INTRODUCTION

When data challenges theory

The analysis of information structure and its paradoxes*

Davide Garassino[1,2] and Daniel Jacob[3]
[1]University of Zurich / [2]ZHAW, Winterthur / [3]University of Freiburg

This article provides a comprehensive survey on current research on information structure so as to clarify some 'paradoxical' effects stemming from the tension between data and theory. Paradoxes are here defined as unexpected data in light of certain assumptions held in mainstream literature. More specifically, we explore two possible sources of paradoxes: certain features of the experimental design and, above all, inadequate discourse models. Mainly considering dislocation and cleft sentences in French and Italian we suggest that some apparent paradoxes (such as non-focalizing clefts or dislocation expressing focus-related functions) can be conceived of as the effect of general pragmatic mechanisms and rhetorical strategies exploited by speakers. We also claim that these effects can be better understood through explicit models of discourse.

Keywords: information structure, non-canonical syntax, prosody, discourse, inferential pragmatics

1. Introduction

Information structure (henceforth IS) concerns the communicative functions of sentences or, more specifically, the linguistic strategies used for sharing and managing information flow. IS has been usually conceived as "information packaging" (Chafe 1976, 1994). As felicitously – and less metaphorically – expressed by Lambrecht, "the study of information structure focuses on the interaction of sentences and their contexts" (Lambrecht 1994: 9). As such, IS can be described,

* This chapter has been jointly conceived and approved by both authors. For academic purposes, Davide Garassino bears responsibility for §§ 1, 2.3, 3 and 4.3; Daniel Jacob bears responsibility for §§ 2.1, 2.2, 4.2 and 5.1. Together, the authors wrote §§ 4.1, 5.2 as well as the introductory parts of §§ 2 and 4.

https://doi.org/10.1075/la.273.00gar

broadly speaking, as the 'interface' between the language system (at different levels: phonological, morphological and syntactic) and discourse.

Over the last decades, IS has established itself as an autonomous field of study, which has been variously integrated into many linguistic theories, both formally and functionally oriented. The success of IS as a research paradigm has contributed to the greater scrutinization of a wealth of experimental and corpus-based data providing valuable information on an increasing number of languages. At the same time, this abundance of empirical material has revealed a number of gaps between certain theoretical assumptions and the data. This is, of course, nothing particularly new or unexpected in research: this tension is a fundamental characteristic underlying the history of science. For this reason, we believe that 'anomalous', hard-to-explain data are key to theoretical development; IS is no exception.

In this article, we shall focus on a specific type of anomalies, namely, 'paradoxical' effects, according to which the pragmatic functions of certain formal devices (such as prosodic features or syntactic structures) in the data seem to contradict or even subvert the ones believed to be more typical in mainstream literature. Significantly, the term 'paradox' must be understood here in its common usage as something unexpected or even counterintuitive in light of a set of specific assumptions.[1]

If the data contradict what a model, or the rules defined within it, predict, one of the obvious consequences would then be to dismiss the model as inadequate.[2] Very often, however, it is not inadequacy *a limine* that provokes 'paradoxical' empirical results. These might also stem from the chosen empirical method or from the incompleteness of the model. In other words, some complications that either result from external conditions influencing the data or that are caused by aspects of the model requiring further elaboration may have gone unnoticed.

---

1. The term 'paradox' as used in this paper should thus not be confused with the more technical term employed in philosophy and logic, in which it usually describes a reasoning process that starts from (apparently) true premises yet produces (apparently) false conclusions. For a survey on the relation between paradoxes and contemporary logic, the reader is referred to Cantini & Bruni (2017).

2. A reviewer observed that 'paradoxes' may put under pressure the *description* of a certain structure, but not the *theory* behind it. The reviewer claims, for instance, that the great frequency of 'non-focalizing' clefts, which will be discussed in § 4.3.2, only creates problems for the description of cleft sentences in some languages, but not for a theory of clefts in general. Following Dryer (2006), we believe that the distinction between description and theory should not be considered so clear-cut and that it is probably more advisable to distinguish between *descriptive theories*, which are about "what languages are like", and *explanatory theories*, which are instead about "*why* languages are the way they are" (Dryer 2006: 207). As for the cleft example: in our view, the frequency of 'non-focalizing' occurrences undermines a descriptive theory of clefts, i.e., one that holds that this structure is focal. In other words, what is contradicted is not simply a language description, but a specific hypothesis on the status and functions of clefts in a certain language.

