Diachronic Studies on
Information Structure

Language Acquisition and Change

Edited by
Gisella Ferraresi and Rosemarie Lühr

De Gruyter
The Role of Information Structure in Language Change: Introductory Remarks

Gisella Ferraresi (Frankfurt /M.) & Rosemarie Lühr (Jena)

The renewed interest in the interaction between ‘information structure’ and grammar is witnessed by many recent publications which try to disentangle the correlation between syntax, prosody and information structure in modern languages. The aim of this volume is to encourage the discussion of the role of information structure in language change and that of operatinal methods which can be applied on corpus languages when working on information structure. Correlations between word order and intonational patterns which, through the question test, enable us to detect the informativity of a sentence (cf. Sgall et al. 1973), for example, are not applicable to corpus languages.

Together with discourse semantics, prosody and syntax are the important ingredients for the production and interpretation of information packaging. Different languages, however, make different use of these components. German, for example, operates in the first place at the prosodic level\(^2\), although it can also make use of marked syntactic constructions like hanging topics or left dislocation (Frey 2005). In Catalan, on the other hand, syntactic constituents are moved in order to constitute the corresponding prosodic constituents (Vallduvi & Vilkuna 1998). Other languages again adopt morphological marking.\(^3\)

In historical linguistics, the prosodic correlate of syntactic constructions cannot be tested directly. However, it is still possible to reach some relevant generalizations about the realization of pragmatic categories like Topic, Focus and their interaction with word order in written texts. This is facilitated also thanks to the intensive research in the last years of the mapping of pragmatic categories in syntax – an interest which has as-

---

1 Some of the papers contained in this volume have been read at the workshop “The role of information structure on language change” during the annual DGS Conference held at the University of Siegen (Germany) in 2007.

2 For a discussion on the prosodic correlates of Topic and Focus in German see Féry (1993); Uhmann (1991)

sumed an important role in formal syntactic theories as well. Chomsky (1995), for example, recognizes that the interface between syntax and discourse is realized in the CP. Likewise coached in generative grammar, the so-called cartographic model (Rizzi 1997 a.o.) holds Topic and Focus to be realized in precise syntactic positions in a split CP. In Rizzi (1997) the CP is therefore further distinguished in

(1) \[
\text{Force TOP FOC FinP}
\]

From a diachronic point of view, many structural changes which involve the left periphery can in this way be put down to changes in the mapping of pragmatic information onto syntax. For the ancient Indo-European languages, Kiparsky (1995) assumes a structural differentiation between Topic and Focus. His proposal is based on Hale’s assumption that topic and wh-elements move to different positions in these languages (Kiparsky 1995: 253):

(2) \[
S'' = \text{CP}
\]
\[
\text{TOPIC } S'
\]
\[
\text{XP}_i
\]
\[
\text{FOCUS } S
\]
\[
\text{XP}_v
\]
\[
t_i
\]
\[
t_i
\]

In ancient Indo-European languages verb fronting is a strategy of emphasis, whereas clause typing is achieved by particles and sentence prosody. Therefore verb-second phenomena are the rule neither in old Germanic nor in old Romance languages. Gothic, for example, shows a rich array of particles – albeit more limited than other ancient Indo-European languages like Ancient Greek –, many of which induce fronting of the finite verb depending on the clause type on the one hand and on the informational status of the moved element on the other (cf. Ferraresi 2005). Verb movement is found regularly only in imperative clauses. In Old High German and Old English, some of the corresponding particles are still attested (Axel 2007; Fuß & Trips in print; Kemenade & Los 2009). However, verb movement in Old High German declarative sentences is more and more generalized, for example in verb-first structures (cf. Hinterhölzl, Petrova & Solf 2005), especially in so-called thetic sentences. There, it has the function of introducing a new discourse referent while changing the discourse situation, nevertheless maintaining the narration line. Verb-

