

A Short Analysis of Information Structures of English Sentences

Shufang Cheng

Zhengzhou Institute of Aeronautical and Industrial Management, Zhengzhou, China

Email: sophiacsf@126.com

Abstract—This paper analyses information structures of English sentences broadly by comprising structural and semantic properties of utterances relating to the discourse status of their content and the states of the discourse participants. It mentions the concept of the notions like presupposition, focus, given vs. new, and topic vs. comment and discusses the relationships between them. Then it introduces the categories of information status and the animacy hierarchy. Finally it states the significance of information structure of English sentences.

Index Terms—information structure (IS), information status, animacy hierarchy

I. INTRODUCTION

Information structure (IS) is one aspect of the textual organization of language. It refers to the organization of a text in terms of the functions Given and New. These are often conflated with Theme and Rheme under the single heading “topic and comment”; the latter, however, is a complex notion and the association of Theme with Given, Rheme with New. IS is the encoding of the relative salience of the constituents of a clause, especially nominals, and is realized as choices among alternative syntactic arrangements. The IS of a particular clause is determined by the larger sentence or discourse of which it is a part (i.e., its context). The communicative effect of the IS is to foreground certain aspects of the message of the clause, but to background others. The need to encode IS is a language universal, but the formal means to do so vary widely across the languages of the world. In English, word order is strongly determined by syntactic conditions, such as the encoding of grammatical relations like subject and object, so here comes an important issue that how English indicates IS.

II. TERMINOLOGY EXPLANATION

Before the discussion of this issue, I will provide the explanations of some terms that will be mentioned later.

Given:

Given refers to “information that is presented by the speaker as recoverable to the listener” (Halliday, 2000, p.298). It is the information mentioned before in the context and shared by both the speaker and the listener). It serves as a link or bridge, or mark of relevance to new information.

New:

New refers to “information that is presented by the speaker as not recoverable to the listener” (Halliday, 2000, p.298). It is not really available in the context. It serves as the speaker’s contribution that is based on the Given.

Topic:

Topic refers to “the entity, in a sentence, about which something is said” (Asher.R.E. and Simpson.J.M.Y, 1994, p.5181). e.g., “the books” in “The books were all on the bookshelf” is the topic here.

Comment:

Comment refers to “the part of a sentence which says something about the topic of the sentence” (Asher R.E. and Simpson J.M.Y, 1994, p.5103). e.g., “were on his head” in “His spectacles were on his head” is the comment here.

Presupposition:

“The assumption made in an utterance or discourse, the given which can be inferred from what is stated (Asher R.E. and Simpson J.M.Y, 1994, p.5160). e.g., from “Has your cat recovered?” it can be inferred that listener owns a cat and that the cat has suffered some illness, accident, or the like.

Focus:

“The new material in a sentence (Asher.R.E. and Simpson J.M.Y, 1994, p.5123).

e.g., if “John fell” is the answer to “Who fell?” “John” is the focus, not so if the question is “What happened?”

III. THE RELATIONSHIPS BETWEEN TOPIC/COMMENT, PRESUPPOSED/FOCUS AND GIVEN/NEW

There are no all-encompassing definitions of topic and focus in the literature. Both terms cover phenomena belonging to the whole spectrum of syntax, semantics, and pragmatics, with an extension to the phonological level. Perhaps the

most controversial concepts lie at the pragmatic level.

With different terminology, these two notions have been opposed to each other as ‘given’ (topic) versus ‘new’ (focus).

It is a fact that has been widely recognized, but for which up till now poor theoretical explanations have been given, that across natural languages there is a high tendency in unmarked sentences to map the contextually bound stretch of the sentence on to the subject and the contextually unbound stretch on the predicate. This tendency as well as the fact that the overwhelming majority of the world’s languages are either SVO or SOV raises a problem concerning linearity. There is in fact no a priori reason why the contextually bound (given/less dynamic/presupposed) part of the sentence should come first.

The same correlation shows itself in the semantic definition associated with the two terms (note, however, that at this level the term “comment” is often found instead of focus). Here topic is “what is being spoken about”, focus (or comment) is “what is being said on what is being spoken about”.