In this introductory article, we shall try to disentangle the different possible reasons for the rise of some paradoxes within the domain of IS. We do so by considering some notions usually employed in the analysis of discourse (§ 2), such as *Common Ground*, *Questions* and *Alternatives*. We then explore the dimensions of IS (its 'building blocks') and their capability of providing adequate analyses (§ 3). In § 4, we consider the linguistic expressions that serve the realization of IS and we try to shed light on some paradoxes, discussing in detail two examples: left dislocation and cleft sentences in French and Italian. Moreover, based on recent attempts to reduce IS categories to more basic pragmatic functions, we shall address the question as to whether the concept of *Information Structure* and its related notions, such as *focus* and *topic*, should (or should not) be dismissed at all as useful tools for describing information flow and its linguistic realization in language (§ 5). Finally, we shall provide a general outlook on the contributions to this volume.

2. Information structure and discourse organization

If we accept the idea that IS is the 'interface' between a sentence and its discourse context, then in order to give an explicit description of this interaction, we need an explicit model of discourse.

Discourse is obviously an abstract and protean notion. Therefore, we should make clear from the start that we are neither aiming at an overarching description (e.g., discourse as the totality of the communicative activity of certain social groups in a certain historical period, as was proposed mainly, but not exclusively, by the French poststructuralist tradition, in particular by Michel Foucault), nor are we intending to wade into the complexity of *ethnomethodology* and *conversation analysis* in the tradition of authors such as Harold Garfinkel, John Gumperz, Gail Jefferson, Harvey Sacks, Emanuel Schegloff and many others.[3] Rather, we restrict ourselves to a more modest and narrow understanding of this term, similar to that adopted by several researchers working in formal and model-theoretic semantics, such as Hans Kamp (cf. among others Asher & Lascarides 2003; Groenendijk & Stokhof 1991; Kamp, van Genabith & Reyle 2011; Roberts 2012 [1996]). In these contributions, discourse is understood as a sequence of utterances (consisting of sentences),[4] in which each utterance relies on the information conveyed by those preceding it, thus forming an "extended but circumscribed piece of language use created for a coherent purpose" (Clark 1994: 985). Clark also highlights another

3. For a detailed overview of the notion of discourse and its definitions and limits in different disciplines, cf. Angermüller, Maingueneau & Wodak (2014).

4. We use the term 'utterances' for any type of communicative contribution in discourse. We consider 'sentences' to be grammatical units containing at least one speech act.

aspect of discourse, namely, its interactional nature of "joint activity" whose purpose is mainly to "accumulate common ground" (Clark 1994: 989).

The main problem for this narrow notion – and, as we shall see, a source of some of the paradoxes to be explored in this article (§ 4) – is that discourse cannot be reduced to the linguistic co-text, but must also consider any type of information on which an utterance is based and to which it may refer, including the contextual setting and a wide variety of sources, ranging from personal, episodic memories to general 'encyclopedic' knowledge (cf. Clark 1996: 92–121). This means that any utterance is related to an enormous wealth of multilayered information that mostly remains implicit.

Furthermore, each utterance does not rely on this implicit background information alone. Very often, the relation that speakers establish between their utterances and this background information remains implicit as well, being instantiated by a variety of pragmatic inferences, such as conversational implicatures and presuppositions. Thus, the most difficult task in analyzing IS is to establish an explicit model of what largely remains implicit in discourse. Since it is impossible to analyze this complex system in a general top-down approach, it seems more promising to start from the linguistic surface and to explore non-linguistic aspects based on the effects observed in linguistic behavior.

Importantly, this strategy of exploring discourse and its cognitive background is the inverse operation of what happens in real-life communication: in real discourse, in fact, it is background knowledge and the situational setting that are the reason for utterances to occur, and not vice versa. This becomes clear, for example, when we reconstruct *questions under discussion*, or when we claim that *alternatives* are *induced* by focalized expressions, as we will see in §§ 2.2 and 2.3.

Even if one adheres to 'armchair linguistics' at least in the phase of establishing models that are later to be tested on real corpus data or experimentally, we must be aware that reconstructing contexts starting from surface structures is a heuristic strategy only indirectly connected with the real tasks of speakers/hearers. In real discourse, utterances *depend on* the context (and are motived by it, at least partly), and language users *rely on the context* in order to interpret them. Thus, we should be careful not to confuse the task of linguists with the task of speakers: while the former start from the linguistic surface, the latter start from a specific discourse situation, from which utterances and sentences result.