\[\text{See a.o. Ferraresi & Goldbach (2002) for Old French, Axel (2007) for Old High German.}\]
second word order, on the other hand, has the function of introducing a
new referent, which predominantly appears in the first position. In
declarative main clauses of Modern German, this word order has been gen-
eralized and has become the unmarked one after different discourse par-
ticles have got lost. In this sense, the question arises if there really is a transition
from discourse-oriented to syntax-oriented languages and, if so, if this transition always takes place in that direction. According to Givón (1979:
98), the word order in the “pragmatic mode” is mostly governed by prag-
matic principles, while the word order in the “syntactic mode” displays
“case functions”. The degree of accessibility or activation of a discourse
referred to a certain phrase, e.g. the correlation with the grammatical func-
tion (obliqueness hierarchy: subject > object1 > object2 > ... > adjuncts),
actuality, definiteness (cf. Givon 1983; Gundel, Hedberg & Zacharski
1993) is, however, influenced by many different factors. Semantic salience
is in any case relevant for the analysis of information structure in ancient
Indo-European languages, as it allows for a differentiation between dis-
course-oriented and syntax-oriented languages (Givón 1979). Personal
pronouns, for example, often appear in the topic position in Old Indic,
even though ancient Indo-European languages are subject-pro-drop and
partially object-pro-drop languages (Luraghi 2003; Lühr 2005; Keydana
2009). Functional tests, especially those operating on the level of text de-
velopment, in this way allow hypotheses on the informational status of a
linguistic unit since they can be supported by semantic determination.
The investigation of corpus languages can also cast light on synchronic
issues in modern languages. There is some significance, for example, in the
fact that Ancient Indic marks contrastive focus specifically (Lühr 2009),
supporting the differentiating analysis between New Information and
Contrastive Focus against the unified semantic interpretation of focus
(Rooth 1992; Krifka 1993; 2007), according to which focus always evokes
a set of alternatives. This differentiation is necessary as the intonations
of these two foci differ from each another also in modern languages like in
German (Kiss 1998; Steube 2002; Späth 2007). In the data material of
corpus languages it is thus possible to define which type of focus is real-
ized5. According to Umbach (2004), it is unclear how the alternatives are
contextually restricted in case of a purely semantically determined focus.
Consequently, it has to be assumed that every sentence adds information
to the preceding one; this is why continuation sentences are subdivided
into established information and new information. The established in-
formation is usually expressed via definite nominal or prepositional groups
(with definite articles in German and Greek), possessive pronouns or per-

5 For a good overview of the different categories cf. Féry & Krifka (2008).
sonal pronouns or via anaphorical, often deictic expressions. Among these, the Aboutness Topic is typically a referential DP. Still, the topic may be contrast-focussed and thus be a Contrastive Topic. As compared to the topic, the focus is related to a specific domain, the focus domain. As long as no contrast foci are present, the focus domain includes new information. Under certain circumstances, the background needs to be added; it usually consists of the established information outside the focus domain. To determine the topic, the approach in Frascarelli & Hinterhölzl (2007) is promising, as Aboutness Topics, Contrastive Topics and Familiar Topics are dealt with separately, which are all operationalizable pragmatical categories that can be found in non-Indo-European languages as well (Ermisch 2007).

When questions are answered in the text itself in corpus languages, a solicited New Information Focus is involved. To determine the unsolicited New Information Focus (new discourse referents as well as new relations between given discourse referents) in narrative texts, a question can be formed that is as universal as possible and that is based on established material (Petrova & Solf 2009 in on the quaestio-thesis by Klein & von Stutterheim (1987). Still, when studying ancient Indo-European languages, this procedure can only be a possible test in addition to semantic analyses, e.g. in connection to Dik’s (1989: 268) strategies for the introduction of a new discourse referent:

1. meta-linguistic information: ‘I’m going to tell you a story about X’
2. as an object or second argument in sentences with transitive verbs: ‘In ..., we saw X’
3. as a subject of “an existential or presentation construction” of the type: ‘Once upon a time, there lived/was ...’; such constructions are typical of sentences initializing a text
4. with verbs of movement indicating an “appearing on the scene”: ‘Suddenly, right before our eyes, X appeared ...’ (Petrova & Solf 2009).