When it comes to a purely syntactic definition of topic and focus, apparently it is the linear dimension of the sentence that is essentially involved. Different models have variously assessed the property of being a topic as the occurrence of a constituent in the first position of the sentence. This generalization, however, can be questioned. The property of being an argument of the verb is no less important as a syntactic criterion than purely linear considerations. By a larger consensus in the literature only those constituents that convey grammatical functions are considered as candidates for the topic function. Thus languages may have topics that do not occur in the first position of the sentence; to take one example, the so-called circumstantial elements that express the temporal or spatial setting may be placed in the first position, as in the sentence “Yesterday Mary was in great depression”. Here the topic is not “yesterday”, but the constituent with the subject function. Thus the idea seems well-founded that in general the first position should be differentiated from topic position, although there will be many cases in which the first position is the topic position. Things are further complicated by the fact that the topic often coincides with a phrase with multiple constituents.

Similar problems are faced in the attempt to obtain a syntactic definition of focus. Here again what is crucial is not merely the position inside the sentence, but also the categorical or functional nature of the constituent involved. Following a purely linear criterion, in fact, one could be led to assume that, at least in unmarked (i.e., nonemphatic) sentences, the focus position is the final one, since the linear dimension of the sentence can be conceived as a serial process of adding information quanta, each quantum conveying a higher information value than its antecedent. This assumption is wrong for two reasons: first, as in the case of topic, focus often does not coincide with a single constituent, but with a configuration of constituents. To resume the preceding example, in the sentence “Yesterday Mary was in a bad mood” the whole string “was in a bad mood” is the focus, or more precisely, the “broader focus”, although inside this domain, some constituents are more focal than others. For example, inside the VP (verb phrase) domain the focus proper is the NP (noun phrase) “a bad mood”; inside the NP domain, in unmarked sentences, the general consensus would be that the focus proper is the modifier “bad” (this is what is called ‘narrow focus’). Second, the final position could be occupied by a circumstantial element, which is a typical nonargument of the verb: the previous assignment of “foci” holds true even in “Mary was in a bad mood yesterday” as well as in “Mary was in a bad mood during her stay in Japan”, when these sentences are uttered with normal intonation contours (i.e., “yesterday” and “during her stay in Japan” have either no nucleus or a secondary one). Similar considerations hold true for other constituents occurring in the final position that do not have a strong dependency relation with the predicate frame and thus are extrasentential (e.g., “he said” in *John was upset by the War, he said*) or appositional (e.g., “Peter” in *I have just met my brother, Peter*; note that in “*I have just met my brother Peter*”, with no pause between “brother” and “Peter” and the nucleus on “Peter”, the latter constituent is the focus proper of the sentence).

Generally, topic occurs initially in a sentence and focus, finally, but not all of English word order is rigidly fixed. Some alternatives are commonly found, for example, “Today Mary is reading” and “Mary is reading today” and “The children are playing in the yard” and “In the yard the children are playing”. That these differences encode information structure is apparent from constructing mini-dialogues. For example:

- | | |
|---------------------------------------|-----|
| Question: When is Mary reading? | (1) |
| Answer: Mary is reading today. | (2) |
| PRESUPPOSED FOCUS | |
| Answer: ? Today Mary is reading. | (3) |
| FOCUS PRESUPPOSED | |
| Question: Is Mary writing today? | (4) |
| Answer: No, today Mary is reading. | (5) |
| PRESUPPOSED FOCUS | |

The normal position for focused information in English is sentence final, but English is flexible and it does allow some deviations from this rule. However, the language does require that this deviation be signaled. Thus, the common way in English to encode focused information that is not sentence final is by a high falling pitch on the constituent (sometimes called “emphatic stress”). For example,

- | | |
|-------------------------------------|-----|
| Question: Who saw Bill? | (6) |
| Answer: ? Bill was seen by John. | (7) |

PRESUPPOSED FOCUS

Answer: John || saw Bill. (8)

FOCUS PRESUPPOSED

Question: Which one is John? (9)

Answer: ? John is || the manager.

