Jadwiga Linde-Usiekniewicz

Constituency and Information Structure

General claim

Many language phenomena, in particular some marginal ones, but also central ones can be explain through interaction of semantics, syntax and information structure
In order to support this claim we need to decide what is the semantics, what is the syntax, and what is the information structure, at least within our model

Rough outline of the model

The model of language I propose is a kind of functional model of language. The way it is designed it is supposed to mirror the activity of the speaker. It is therefore an encoding model of text production.
The model resembles MTM in that it takes as a starting point the sense the speaker is trying to convey, and as the result – the text produced. However, while MTM model is reversible in terms that it can mirror both the encoding procedure and the decoding one, here only the encoding mode is taken into account.
The main reason for that is that in real life the speaker bears the responsibility for his or her message. The speaker has all the encoding, i.e. linguistic means to hand.

The starting point is the speaker's choice of what he is going to convey, by means of linguistic encoding. It is important to establish a difference between what is conveyed by linguistic means and what is to be inferred extra-linguistically.

The speaker's choice constitutes the semantic contents of the message.
Before the encoding starts, the speaker has organize the contents in view of subsequent encoding. There are several decision he makes, concerning
- the degree of explicitness
- relations between chunks of sense
- relative prominence of particular chunks
- relation to previous discourse
- the information structure

The degree of explicitness covers the whole range of choices including encoding some chunk of sense by lexical or by grammatical means, if both are available. For example, in some languages the speaker may choose a temporal conjunction that explicitly marks temporal relations, for example anteriority,. e.g. Polish *jak tylko* 'as soon as' or a general temporal conjunction, e.g. Polish *kiedy* 'when'. Consequently his final text may read:

Jak tylko usłyszał tę historię, pomyślał że….
'as soon as he heard the story, he thought that…

Or

Kiedy usłyszał tę historię, pomyślał, że….
'when he heard the story, he thought that….

where the anteriority is marked by perfective aspects of both verbs.

Other choices include introducing a full-blown adverbial clause through a conjunction that marks it as temporal, causal, conditional, concessive or other or introducing the same contents by means of a gerund

Śmiał się, rzeźbiąc twarz.
'lit. He was laughing, sculpting the face'

Śmiał się, kiedy rzeźbił twarz.
'He was laughing while sculpting the face

 Codziennie je oglądał, nie mogąc się nacieszyć ich kształtem, ich ładnością.
'He watched them every day, unable to get over the joy their shape and beauty were giving him
Codziennie je oglądał, bo nie mógł nacieszyć ich kształtem, ich ładnością
'He watched them every day, because he was unable to get over the joy joy their shape and beauty were giving him

Nie będąc zdolny do naśladowania, byłem przecież zdolny do miłości.
'Unable to emulate, I was however able to love

Mimo że nie byłem zdolny do naśladowania, byłem przecież zdolny do miłości.
'Although I was unable to emulate, I was however able to love'

Przyjmując taki urząd, postąpiłby wbrew woli i rozkazowi Rzplitej.
'Accepting the post he would have acted/he would act against the will and the order of the Commonwealth'

Gdyby przyjął taki urząd, postąpiłby wbrew woli i rozkazowi Rzplitej
If he accepted/had accepted the post he would act/have acted against the will and the order of the Commonwealth'

In some cases the speaker may opt for nominalizing the relevant chunk instead of phrasing it as a gerund or clause,

Po usłyszeniu tej historii, pomyślałem…
Upon hearing the story, I thought…

Relations between chunks of sense may be framed alternatively, for example the speaker may opt for a compound sentence bearing a concessive clause or a coordinated sentence, with adversative clause, e.g.

Nie byłem zdolny do naśladowania, ale byłem przecież zdolny do miłości.
'I wasn't able to emulate, but I was however, able to love.

These choices are similar to what within the MTM is called conceptualization

Differences in relative prominence of sense chunks is best correspond to narrative foregrounding/backgrounding techniques, and includes encoding some parts of the message as parenthetical, e.g.as a nonrestrictive relative clause or other parenthetical structures.