To be able to comment, in a second step, on the part of the linguistic system of an ancient Indo-European language which concerns the information structure, the analysis units of information structure first have to be synchronically determined and to be assigned to the language material. When one has undertaken several synchronic sections, diachronic processes become evident and explicable as well. Moreover, comparing different languages makes language-specific and general traits of information structure in ancient Indo-European languages become apparent. As mentioned above, however, the study of ancient Indo-European languages can support the analysis of modern ones.
These are only some of the problems which one is exposed to when working on pragmatic categories in corpus languages, and which the contributions of this volume partly address. The papers analyse different phenomena of language change connected to information structure in many different old languages and language families, applying some of the techniques described above for corpus languages. The languages analyzed in the papers belong to the Uto-Aztecan (Armendáriz), Vedic (Viti), Slavic (Matić), Romance (Gabriel & Rinke, Eide) and Germanic language families (Wratil, Westergaard, Petrova & Hinterhölzl). The phenomena considered range from word order to pronouns and verb movement.

The paper ‘Information structure, constituent order, and case marking in Waríhi’ by Rolando Félix Armendáriz discusses word order in Waríhi, an Uto-Aztecan language spoken in Mexico. The methods applied for this study are both text analysis and direct elicitation. Armendáriz shows that syntactic accounts are not sufficient for understanding the principles underlying the various patterns of word order in Waríhi. He therefore proposes a pragmatic explanation based on Lambrecht (1994). Accordingly, focus determines the fronting of constituents, while topical material usually follows. The analysis is mainly conducted on a synchronic level, as Waríhi does not possess a long written tradition that permits a reconstruction of previous linguistic stages. However, Armendáriz indicates that some diachronic evolution is observable in the use of a suffix, which is shifting from a syntactic to a pragmatic function of signalling definiteness.

Carlotta Viti’s paper ‘The information structure of the OVS order in Vedic’ discusses OVS word order in Vedic, which is marginally attested in the Rig-Veda, the earliest Vedic text, and completely unattested in the subsequent Indian linguistic varieties. Owing to this fact, OVS has been considered as being an exception to the basic SOV word order, related to the poetic register of the Rig-Veda. However, data point out that OVS is consistently used to convey a certain type of information structure, where the referent of the object is more topical (i.e. human, specific, definite, persistent, etc.) than the referent of the subject. Accordingly, the synchronic rarity and the diachronic disappearance of the OVS order are due to the fact that its information structure has fewer chances to appear in discourse.

In their article ‘Information packaging and the rise of clitic-doubling in the history of Spanish’, Christoph Gabriel & Esther Rinke investigate the diachronic development of clitic-doubling (CD) in Peninsular Spanish. They show how this construction differs from structures with right-dislocated objects (clitic right dislocation; CLRD). Besides prosodic and syntactic differences, both structures differ in their informational status. In CLRD constructions, a right-dislocated object is separated from the core
sentence by a prosodic boundary being located in a position outside rP, whereas in CD constructions a nominal or strong pronominal object forms, together with a co-referential clitic pronoun, one prosodic domain – the doubled object remains in its rP-internal base position. Here, object clitics function as agreement markers. As for information structure, the main difference between CLRD and CD constructions is that in a CLRD structure the object constitutes a dislocated topic, whereas in CD constructions it belongs to the focus domain. In modern Spanish, elements displaying a high degree of ‘topicality’, such as pronouns, definite noun phrases, experiencer arguments and proper names, are more likely doubled than those situated lower in the topicality hierarchy. This seems to confirm that CD constructions follow the topic hierarchy given by Givon (1976). As a whole, contexts with obligatory CD are quite rare and the appearance of the clitic is due to optionality as it is in modern Peninsular Spanish. In the early periods, in the 12th and 13th centuries, resumptive clitics appear in texts that reflect spoken language. However, such examples are far from being systematic or obligatory. Resumptive clitics become more and more frequent, especially during the 16th century. Gabriel & Rinke agree with Fontana (1993) that CD should be analyzed as an agreement phenomenon. They provide evidence from texts from the 15th up to the 18th century for the assumption that there was a diachronic tendency from more topical to less topical elements to co-occur with a resumptive clitic, whereby a topicalization strategy (clitic right dislocation) was reanalyzed into a doubling structure. The reason behind this development is that right dislocation is more marked and less economic than CD.

