zjistit, zda se princip end-focus uplatňuje stejným způsobem v psaném a mluveném jazyce. K tomuto účelu byly vyhledány příklady dvou běžných syntaktických konstrukcí: existenciální konstrukce *there is* a tzv. tough

movement. První z nich slouží k uvedení rematického podmětu v postverbální pozici, a to bez ohledu na umístění příslovečného určení, které se vedle podmětu v této konstrukci typicky vyskytuje, druhá umožňuje konstruovat tematický prvek jako podmět, tedy dosáhnout základního rozložení výpovědní dynamičnosti s rematickým prvkem na konci.

U existenciální konstrukce bylo zjištěno, že pokud se v ní explicitně vyskytuje příslovečné určení, je v psaných textech poměr jeho koncové a počáteční pozice téměř vyrovnaný, kdežto u mluveného jazyka je kanonická koncová pozice příslovečného určení téměř osmkrát častější ve srovnání s počáteční pozicí. Zdá se tedy, že v psaných textech se tendence k uplatnění principu end-focus u této konstrukce projevuje mnohem silněji než v textech mluvených.

Frekvence konstrukce tough movement se ukázala být téměř třikrát vyšší v jazyce psaném než mluveném. Je-li tato konstrukce chápána jako prostředek dosažení lineárního růstu výpovědní dynamičnosti, svědčí to rovněž o silnějším uplatnění principu end-focus v psaném jazyce. Společným jmenovatelem obou zjištění může být skutečnost, že zatímco v mluveném jazyce funguje jako důležitý indikátor intonace, která je schopna převážit nad ostatními neprozodickými indikátory FSP, v psaném projevu má při absenci intonace autor potřebu naznačit rematičnost větného členu jeho koncovým umístěním.

Vladislav Smolka
Department of English Studies
University of South Bohemia in České Budějovice
smolka@pf.jcu.cz