Nevertheless main decisions of this kind affect the choice of lexical items among synonymous expressions existing within language, cf.

Ocalałem vs. *Uratowałem się*
(the example is drawn on different Polish version of Frodo's narrative in *The Lord of Rings*, the original sentence reads *I was saved.*

Lexical items are chosen for their sense but also for their semantic and syntactic valency, as they play important part of how the final encoded version will read. In our example it is illustrated by choosing either an intransitive verb or a transitive verb constructed as reflexive.

Here is where syntax intervenes

In our model, as in MTM, syntax is two-level: deep syntax and surface syntax. Our claim is that for our purposes deep syntax is best represented as dependency syntax, while the surface syntax is best represented by syntax of immediate constituents.

Some of the reasons for this are based on observations concerning order of syntactic elements when they are readjusted for reasons of information structure, cf.

Do Krakowa pojechał Jan (+)
Do Krakowa Jan pojechał (++)
 Do Krakowa pojechał Jan na koniu (++)
 Do Krakowa Jan pojechał na koniu (+)

The information structure

Information structure is, in my opinion, present at all the levels of language and in all the linguistic subsystems within which it is defined. It can also be interpreted on all those levels. In particular, it is very important to analyse information structure within the language system, i.e. *langue*, mainly within semantics and syntax. It is the language system that possesses all the means for arranging the information structure of an utterance.

Some of these means are non-specific. There is no single level on which information structure can be analysed, that could then be added to the phonological, morphological, syntactical and semantic level of the language. Quite the opposite, in many languages, the linguistic means belonging to various subsystems are employed in the arrangement of the information. Some languages, however, posses special grammatical markers of the information structure, e.g. the Japanese particle *–wa*, and its Korean counterpart *–ŭn-/nŭn.* Other languages employ specific syntactic means, for example cleft sentences (Bogusławski 1977: 221; Linde-Usiekniewicz 2004, 2006a, b).

Other means are: left and right dislocation, e.g.

A mi madre no le digas esto
'lit.: to my mother do not tell her that'//
No se lo digas a mi madre

'lit: do not tell her that to my mother'; and particles including Polish *to*, e.g.

Jan to ma szczęście
'John is one lucky bloke'
To Jan ma szczęście
'It's John who is lucky',

and Portuguese *é* (*que*), e.g.

Tu é que estas doido
'You must be really crazy'

Both the Polish sentence-internal *to* and Portuguese *é* (*que*) mark the rheme. Some languages, e.g. Chadic, have morphological markings for rhemes (Pawlak 1984).

It should be noted that languages may have some secondary means of marking the information structure. It has been proposed, both in Derwojedowa, Linde-Usiekniewicz (2004), and in Linde-Usiekniewicz (2006b) that variance in morphological agreement can be analysed as one of its markers. The first paper discusses Polish sentences with numerals, e.g.

Pięć dziewcząt było wesołych/wesołe,
'Five girls were merry'

the second mentions Spanish variance in agreement in topicalizing clefts, e.g.

Quien llevo (sg) // *Quienes llevaron* (pl) *al niño fueron los abuelos*
'Who took the child were the grandparents'.

Consequently it is no longer reasonable to relate the information structure to the actual sentence division, as was originally proposed (Mathesius 1939).

Some attention has been paid to the notion of the emphatic (or marked), i.e. contrastive rheme, less so to the emphatic (or marked), i.e. contrastive theme. Opinions differ as to whether all topicalizations should be considered emphatic, or not.

As mentioned before, in the majority of the theories, themes have to correspond to nominals. However, some data seems to indicate that some languages, but not all, allow for non-nominal themes. One of them seems to be Polish (Huszcza 2003), which allows for verbal themes, and themes including complex verbal groups, e.g.

Złoty medal zdobył Carl Lewis
 'The gold medal winner is Carl Lewis;
lit. the gold medal won Carl Lewis (sic!)'.
The gloss proves that English is not this kind of language.