PRESUPPOSED FOCUS

Answer: The manager is || John. (10)

FOCUS PRESUPPOSED

Another way of altering the normal English ordering of presupposed information before focused is through it-cleft constructions. Thus the meaning of the sentence "Mike ate an icecream" could be expressed in it-cleft constructions, indicating unusual encodings of information structure. For example:

It was Mike || who ate an icecream. (11)

FOCUS PRESUPPOSED

It was an icecream || that Mike ate. (12)

FOCUS PRESUPPOSED

The relative clauses in it-cleft constructions are typically used to reintroduce information already known, but not foregrounded in the present discourse context. They are presupposed information, but are now being reintroduced and partially highlighted in this context; hence, they appear at the end of the sentence and in the position of focused information. It-clefts, contrast with wh-clefts, which have expected ordering of presupposed information before focused. For example:

The one who ate the icecream was || Mike. (13)

PRESUPPOSED FOCUS

What Mike ate was || an icecream. (14)

PRESUPPOSED FOCUS

Given their different information structure, it should come as no surprise that wh-clefts are found in different discourse contexts than it-clefts. The relative clauses of wh-clefts (presupposed) express information already foregrounded, assumed to be in the hearer's mind at the moment. Hence they have no need to appear in focus position, in fact, rather properly belong in the topic position they do indeed occupy.

The encoding of the topic-comment information structure in English is complicated by the common sentence initial position for subject nominal. There is, in fact, a very strong correlation between the concepts of topic and subject in English. Thus, the typical way to express alternatives of topic choice is to select different subjects. This is very common in English. For example:

Tears || streamed down her face. (15)

TOPIC COMMENT

Her faces || streamed with tears.

Blood || flowed in the streets.

TOPIC COMMENT

The streets || flowed with blood. (16)

In these two examples the same predicate is used with different subject/ topic choices. In other cases different predicates must be used. For example,

Question: Where is the dot? (17)

TOPIC COMMENT

Answer: The dot || is || inside the circle. (18)

PRESUPPOSED FOCUS

Answer: ? The circle surrounds the dot. (19)

Question: Where is the circle? (20)

TOPIC COMMENT

Answer: The circle || is || around the dot. (21)

PRESUPPOSED FOCUS

Answer: ? The dot is inside the circle. (22)

Any construction in English that affects subject selection also typically encodes the topic-comment information structure. For example, extraposition is a topic-altering device:

That John will attend || is certain. (23)

TOPIC COMMENT

John || is certain to attend. (24)

TOPIC COMMENT

And, perhaps, the most common information-structure encoding device in English is the alteration between active and passive voice, for example,

The manager || sacked the pilots. (25)

TOPIC COMMENT

The pilots || were sacked by the manager. (26)

TOPIC COMMENT

Thus, the immediate discourse context determines the proper choice between active and passive, as with all information-structure alternatives. Consider these mini-dialogues:

Question: Who saw Bill? (27)

Answer: Bill || was seen by John. (28)

TOPIC COMMENT

Question: Whom did Bill see? (29)

Answer: John || was seen by Bill. (30)

TOPIC COMMENT

The second answer is bizarre because “Bill” is the natural topic, as established by the question. Hence, Bill should be subject and the answer is proper in the active voice: “Bill saw John”.

While there is a very close correlation in English between the notions of topic and subject, it is not the case that they are isomorphic; these are clear cases of topics which are not subjects. These are topics occurring sentence initially and preceding the subjects as the following sentences:

TOPIC COMMENT

Last night || I saw three movies in town. (31)

As for John || he is such a clown. (32)

Soukous || I think it’s the great African twentieth-century contribution to civilization. (33)

Such constructions are especially common in informal spoken speech styles. The first example is an ordinary topicalization, while the second and third examples illustrate left-dislocations, in which a pronoun marks the position where otherwise the topic nominal would be found. Left-dislocations are used in English to express a change in topic in discourse, such as when a new topic is introduced which supersedes a previous one (this is especially noticeable with “as for”). Ordinary topicalizations, on the other hand, can be used to prevent as topic any constituent which is presupposed. It may not be explicitly mentioned previously, but must be presupposed. For example:

Question: What’s your favorite dance music? (34)

TOPIC COMMENT

Answer: Soukous || my feet find irresistible. (35)

Soukous, a type of African dance music, is presupposed by the general cover term “dance music” in the question.