The paper ‘Cue-based acquisition and information structure drift in diachronic language development’ by Marit Westergaard evaluates information structure from another point of view. Within a model of language acquisition and change which recognizes the importance of cues and so-called ‘micro-cues’ in the primary linguistic data (PLD) that children are exposed to, she shows that information structure may be a factor for word-order change. Following Lightfoot (1999, 2006) she assumes that there is a UG requirement for cues to be obligatory in children’s I(ternalized)-language grammars. Westergaard analyses word-order variation that is dependent on linguistically relevant sub-categories as pertaining to the I-language. The data from English and Scandinavian languages present mixed grammars which allow two different subject or object positions. One of the positions is preferred for discourse-given elements (typically pronouns) and the other for informationally new or focused elements (typically heavier elements such as full DPs or clauses). The paper shows that word order variation is relatively common in present-day grammars,
whereby the choice of a word-order is at least partly dependent on information structure. In accordance with Behaghel, Westergaard demonstrates that discourse-given elements, typically pronouns, appear in a high position, whereas informationally new or focused elements stay in a lower position. Historical data, too, attest that word order variation based on mixed VO/OV is caused by information structure. With respect to child language, Westergaard demonstrates on the basis of Norwegian and Russian data that children have an early sensitivity to patterns of information structure, producing both word orders in a more or less target-consistent way in e.g. Norwegian wh-question and subject-shift constructions. In the second part of her paper, Westergaard investigates the realization of subjects and objects in a sample of Norwegian conversational speech and in a sample of Norwegian and English child-directed speech. The results show a clear preference of putting discourse-given information in subject position and discourse-new or focused information in object position. Thus, for mixed word orders where the choice depends on information structure, word orders linked to informationally given (typically pronominal) subjects should be naturally frequent in the E-language. Word orders linked to new or focused (full DP) subjects should be correspondingly infrequent. For mixed grammars involving two object positions, the situation should be the other way round. The mixed word order system should be easily learnable, and from an acquisition perspective such grammars should also be relatively stable diachronically. However, in this scenario the input varies, since not all children are exposed to exactly the same linguistic data. The low frequency of a construction may fall below the threshold for acquisition in some children, who will develop a grammar with only one subject or object position. The decline in the frequency of word order would be the crucial factor for change. Certain developments can be reversed if external factors and/or internal language development cause the PLD to change as in some present-day Norwegian dialects, where the interference from the standard language seems to cause a return from non-V2 to V2 word order.

In his contribution ‘Discourse and syntax: the decline of postverbal topics in Serbo-Croat’, Dejan Matić deals with a word order change that cannot be attributed to syntactic processes and is therefore impossible to phrase in terms of competing grammatical choices. Postverbal topical subjects (PTS), i.e. subjects with topical interpretation placed to the right of the verb, have been attested throughout the history of Serbo-Croat (SC) and seem to be as frequent and of the same range of functions until the 19th century. These constructions are a) inversion, a verb-second-like construction in which subjects appear postverbally after a fronted wh-word, relative pronoun, quote, and a fronted focus, b) vS, a construction
with a focused postverbal subject, c) V±X, a construction with a topical postverbal subject. They differ in intonation contour, position of sentence adverbials, position of the subject relative to the copula/auxiliary and function as continuous topics. During the 20th century, the frequency and variability of PTSs rapidly declined, even though speakers are still ready to construe all types of clauses with PTSs like the ones attested in the sources from the 19th century. The reasons for this language change are a number of interrelated sociolinguistic factors, initiated by pragmatic changes. In the 20th century, sixty to seventy per cent of all instances of PTSs are more or less petrified verb-subject collocations, which are regularly used in a particular discourse function. Marić suggests to introduce a process of automatization into the repertoire of diachronic changes, a kind of binding of syntactic structure to certain lexemes. The only change between the 19 and the 20th centuries that seems to have led to something resembling a categorical difference is the change in the function of V±X clauses with PTSs, which are no longer markers of resultative and consecutive events in narratives. The syntax remains unchanged, only discourse-organization principles change.

Kristine Eide, in her paper ‘Prosody, information structure and word order change in Portuguese’, considers the word-order changes which have taken place in the period from Classical Portuguese to Modern European Portuguese, which – according to Eide – have their reason in a prosodically driven change of discourse patterns. This change caused a fixation of the subject in the preverbal position, which then developed from a topic position to a more neutral one. Classical Portuguese is a language with topic-verb-XP structure, in which inversion takes place in case the first position is not occupied by the subject. Eide observes a decrease in the postverbal position of subjects of unaccusative verbs which represent old information. Topics in Classical Portuguese always move to the beginning of the sentence, whereas the rest of the background follows the finite verb. This creates many inversion structures, in case the topic is not a subject. In Modern European Portuguese only new information can occur postverbally.