Our impression is that some misunderstandings and paradoxes stem exactly from this confusion between linguistic heuristics and the intuitive strategies of language users. One example is provided by Matić (this volume) concerning made-up questions used as a diagnostic tool for IS categories (cf. also § 2.2 below). In the process of reconstructing contextually relevant questions for an utterance, linguists may be baffled by the fact that the same utterance can answer different

questions, whereas speakers typically recognize the IS of an utterance in discourse without effort.

Although mindful of this warning, we are still convinced that in a research scenario, reconstructing the context from realistic (or, even better, ecologically valid) linguistic examples is the best (if not the only viable) strategy for exploring the "sea of silent but unequivocal communication" (Bühler 2011 [1934]: 176) on which linguistic surfaces emerge. In order to address this task, three major concepts have been employed in linguistic pragmatics: *Common Ground* (§ 2.1), *Question Under Discussion* (§ 2.2) and *Alternatives* (§ 2.3).

2.1 Common ground

The analysis of discourse, of its non-linguistic aspects and the complexity of background knowledge, requires a dynamic model of how single utterances interact with this background:

> […] the speaker-hearer knowledge at a given moment in the conversation can be viewed as a set of propositions that represents the relevant assumptions that the speaker and the hearer share at that moment. […] the speaker-hearer knowledge seems to have more internal structure than a set of propositions […].
>
> (Jacobs 2001: 650–651)

A systematic model of this "internal structure" has been proposed in Relevance Theory (Sperber & Wilson 1995; Wilson & Sperber 2004), in which a 'deductive device' is put forward, as a module where different sources of information (linguistic input, background knowledge, situational information as well as the inferences drawn by speakers/hearers) are integrated into a set of relevant assumptions.

In IS research, a similar model of shared knowledge as a set of assumptions has been established under the term *Common Ground* (henceforth CG):

> If we are to talk about communication as transfer of information and its optimization relative to the temporary needs of interlocutors, it is useful to adopt a model of information exchange that makes use of the notion of **Common Ground**. The original notion of CG […] saw it as a way to model the information that is mutually known to be shared and continuously modified in communication. This allowed for a promising way of modeling the distinction between **presuppositions**, as requirements for the input CG, and assertions or the **proffered content**, as the proposed change in the output CG. (Krifka 2008: 245, emphasis original)

We would like to clarify that the term *shared*, which is common in the definitions of CG, is in a way metaphorical: speaker and hearer are not computers in a network, connected to each other or to a common memory, in which they can directly access their mutual knowledge. Instead, it is more correct to state that each participant in

a given discourse *presumes* that her/his interlocutors have the same set of assumptions as herself/himself. Consequently, Krifka & Musan (2012: 3) consider CG to be "a set of propositions that *is presumed to be mutually accepted*" (our emphasis).

Very often, utterances are not just meant to add propositional information to the CG. This typically holds for speech acts other than assertions. Questions, for instance, are described by Krifka & Musan as a speech act that "specifies the way in which the common ground should develop in the immediate future" (Krifka & Musan 2012: 10).

But even assertions do not always add propositions to the CG. As we shall see below (§ 2.3), the goal of focalized utterances is to *maintain* one alternative among several propositions that are contextually generated, causing the other alternatives to be removed from the CG. Krifka (2008: 246) introduced the distinction between *CG content*, the dimension of CG concerned with semantic, 'truth-conditional' information, and *CG management* for the dimension concerned "with the way in which the common ground content should develop", to which he assigns a more 'pragmatic' function, having to do with "information about the manifest communicative interests and goals of the participants" (Krifka 2008: 246).[5]

2.2 Information structure and (implicit) questions

In current research, the interaction between a sentence and its immediate discourse context is commonly described by means of question-answer pairs (cf. van der Wal 2016: 265). Hermann Paul (1898: 260) is often cited as the first author to have related information structure to its context via potential questions (e.g., in Klein & Von Stutterheim 1992: 75; Matić, this volume). Paul claims that a sentence like (1a) could be an answer to different potential questions (1b)–(e):

(1) a. *Peter fährt morgen nach Berlin.*
'Peter is traveling to Berlin tomorrow.'
b. *Where will Peter travel tomorrow?*
c. *When will Peter travel to Berlin?*
d. *How will Peter travel to Berlin tomorrow?*
e. *Who will travel to Berlin tomorrow?*

5. For extensive treatments of CG, the reader is referred to, among others, Clark (1996: 92–121), Krifka & Musan (2012) and Stalnaker (2002, 2009). In particular, Clark (1996) strove to provide a psychologically sound description of this notion, linking CG to memory (a detailed psycholinguistic account on this aspect has also been recently offered by Horton & Gerrig 2016).