IV. CATEGORIES OF INFORMATION STATUS

A. *Given versus New Information*

Topics are typically presupposed information and they are the starting point of the sentence. Focused nominals, on the other hand, are the end goal of the sentence, the information which the speaker intends to introduce into the discourse. Hence the topic tends to occur toward the beginning of a sentence and focus toward the end. Topics are closely correlated with the given or old information, which is currently in the speaker’s awareness, while focused constituents are new information, just being introduced into the discourse. The concept of given information is more or less equivalent to presupposed, but new information need not (although it usually does) correspond only to the focused constituent.

Question: What happened? (36)

Answer: An enormous storm devastated Manila. (37)

Arguably, all the information in the answer is new, but only the nominal “Manila” is really the focus, which is usually demonstrated by high falling pitch.

There are other types of mismatches between topic and presupposed/given and focus and new. Some of these are exemplified by contrastive nominals. These are usually marked like focused constituents by a high falling pitch. Contrastive nominals are typically focused, but presupposed. For example:

Considering the suspects, only John has a motive. (38)

John is one of several suspects in a crime investigation; he is thus presupposed, yet new information is provided about him, as focus, as being the only person among them with a motive. Thus, he is set up in contrast to the other suspects, presupposed but focused. Another example of a focus-presupposed nominal would be in the following sentence in a discourse about dances. For example:

I like the lambada, but Sam likes the Cha-cha. (39)

People who like dancing are presupposed, and Sam is among them. But he is focused as being unusual in liking the Cha-cha; the goal of the sentence is to communicate this piece of gossip.

There are also examples of topics which are new information. Perhaps the best-known examples of these are found in the opening lines of stories, such as fairy tales. For example:

Once upon a time, a handsome prince lived in an old castle. (40)

“A handsome prince” is the topic of this sentence, but as it initiate a discourse, it is obviously not given information, but new.

B. Referential versus Nonreferential

Besides given/presupposed versus new, nominals may bear other kinds of information statuses, depending on the speaker's view of the listener's knowledge of their referents. A nominal is referential if the speaker intends that it refers to a particular entity in the world. For example, if someone rings a friend and asks "What are you doing?" and the response is "I'm looking for a dog". The nominal "dog" could be either referential or nonreferential. If the respondent has a particular dog in mind, for example, a pet, then "dog" would be referential as in

I'm looking for a dog. Shelby escaped from his cage. (41)

Here, Shelby is the dog's name. If, on the other hand, she/he is lonely and just wants a pet dog, then "dog" would be nonreferential, as in

I'm looking for a dog. I want a pet.

Pronouns like "I" "we" "you" "they" are typically referential, but at least two, "you" and "it", have nonreferential uses. For example:

You pay your money and you take your chances. (42)

It's raining in Melbourne, as usual. (43)

C. Definite versus Indefinite

A nominal is marked as definite when the speaker presupposes the listener can uniquely identify its referent; otherwise a nominal is indicated as indefinite. So if Mary goes up to John and announces, "Mike bought the dog", the definite article indicates that Mary presupposes John knows which dog she is talking about. If she does not, she would probably come back with additional information to help identify the referent of "the dog". On the other hand, if Mary believed that John had no knowledge of the particular dog, she would have initiated the conversation with the nominal indicated as indefinite: "Well, Mike just bought a dog."

Topic nominals are closely correlated with definiteness and because subject selection is largely equivalent to topic choice in English, indefinite subjects are sometimes impossible. For example:

| | | | |
|-------------|--|---------|---------------|
| TOPIC | | COMMENT | |
| John | | has | a new camera. |
| PRESUPPOSED | | FOCUS | |

(44)

| | | | |
|----------------|--|---------|--------|
| TOPIC | | COMMENT | |
| The new camera | | is | John's |
| PRESUPPOSED | | FOCUS | |

(45)

| | | | |
|---------------|--|---------|---------|
| TOPIC | | COMMENT | |
| ?A new camera | | is | John's. |

(46)

Note that the variant with the indefinite nominal as subject and, hence, topic, is ungrammatical.