 Moreover, the so-called topicalizing clefts, e.g.

What John was doing was waiting on the station,

seem to prove that an entire clause can be a complex theme.

I have proposed elsewhere (Linde-Usiekniewicz 2007) that the notion of marked (emphatic) theme-rheme division should be added to the notions of an emphatic theme and an emphatic rheme (Huszcza 2000, 2003). Such a division is indicated by a pause in speech, while in writing it can be marked by word order change, punctuation, and, both in speech and writing, by cleft-sentences, if available. Its function is to prevent a possible neutralization of the division.

Furthermore, it seems that languages possess procedures for the reverse process, i.e. facilitating the thetic reading of a sentence. Bartschev, Partee (2002) argue that the Russian genitive of negation, e.g.

Otveta iz polka nie prišlo, vs. *Otvet iz polka nie prišel*
'The answer from the regiment didn't come'
(the authors' original example)

serves this purpose.

Huszcza (1990) and Ishigawa (2007) independently show that in Japanese this procedure is carried out by substituting the subject particle *–ga* for the thematic particle *–wa*. It is possible that in Polish, whenever a verb can be used both with a nominative subject, *e.g.*

Woda nakapała na parapet 'The water dripped on the window sill'

and with a genitive, e.g.

Nakapało wody na parapet 'There is some water on the window sill'

the choice between the two renderings depends on whether it is the speaker's intention to make their sentence a thetic one or not.

Other languages possess similar means, e.g.

English *There is a policeman in the kitchen*,
French *C'est que j'ai parlé à Pierre*
        'But I did speak to Pierre' (Tollis 2006)
        Il y a un téléphone qui sonne
        'There is a telephone ringing', (Axel Holvoet, pc).
 Spanish, *Hay agua hirviendo*
                'The water is boiling, lit. There is water boiling'.

All in all, the controversies concern the non-nominal themes, theme-less sentences and the semantic vs. syntactic vs. pragmatic nature of the information structure. It seems that they are geared by a very fundamental difference in perspective adopted by each linguist for their analyses. Classical approaches, and also the approach proposed by Huszcza (1980, 1990, 2000, 2003), have followed, implicitly or explicitly, the speaker's perspective. As can be seen in the definition quoted above, the speaker intends the audience to do something with the message. Other approaches seem to follow the audience's perspective. The advantage of this perspective is that the audience's perspective normally mirrors the way a linguist looks at the linguistic data.

The two perspectives and the resulting statements about the information structure of the utterance are quite distinct because in a deictic situation each participant has his own task to accomplish. The audience's task is to decode and interpret the message. It is, therefore, natural for them to think in terms of 'what the message is about'. The way this question is answered influences the way the audience construes the theme of the utterance. And the very form of the question favours nominals as themes. By contrast, the speaker assumes a different role: he remains in charge of the communication process and does his best, using all the linguistic means available, to signal to the audience how the message should be processed.

The two perspectives should be separated, not in terms of theoretical and methodological independence, but as two distinct stages of communication. The speaker is trying to do something with the message he is conveying, but his responsibility ends the moment he closes his mouth. The audience is omnipotent and can process the information any way it chooses. In particular, it can decide, contrary to the speaker's intention, that the message is in fact about something. In the much discussed example *Woda zalewa mieszkanie!* the speaker may be indicating that he does not wish his message to be treated as having the theme 'water' or the theme 'flat' and the audience may decide that the message carries the information that 'the flat is flooded'. The audience arranges the information any way it wishes, even treating the contents of the message in a way that was not planned by the speaker. In fact, it may act in the way described in *Foucault's Pendulum,* (when one of the protagonists decides to become "a kind of private eye of learning" (p. 189):

> Still, I was accumulating experience and information, and I never threw anything away. I kept files on everything. I didn't think to use a computer […]. Instead, I had cross-referenced index cards. Nebulae, Laplace; Laplace, Kant; Kant, Königsberg, the seven bridges of Königsberg, theorems of topology… It was a little like that game where you have to go from sausage to Plato in five steps, by association of ideas. Let's see: sausage, pig's bristle, paintbrush, Mannerism, Idea, Plato. Easy. Even the sloppiest manuscript would bring twenty new cards for my hoard. (p. 199).