Paul's example is problematic in several respects: in particular, it works at best in written language, while it seems somehow unnatural as an example of spoken language. In fact, this example was not meant to show how sentence structure reacts to context, but rather to establish a discussion about what Paul called the 'psychological predicate' (the *Comment* in modern terminology). His argument is nonetheless relevant for the relation between sentence and discourse, since Paul postulates questions as plausible contexts for certain (made-up) utterances.

This argumentation has become a common strategy in linguistics, namely, reconstructing a context on the basis of a given sentence. In this respect, Matić (this volume) speaks of 'critical contexts' "which are believed to more or less unequivocally reveal the IS of the utterances/sentences which are inserted in these contexts."[6]

In particular, Question Under Discussion (henceforth QUD) models consider questions (both explicit and implicit ones) to be the basic elements of discourse and seek to describe the 'inquiry strategies' followed by interlocutors: e.g., answering a question by addressing another question entailed by the first one or splitting a general question into more specific sub-questions (cf. among many others Beaver & Clark 2008; Ginzburg 2012; Onea 2016; Riester, Brunetti & De Kuthy 2018; Roberts 2012 [1996]; van Kuppevelt 1995; Velleman & Beaver 2016). In Roberts' perspective, for example, discourse is conceived of as a language game, whose goal is to "come to agree on the way things are in the world" (Roberts 2012 [1996]: 4), and whose basic moves are questions and assertions.

It is often observed that IS-related categories such as *contrastive topics* (see §§ 3 and 4.3.1, see also Büring 2003, 2016) serve a "strategy of incremental answering" (Krifka & Musan 2012: 30). Such strategy involves a hierarchy of questions, as shown by Example (2) (Büring 2016, following Krifka 2008: 269):

(2) *(How is business going for Daimler-Chrysler?)*
[*In Germany*]$_{\text{Frame}}$ *the prospects are* [*good*]$_{\text{F}}$, *but* [*in America*]$_{\text{Frame}}$ *they are* [*losing money*]$_{\text{F}}$.

In (2), the main question is broken down into the two sub-questions 'How is business going in Germany?' and 'How is business going in America?', to which the two utterances provide a partial answer contributing to the resolution of the more general question. This strategy might account for some paradoxes, as we shall see more specifically in § 4.3.1.

6. Matić (this volume) and Ozerov (2018, 2021) levy severe criticism against this procedure, in particular concerning its inherent circularity in the recognition and analysis of IS-related categories. Cf. also van der Wal (2016: 265–267) for a discussion on several weaknesses of question-driven approaches in IS.

While QUD-based models were mostly used to explore the formal constraints of QUD representations in discourse (as in Roberts 2012 [1996]), more recently they have been increasingly applied to the analysis of real corpus data. Riester, Brunetti & De Kuthy (2018), for instance, have put forward explicit guidelines for annotating corpora on the basis of QUDs, that is, for reconstructing implicit questions 'retrospectively', starting from a careful analysis of the context. Although this empirically driven endeavor is still in a relatively early phase, it has already offered some interesting insights into the hierarchy of discourse (cf. De Kuthy, Brunetti & Berardi 2019; Riester 2015, 2019; Riester, Brunetti & De Kuthy 2018), based on data representative of different text genres (Riester & Baumann 2013) and languages, and also including lesser-known language varieties (Riester 2015; Riester & Shiohara 2018).

2.3 Alternatives and types of focus

Considering a sentence as the answer to an (implicit) question entails that what is truly essential in an utterance is the information required by the question, which can either be the alternative between a 'true' and an 'untrue' proposition in the case of polar questions, or the element that satisfies the variable represented by the *wh*-form in a partial question. Thus, when at an award ceremony, somebody declares 'The winner is JOHN', this statement provides an answer to the question 'Who is the winner?', which may remain implicit but is quite compelling in such a situation. The focalized constituent *John* replaces the interrogative *wh*-pronoun in the (implicit) question. Thus, the main function of this replacement is to exclude all other possible candidates who might have won the prize.