Melani Wratil’s paper ‘The development of imperative V-to-C movement in the West Germanic and Romance Languages’ contributes to the discussion about the movement of the imperative into the C-position. In the Indo-European languages the imperative as a directive speech act constitutes a genuine directive operator represented by a zero-expression. With respect to head movement, imperative and indicative behave in the same way. In VSO languages they occupy the regular clause-initial position, in SVO languages they are positioned immediately after the subject, and in SOV languages they appear in sentence-final position. But in the
West Germanic SOV languages, as for example in German, as well as in the West Germanic SVO languages, as for example in Yiddish, imperative verbs normally precede their complements and adjuncts. Here, the exceptional structure of the West Germanic and Romance imperatives is a result of the emergence of V2 phenomena. Root clauses are reanalyzed as V2 structures as soon as the finite auxiliary or lexical verb follows an initial focalized or topicalized constituent in the second position, whereby the fronted focus phrases are considered as the scope-marking specifier of a new functional projection CP. The V2 properties of the modern West Germanic and Romance languages are strongly connected with the rise of V1 clauses. V1 structures are the structural outcome of a focalization process. Head movement with imperative verbs had been solidified by analogy with the finite verbs of V1 interrogative and declarative clause before V2 structures became basic structures. Wratil shows why especially the V1 structures of imperative clauses and therefore the imperative verb movement to C has been able to survive in the West Germanic and Romance languages up to this day. In modern Standard English, in all non-negated and non-emphasized imperative clauses V-to-C movement applies at LF, whereas in negative and emphasized affirmative imperative clauses do moves to C before Spell out.

In their paper ‘Evidence for two types of focus positions in Old High German’, Svetlana Petrova & Roland Hinterhölzl – using data from the Old High German Tatian translation – propose a basic VO order and an informational account in explaining word-order variation in subordinate clauses in Old High German. This view is in contrast with two other positions found in the theoretical discussion on basic word order in (old) Germanic languages. The first one is based on a basic OV order with different possible extraposition movement operations, the other one assumes two underlying base orders, which are due to different grammars (the so-called ‘Double Base Hypothesis’). Petrova & Hinterhölzl’s argumentation is supported by a battery of syntactic tests which show a contradictory picture for both hypotheses. These tests include the analysis of materials following the finite verb in subordinate clauses, such as DPs with different functions, PPs, predicative adjectives or nouns, Verb Raising and Verb Projection Raising. Root patterns (Verb-first and Verb-second) in subordinate clauses do not confirm the hypothesis of root patterns, either. The information-structural approach proposed in this paper considers the distribution of background material in comparison to asserted information and narrow focus patterns. Background materials tend to occupy the Wackernagel position, which in subordinate clauses is the position immediately following the subordinating conjunction. In most of the cases, pronouns in the corpus data are positioned in the Wackernagel position.
also against the Latin original. A similar observation is made for accessible
or discourse-anaphoric full lexical phrases: they are moved from a postver-
bal position to the position after C'. In postverbal position of subordinate
clauses, arguments of the verb and non-finite parts of complex verbal predi-
cates are considered. Arguments in postverbal positions represent new
information, non-finite parts of complex verbal predicates constitute the
asserted part of the proposition; in particular it is shown in the paper that
contrastive focus regularly appears left-adjacent to the finite verb. Petrova
& Hinterhöhl propose a structure where the verb moves to the head of a
Focus phrase at the left edge of the middle field. Whereas contrastive ele-
ments move to SpecFoc, new objects remain in the scope of the focus
head.

Last but not least, we would like to thank all our authors, who cooper-
ated with us with great patience. We also thank the series editors, above all
Anita Steube. Many thanks also to the person in charge of linguistics,
Heiko Hartmann, as well as to Andreas Brandmair at de Gruyter’s, who
assisted us in all steps of the publication process.

References

Cambridge University Press.
Ferraresi, G. & M. Goldbach (2002): V2 Syntax and Topicalization in Old


Kiss, K.É. (1998): Identificational focus versus information focus. Language 74.2, 245-273