(The quotes and the page references come from the English version, translated from the Italian by William Weaver, and published by Ballantine Books, New York, 1990)

The quote is not accidental. There is nothing to prevent the audience from treating every piece of information present in the message as a source of file-cards for their mental catalogue and – should there be no drawer for some esoteric piece of information – from creating a new one.

The distinction between the speaker's and the audience's perspective allows for the temporary character of the theme-rheme division for the purpose of a particular exchange. As the discourse develops, the thematic or rhematic character of the items may shift – as described in Daneš (1974).

Nevertheless, since it is the speaker who is in charge of all the linguistics means, it follows that the distinction between theme and rheme has to be described from the speaker's and not the audience's perspective. Since this involves the speaker's intention, it should be couched in pragmatic terms. An ideal framework for such a distinction is provided by the Relevance Theory of Sperber and Wilson (Sperber, Wilson 1986). Roughly speaking, what is proposed here follows the premise that each speaker's utterance changes the shared cognitive environment of the speaker and the audience because it alters the audience's cognitive environment. However, I would like to propose two types of changes that the speaker may intend to produce. He may want to introduce some quantitative changes in the shared environment, or some qualitative changes. A quantitative change would consist in introducing a new fact or making a new fact manifest, to use the authors' terminology. A qualitative change would, by contrast, consist in altering the relative prominence of the facts already present and accessible in the shared cognitive environment, by bringing some fact to the fore.

If the speaker marks any part of his utterance as thematic, it means that he intends to change the shared cognitive environment qualitatively only. That is why, in many cases, the thematic part of the utterance corresponds to what used to be called "old information". However, the assumptions the speaker has about the shared cognitive environment may be wrong, and what is intended as a qualitative change may become a quantitative one. To further elaborate the quote from *Foucault's Pendulum*: it is perfectly safe to assume that the readers of the present text have heard of Nebulae, Laplace, Kant, Königsberg, and topology. Therefore, the reference to them in the quote above has just brought them to the fore of their minds. However, if a reader has never heard of the seven bridges of Königsberg, mentioning them changes his cognitive environment in a qualitative way. The reverse is also true: A professor of linguistics, teaching about information structure and using the *Columbus discovered America* example to show the distinction between theme and rheme, and pronouncing the sentence in such a way that *discovered America* constitutes the rhematic part of the sentence, does not assume that his students have never heard about the discovery of America. Therefore, what would be a quantitative change in a normal discourse (let's say a history class in elementary school) becomes a qualitative change in a university lecture hall.

The terminological switch from theme and rheme to thematic and rhematic part of the utterance in the previous paragraph is not accidental. It allows for all kinds of syntactic structures, including verbs, verb phrases and clauses in the thematic position. Moreover, it also provides an explanation as to why the information structure is so similar to the distinction between what is presupposed and what is posed and why it affects that distinction. Namely, the elements of meaning of lexical items that correspond to what is presupposed should alter the cognitive environment of the audience in a qualitative way. By contrast, what is posed should alter it in a quantitative way. The main difference between the traditionally conceived information structure and the presupposed vs. posed distinction consists in the speaker having his say (pun intended) in what constitutes the thematic part of his utterance and the intended qualitative change in the former, and in the case of the latter remains at the mercy of the lexical resources the language places at his disposal.

Constituency vs. Information structure

Information structure is binary, i.e. utterances are divided into *theme* and *rheme*, The distinction I posit is not between an element marked for some purpose and the unmarked remainder but between the two counterbalanced, equally important parts of the utterance and the information it contains. The scope of the theme and the rheme of an utterance are marked by a TR boundary.