In the literature, it is commonly accepted that the occurrence of focus is tightly correlated to alternatives. As Krifka & Musan (2012: 10) claim, for instance, "focus especially stresses and points out the existence of particular alternatives". This intuition was explicitly defined and formalized by Rooth (1985, 1992) in the framework of *Alternative Semantics*, in which the 'ordinary semantic value' of the focalized term is accompanied by its 'alternative semantic value', consisting of the alternatives to the element to which the ordinary semantic value refers. This approach intended to provide a formal account of the role of alternatives in a perspective centered on sentences considered independently of context. However, the original approach of *Alternative Semantics* proposed in Rooth (1985) turned out to lack constraints on the size and specificity of alternative sets (see Kratzer 1991; Krifka 2006; Zimmermann & Onea 2011: 1657). The solution was ultimately found in considering the alternatives as being discourse-bound. As Rooth (1996: 1279) claims, "focus interpretation introduces a variable which, like other free variables, needs to find an antecedent or be given a pragmatically constructed value".

Once again, it seems relevant to us to differentiate the sentence-oriented perspective of many linguists (especially formal linguists) concerned with IS from the context-oriented perspective of real-life communication: while sentence-oriented linguists usually deduce alternatives from a focalized sentence, speakers rely on the alternatives raised by QUDs in the context. From the perspective of hearers, we can claim that they treat focalized items as a cue for ruling out alternative propositions.

Thus, this means that speakers focalize an element not to *evoke* alternatives (this would only be the case in certain rhetorical strategies, see § 4.3.1), but rather to eliminate all other alternatives raised by the context (cf. Masia, this volume; Rosemeyer, Jacob & Konieczny, this volume). As observed by Rosemeyer, Jacob & Konieczny (this volume), the relationship between alternatives, focus and the CG is better conceived in a 'subtractive' rather than in an 'additive' fashion. In other words, it is more likely – and psychologically more plausible – that the CG develops in discourse by suppressing the alternatives that do not apply to the current situation, rather than constantly incorporating new items. Thus, focus seems to select some specific alternatives, while discarding others.[7]

A further aspect of focus alternatives, on which many authors agree, is that the constellation of the alternative sets, defined by their domain, size and even internal ordering, is the basis for defining the different types (or 'uses') of focus observed in the literature (cf. among others Krifka 2008; Riester & Baumann 2013; Zimmermann & Onea 2011: 1662–1667).

For instance, the distinction between *expression* (or *metalinguistic*) *focus* and *denotation focus* relates to the domain of the alternative set (Krifka & Musan 2012: 7–9). One can recognize, in fact, alternatives of form, which become evident when a formal aspect of a word or a structure is corrected (such as, for example, its accent structure), or alternatives of denotation, which involve instead the level of semantic content.

Concerning the latter, alternative sets can have different sizes. *Polarity focus*, for example, is based on an alternative set containing only two members, a true and an untrue proposition (i.e., p vs. $\neg p$, cf. Höhle 1992; Leonetti & Escandell-Vidal 2009: 177).

The definition of *novelty* or *information focus*, on the one hand, and *contrastive focus*, on the other hand, can be partly described by the size of the alternative set and partly by the prominence degree of the alternatives within the CG. As suggested by

7. Similar ideas have been expressed by Zimmermann & Onea (2011: 1652) and Roberts (2004) in terms of 'possible worlds'. According to Roberts (2004: 208), for instance, "drawing on Stalnaker's notion of COMMON GROUND and the related CONTEXT SET (i.e., the set of worlds in which all the propositions in CG are true), we can say that our goal is to reduce the context set to a singleton, the actual world".

Riester & Baumann (2013: 233), *novelty focus*, which corresponds to discourse-new information, relies on an open set of "unidentifiable" or "anonymous" alternatives. *Contrastive* focus can instead be defined as being based on a closed set, "whose alternatives can be unanimously identified in the context (i.e., discourse context, encyclopedic knowledge, lexicon etc.)" (Riester & Baumann 2013: 233; cf. also É. Kiss 1998).[8]

Other types of focus recognized in the literature are based on specific discourse-pragmatic functions. *Corrective focus*, for instance, is considered as a subtype of contrastive focus used to correct "part of a previous assertion" (Bianchi, Bocci & Cruschina 2015: 5), or as an "extreme case of contrast" (Riester & Baumann 2013: 230). An interesting instantiation of focus is the so-called *mirative focus* (among others, Bianchi, Bocci & Cruschina 2015, 2016; Cruschina 2019), which has been analyzed as conveying a conventional implicature of unexpectedness or surprise. This type of focus is usually expressed via movement of a constituent to the left periphery of the sentence in Romance and in other language families. It presupposes an alternative set in which alternatives are ranked according to their likelihood in a certain context and which signals that the asserted proposition is less likely than other alternative propositions.