D. Generic versus Specific

This distinction indicates whether a nominal refers to the entire class of its possible referents or a specific one. For example:

Dogs are easy pets to care for. (47)

The dog is my favorite among my pets, but it keeps escaping. (48)

"Dogs" in the first example is generic; the statement is meant to cover all animals classed as dogs. The nominal "the dog" in the second example is specific, as the sentence only applies to a particular dog, my pet dog. Generic nominals can be definite or indefinite. For example:

The dog is an easy pet to care for. (49)

A dog is an easy pet to care for. (50)

Dogs are easy pets to care for. (51)

V. THE ANIMACY HIERARCHY

The distinctions of information status for nominals considered above were all established by the discourse context. Now it is time to assess the information status of nominals as determined by inherent properties of their referents, the most significant of which properties is being one of the immediate speech act participants: the speaker or the addressee. Speaker and addressee generally correspond to the personal pronouns "I" and "you". The traditional definition of a pronoun as a word which stands for a noun is inaccurate in the case of "I" and "you" in that there is no possible nominal for which they stand. The referents "I" and "you" are not constant, but rather they change in the course of interaction, depending on who is doing the speaking and who is being spoken to. This interplay of shifting referents of "I" and "you" in the continuing speech act is a fundamental fact of language.

The elements which do fit the traditional definition of pronouns as the forms which take the place of nouns are the third-person pronouns: in English "he, she, it, they". These are fundamentally different from "I" and "you". Whereas "I" and "you" have the present speech participants as referents, a third-person pronoun may refer to any referent, other than the speech act participants. The third person is, in fact, a non person, its possible referents being restricted to non participants in the speech act.

There is a fundamental principle of salience in the system of persons. The speech act participants, speaker and

addressee, are more salient than the absent participants of the third person. In some languages the addressee is more salient than the speaker. In other languages the speaker is more salient than the addressee. And, in still other languages, speaker and addressee have equal salience. Many languages make further distinctions between different types of third person nominals. Nominals with animate referents are more salient than those with inanimate referents, and among animates, human referents are the most salient. Some languages make a further distinction among nominals with human referents, with proper nouns more salient than common nouns. Finally, third-person pronouns are generally more salient than full nominals. A hierarchy of inherent salience can be established:

Speaker/listener > third-person pronouns > human proper nouns > human common nouns > other animate nouns > inanimate nouns.

The inherent salience of a nominal often determines the packaging of a particular expression. Nominals higher on the animacy hierarchy tend to occupy more prominent syntactic positions than nominals lower on it.

VI. CONCLUSION

Information structure construes a piece of information from the complementary point of view, as something having news value---something the listener is being invited to attend to. It may not contain anything the listener has not heard before; a great deal of "news" is totally familiar, being simply contrasted or even reiterated. On the other hand, the entire message may consist of unknown information. The message is construed along prototypical lines as an equilibrium of the given and the new, with a climax in the form of a focal point of information. This focal point usually comes at the end: but unlike the Theme+Rheme, the Given+New structure is not signaled, in English, by word order---it is signaled by intonation, and specifically by pitch prominence, the point of maximum perturbation (falling, rising or complex) in the intonation contour. The principle behind this is clear: if the Theme always came first, and the New always came last, there would be no possibility of combining them; whereas one powerful form of a message---powerful because highly marked---is that in which the two are mapped on to one another.

REFERENCES

- [1] Asher, R.E. and Simpson (eds), J.M.Y. (1994). *The Encyclopedia of Language and Linguistics*, Volume 10. Oxford: Pergamon.
- [2] Keith Brown and Jim Miller (eds). (1999). *Concise Encyclopedia of Grammatical Categories*. Oxford: Pergamon.
- [3] M. A. K. Halliday. (1994). *An Introduction to Functional Grammar*. London: Edward Arnold
- [4] M. A. K. Halliday. Jonathan Webster (ed.). (2002). *On Grammar*. London: Continuum
- [5] M. A. K. Halliday. (2000). *Introduction and Grammar in British English*. The Hague: Mouton.
- [6] Thompson, G. (2000). *Introducing Functional Grammar*. London: Edward Arnold
- [7] Yongsheng Zhu and Shiqing Yan. (2005). *Reflections on Systemic-functional Linguistics*. Shanghai: Shanghai Foreign Language Education Press.

Shufang Cheng was born in Shanxi province, China in 1976. She received her M.A. degree in linguistics from Shanghai International Studies University, China in 2008.

She is currently an instructor in the School of Foreign Languages, Zhengzhou Institute of Aeronautical and Industrial Management, Zhengzhou, China. Her research interests include contrastive linguistics, pragmatics and college English teaching.