Themes and rhemes can be further divided into secondary themes and rhemes , by virtue of a secondary TR boundary. This division can be even more granulated, however, there might be pragmatic limitation of how far one can go. Lower order theme or rheme may be so marked for many reasons, including contrast. Nevertheless, assigning theme of rheme values to sentence parts has to conform to constituency relations. If two utterance elements do not form a constituent, they cannot constitute a single theme or a single rheme. For example internal and external arguments of a verb do not form a single topic. This conundrum is also solved by marking one of them, as an upper order theme/rheme and the other as a lower order theme/rheme, not necessarily contrastive.

Speakers may refrain from dividing their utterances into topic and focus, and the resulting utterances are considered thetic. They may also decide not to mark the division explicitly, if the division is otherwise retrievable or if such marking interferes with whatever effect, be it stylistic, semantic, pragmatic or other, they want to achieve.

Some languages have special resources, morphological or syntactical, to mark themes. rhemes or the division itself. These include topic particles in Japanese and Korean, focalizing particles in Chadic and in Gungbe (Aboh 2007), and give rise to distinguishing so-called Subject-driven languages from Topic-driven and Focus-driven. Syntactical resources include, among others, focalizing cleft-sentences and, less often, topicalizing (pseudo)cleft sentences); and left-dislocation Other languages have means of obviating the division.

The most universal resource for marking themes and rhemes is intonation, as a stand-alone or in combination with other resources. Nevertheless languages vary strongly in actual phonological patterns of intonation markings and in what can be marked by intonation. For example, French does not allow for initial intonation marked rhemes

Some languages apparently allow for marking TR division through changes in word order. In oral language this is accompanied by appropriate intonation pattern. In written language word-order changes alone stand for the division and the reader is left to reconstruct the possible pattern. Word-order changes alone (without intonation) rarely serve to distinguish theme and rheme. By contrast, they tend to mark the TR division itself. However, they may also serve to obviate the division obligatory by virtue of constituency requirement.

Jan dał książkę Dorocie
'John gave the book to Dorota'

Jan dał Dorocie książkę
John gave Dorota the book

Languages vary as to the degree to which they accept TR divisions playing havoc with the linear structure of constituents. This translates into another typological distinction, that into constituency-driven languages and TR-boundary driven languages. Polish appears to be fairly strongly TR-boundary driven, which may account for the fact that it allows for deep left branch extraction.

Jaką Zadie Smith książkę czytasz?
'lit. What of Zadie Smith book are you reading?'
Jaką czytasz książkę Zadie Smith?
Lit. What are you reading book by Zadie Smith?'
Lwa czytam Tołstoja, nie Alekseja
'I am reading a book by LEW Tolstoj, not by Alexy Tołstoj

Moreover, this is not limited to nominal structures, cf.

 Jak to było wczesnie?
'How it was early?

Other phenomena that could be or are described in similar framework:

- Full lexical verbs vs. support verbs with nominals

  Podjął jedyną możliwą w takich wypadkach decyzję, że ....
  'He made the only possible decision in these circumstances, that…
  [*Decyzję, że druk zostaje wstrzymany,*]$_T$ [*podjął dyrektor.*] $_R$
  'Lit. The decision to stop printing made the manager'
  Że druk zostaje wstrzymany, zdecydował dyrektor
  'Lit. To stop printing decided the manager'

- Ordinary and marked order of adjectives in Polish and Romance and adjectives as nominal adjuncts vs. predicatives in relative clauses

  To jest parking dla CZYSTYCH samochodów.
  To jest parking dla samochodów CZYSTYCH

  This is a parking place for cars that are CLEAN
  kolorowa drukarka atramentowa vs. *drukarka atramentowa kolorowa*
  ' color ink-jet printer'

- Possessives vs. Indirect objects
  Cf. Polish *Połamią mu nogi* vs. *They will break his legs // ?They will break his legs for him*

- Elective constructions
- Superlative constructions
- Metatextual, discourse, evidential etc. markers
- Parentheses

Literature

Bogusławski A., 1977, *Problems of the Thematic-Rhematic Structure of Sentences*, Warszawa: PWN.