3. The 'building blocks' of information structure

In § 2, we discussed different ideas and concepts related to IS (such as focus and (contrastive) topics), without delving into how they relate to each other and which status they have in current thinking on IS. In this section, we shall explore the different conceptual levels considered to be essential components of IS architecture.

A first basic distinction reflects the difference between what is usually referred to as referential and relational givenness (Gundel & Fretheim 2004). The former is the level of given or new information in relation to the previous discourse context or the knowledge status of the hearer/reader (cf. Baumann & Riester 2012; Prince 1981). The latter concerns instead the partition of the sentence into two portions, X and Y:

8. For a typology of focus in which *information focus* or *new information* is interpreted against a set of alternatives, cf. also Zimmermann & Onea (2011: 1662–1663) and the references provided there. It is important to observe, however, that the notion of *information focus* is not unanimously accepted in the literature. According to Kratzer & Selkirk (2020), considering information focus as a 'proper' kind of focus is misguided. Based on Standard American and British English, these authors observe that new information does not receive any prosodic or syntactic marking, unlike (contrastive) focus. Semantically, the interpretation of new information does not require alternatives, whereas (contrastive) focus does. Kratzer & Selkirk thus suggest that information focus be considered rather as (unmarked) new information.

> X is what the sentence is about (the logical/psychological subject) and Y is what is predicated about X (the logical/psychological predicate). X is given in relation to Y in the sense that it is independent of, and outside the scope of, what is predicated in Y. Y is new in relation to X in the sense that it is new information that is asserted, questioned, etc. about X. (Gundel & Fretheim 2004: 177)

In the IS model outlined by Krifka (2008) and Krifka & Musan (2012), relational givenness is further divided into two distinct levels: one regarding the topic (and its comment) and one concerning the focus (and its background). This three-level model thus relies on the following partitions: (i) *given-new*; (ii) *topic-comment*; (iii) *focus-background*.[9] Let us briefly illustrate these notions by analyzing Example (3):

(3) *Along with ammonia, hair dye formulae contain hydrogen peroxide, a bleaching agent. Peroxide* ***serves two purposes:*** *it* ***reacts with the melanin in hair, extinguishing its natural colour****, and* [ø] ***provokes a reaction between PPD molecules.***
(Rebecca Guenard, 'Hair Dye: A History,' *The Atlantic*, 02.01.2015)[10]

GIVEN-NEW: *Peroxide* and *it* represent in (3) discourse-given information which refers to the NP *hydrogen peroxide* introduced in the first sentence. The content of the VPs (*serves two purposes*; *reacts with the melanin in hair…* and *provokes a reaction between PPD molecules*) are instead discourse-new.

TOPIC-COMMENT: in the very first utterance, *hydrogen peroxide* is contained within the comment (the aboutness topic being *hair dye formulae*). In the subsequent sentence, *peroxide* is an aboutness topic; since the pronominal anaphora *it* and the null subject in the sentence *and provokes a reaction between PPD molecules* are used to maintain the same aboutness topic in discourse, they can be analyzed as instances of given topics (*Aboutness G*[*iven*]*-Topic* in Frascarelli 2017: 477–478). The whole VPs represent instead the comment, in bold in (3) (i.e., what is predicated about peroxide).

FOCUS-BACKGROUND: Following the first sentence in (3), an implicit QUD such as 'What does hydrogen peroxide do?' can be easily reconstructed. Thus, the topical element *peroxide* constitutes the background, whereas the whole VP represents

---

9. In the current literature, IS models have been put forward that differ considerably from the one described here. Several authors in this volume rely, for instance, on Cresti's *Language into Act Theory* (cf. among others Cresti 2018). In this model, rooted in an Austinian perspective (Austin 1962), the Comment is defined as the portion of an utterance that accomplishes its illocutionary force, whereas the Topic is meant to be an optional element that defines the Comment's 'field of application' (for a more detailed presentation, cf. De Cesare, this volume; Lombardi Vallauri, this volume; Masia, this volume).

10. <https://www.theatlantic.com/health/archive/2015/01/hair-dye-ahistory/383934/> (11 June 2021).

information focus. The immediately following text (starting with *it serves two purposes…*) is an answer to another implicit QUD such as 'What are the two purposes of hydrogen peroxide?' Again, the topical elements *it* and the null subject are the background, while the two VPs in the coordinate sentences are the information focus.

This example represents a very simple situation, in which the different IS levels overlap (given information-topic-background; new information-comment-focus). However, 'mismatches' between the three levels are often to be found. As we can observe in the same article, for instance, topical elements are not necessarily given (this confusion was typical of earlier stages of research on IS). In (4), *Barclay Cunningham* is the discourse-new information on which the sentence is built:

(4) *Every two months,* ***Barclay Cunningham*** *goes through a process that begins with taking an antihistamine tablet.*
(Rebecca Guenard, 'Hair Dye: A History,' *The Atlantic*, 02.01.2015)

Moreover, it is also important to observe that the levels 'topic-comment' and 'background-focus' do not involve co-extensional categories. For example, in the made-up Example (5) the syntactic domains of focus and comment are not identical (Krifka & Musan 2012: 28):

(5) A: *Tell me something about Onassis. When did he marry Jacqueline Kennedy?*
B: [*He*]$_{\text{Topic}}$ [*married her* [*in 1968*]$_{\text{Focus}}$]]$_{\text{Comment}}$

Moreover, foci can be also discourse-given (or discourse-inferable), as is commonly the case for contrastive foci (cf. § 2.2). Contrastiveness is another notion that might appear puzzling in the study of IS: several researchers have suggested that 'contrast' should be considered an additional, independent level in the architecture of IS (cf. among others Neeleman & Vermeulen 2012; Torregrossa 2018; Vallduví & Vilkuna 1998). This proposal relies on the observation that contrastive elements tend to receive cross-linguistically specific prosodic, syntactic and morphological markings as opposed to non-contrastive (either focal or topical) elements (cf. also Kratzer & Selkirk 2020).

Other researchers suggest instead that contrast should rather be defined in merely pragmatic terms as a by-product of specific configurations of the alternative set (cf. Repp 2016 for a detailed overview). In other words, as already seen in § 2.3, a restricted and closed alternative set would naturally lead to a contrast between its alternatives. In agreement with this view, both a focus and a topic (as in the case of contrastive topics) can thus be considered contrastive if they evoke restricted sets of alternatives in discourse. As such, contrastive topics fall under the definition of focus, or, as suggested by Krifka, topics "can contain a focus" (Krifka 2008: 267). We shall return to this 'paradox' in § 4.3.1.

4. The realization of information structure and its 'natural paradoxes'

Krifka & Musan (2012: 5–6) warn against a common problem in IS descriptions: notions such as *topic* or *focus* are often used ambiguously, sometimes referring to syntactic entities, sometimes to the pragmatic functions of the denotata of these entities.

While in the generative cartographic framework (e.g., Rizzi 1997; cf. Shlonsky & Bocci 2019 for a recent survey), the terms 'topic' and 'focus' relate directly to specific functional heads in the syntactic structure, functionalist approaches by contrast claim that syntax is just one means for 'expressing' the discourse-pragmatic functions usually assigned to IS. In this perspective, it is generally agreed that there are 'competing' surface strategies (prosodic, syntactic and also morphological and lexical) that convey these discourse functions. This is true not only cross-linguistically, but also language-internally.

4.1 Information structure and prosody

While prosody seems to be the most straightforward means for fine-tuning an utterance to certain pragmatic functions, the analysis of its relationship with IS is one of the most challenging tasks in linguistics. Although it is a common practice in the literature to rely on examples where prosodic prominence is assigned more or less intuitively (usually by means of small capitals), moving away from such phonological idealization towards real, observable phonetic patterns pertains to a level of complexity that we cannot address here.

This complexity does not only stem from the different kinds of prosodic features, such as pitch, intensity and duration, but also from their interaction with one another in signaling different types of phonological events (e.g., the tones identified in the *Autosegmental-Metrical* approach; cf. Arvaniti in press for a survey). On the other hand, a further complication lies in the need to integrate the different levels of prosodic prominence, from the 'basic', lower-level notions of mora, syllable and foot, to the higher-level categories represented by prosodic words, phrases, sentences and even inter-sentential relations, in which speech acts, propositional or interactional attitudes and other pragmatic phenomena also come into play. Consequently, in the prosodic marking of IS, no co-extension of phonological accents with their corresponding syntactic domains can be taken for granted. A well-known case in the literature is represented by English and other West Germanic languages, in which a pitch accent on the rightmost argument in a VP, as in (6), can signal different focus sizes:

(6) I called MARY

This utterance can be used as a plausible answer to different QUDs: *Whom did you call?* (indicating narrow focus on the object NP), *What did you do?* (narrow focus on the whole VP) as well as *What happened?* (broad focus on the whole utterance). According to several authors (cf. among others Baumann, Grice & Steindamm 2006; Ladd 2008), this is likely due to a tendency shown by West Germanic languages to prefer arguments rather than predicates as focus exponents, independently of the specific communicative function intended by the speaker.