Borschev V., B. H. Partee, 2002, "The Russian genitive of negation: Theme-rheme structure or perspective structure?", *Journal of Slavic Linguistics* 10, pp.105-144.

Derwojedowa M., J. Linde Usiekniewicz, 2004 "O pewnych własnościach zdań z przymiotnikowym orzeczeniem imiennym i grupą liczebnikową w podmiocie", *Biuletyn Polskiego Towarzystwa Językoznawczego* LIX, Kraków, pp. 36-45.

Derwojedowa, M., Linde-Usiekniewicz, J. (2007), "O konstrukcjach elektywnych – inaczej", en A. Mikołajczuk, R. Pawelec (eds.) *Na językoznawczych ścieżkach,* Warszawa: Semper, 271-277.

Huszcza R., 1980, "Tematyczno-rematyczna struktura zdania w języku polskim", *Polonica* VI, pp. 57-71.

Huszcza R., 1990, "Czy istnieją zdania czysto rematyczne?", in T. Dobrzyńska (red.) *Tekst w kontekście,* Wrocław: Ossolineum, pp. 97-103.

Huszcza R, 2000, "Nie ma, żeby nie było - o segmentalnych wykładnikach tematyczno-rematycznej struktury zdania w polszczyźnie", *Poradnik Językowy*, z. 8, pp. 1-8.

Huszcza R, 2003, "Wykładniki tematyczno-rematycznej struktury zdania w polszczyźnie – próba typologii", in R. Huszcza i J. Linde-Usiekniewicz (red.) *Prace językoznawcze dedykowane Profesor Jadwidze Sambor*, Warszawa, pp. 84-95.

Ishikawa K., 2007, "Overriding effects of focus on weak NPs in thetic sentences", *CLS 40 proceedings*, in print, draft version accessed at www.kisc.meiji.ac.jp/~kuni5963/papers/Ishikawa-CLS40-Procdngs2005.pdf, in December 2006.

Linde-Usiekniewicz J., 2004, "El verbo en las oraciones escindidas: implicationes teóricas", *Studia romanica posnaniensia* XXXI, Poznań, 2004, pp. 321-328.

Linde-Usiekniewicz J., 2006a, "Las oraciones escindidas: ¿qué es lo que escinden?", in. J. Linde-Usiekniewicz, A. Henriquez Vicentefranqueira (coords) *Estudios en lingüística ibérica e iberoamericana,* Warszawa, pp. 31-40.

Linde-Usiekniewicz J., 2006b, "On the relative depth of cleaving", in. H. and A. Włodarczyk, *La focaliation dans les langues*, Paris, L'Harmattan, pp. 81-94.

Linde-Usiekniewicz J., 2008a, "O pewnych interesujących własnościach wybranych analitycznych konstrukcji leksykalnych," in. R. Grzegorczykowa, K. Waszakowa (ed.). *Studia z semantyki leksykalnej. Słowo – pojęcie – tekst,* Warszawa, 2008, WUW, 255-264

Linde-Usiekniewicz, J. 2008b "On some controversial issues in the description of information structure in language and utterances". Etudes Cognitives 8, Warszawa, 2008, SOW, 129-138

Mathesius V., 1939, "O tak zvaném aktuálnim členění větném", *Slovo a slovestnost* 5, pp. 171-174.

Mel'čuk I.A., 1988, *Dependency Syntax: Theory and Practice*, New York: Sunny Press.

Pawlak N., 1994, *Syntactic Markers in Chadic*, Warszawa: Wydawnictwa UW.

Sperber D., D. Wilson, 1986, *Relevance: Communication and Cognition*, Oxford: Oxford University Press.

Tollis, F. 2006, "Les «constructions expressives » en *C'est* dans les écrits publiés de Gustave Guillame" in. H. and A. Włodarczyk (eds.) *La focalisation dans les langues*, Paris, L'Harmattan, pp. 95-109.