This observation also points to a further complication, namely, that the use and distribution of (phonological) accents as a means of marking IS categories such as focus are also subject to language-specific preferences and restrictions. French, for instance, is well-known for its lack of a lexical accent and its strict rules of compulsory accent placement, which lead to its rightmost position within a word, an intonational unit or a sentence (Destruel & Féry 2020: 3), and, consequently, to its reduced flexibility in signaling IS-related categories. Many authors (cf. among others Destruel 2012; Dufter & Gabriel 2016; Lambrecht 1994, 2001) observed that French tends to rely on a syntactic strategy, i.e., cleft constructions, in contexts in which a subject is assigned a focal reading. In such a case, English would prefer instead a prosodic strategy (*in situ* accentuation on the subject constituent), while German and Spanish (as well as other languages such as Italian) would rather displace the focalized subject to post-verbal position (via subject-verb inversion), as in Example (7):[11]

(7) A: *How do you know?*
B: a. *HUMA told me.*
b. *Das hat mir HUMA erzählt.*
c. *Me lo ha contado HUMA.*
d. *C'est HUMA qui me l'a dit* (Lambrecht 2001: 490)

11. This is an obviously simplified description. As observed by Dufter & Gabriel (2016), subject-verb inversion in Spanish examples such as (7c) is favored when the direct object is represented by a clitic pronoun (the same pattern is also observable in Catalan, European Portuguese, Italian and Romanian). When the direct object is a full object NP, the preverbal position is instead preferred. Heidinger (2018) shows that certain varieties of Spanish also select cleft or pseudo-cleft sentences to convey focus on the subject, independently of the focus type (information or contrastive focus). In general, Spanish speakers tend to align the focalized element to the rightmost edge of prosodic constituents, even if they use different strategies, also depending on diatopic variation (Heidinger 2018: 63). As correctly suggested by a reviewer, subject-verb inversion in Spanish may be another interesting case for our inventory of IS 'paradoxes'. A discussion on this paradox would also help highlight the role of language variation (in this case, diatopic variation) in understanding the tension between data and theory (see also García & Uth 2018).

However, Turco, Dimroth & Braun (2013) showed that, at least in certain circumstances, French speakers can exclusively rely on prosody in marking focus. Turco and colleagues carried out a production experiment on the realization of *Verum Focus* in French and German. *Verum Focus* (or polarity focus, cf. §§ 2.3 and 4.3.1) is usually conceived of as a focus on the assertive component of the utterance and is typically marked in German with an accent on the finite form of the verb or the auxiliary, whereas French and other Romance languages prefer different morpho-syntactic and lexical strategies (Garassino & Jacob 2018). In this production experiment, participants were supposed to look at a scene in a picture (e.g., a child tearing a banknote) and contrastively reply to a statement made by a confederate ('the child is not tearing a banknote') by focalizing the assertive component of the sentence ('the child IS tearing a banknote').[12]

Against expectations, Turco and colleagues found that in about 33% of the cases, French speakers phonologically distinguished *Verum Focus* from non-*Verum Focus* by placing a focal accent on the auxiliary, thus violating language-specific constraints (Turco, Dimroth & Braun 2013: 484). This experiment thus shows that even if French speakers preferentially use syntactic or lexical strategies to express focus in normal circumstances, they have nonetheless the intuition of prosodically marking IS, at least under certain conditions.

The results of Destruel & Féry (2020) on dual focus marking in French point in a similar direction. On the one hand, this study shows, by means of a production experiment (involving a reading task), that a high degree of individual variation exists in the prosodic cues used for realizing different IS configurations. On the other hand, a perception test also revealed that French participants are less reliable than English and German speakers in assigning adequate contexts (in the form of questions) to different types of prosodically marked foci.

In conclusion, although the findings of Turco, Dimroth & Braun (2013) and Destruel & Féry (2020) may seem 'paradoxical', they can be better understood if one takes into account the experimental setting. In both studies, the participants' linguistic options seem in fact limited by the experimental design, involving read speech and some possible bias (namely the priming effect of the confederate's sentence in Turco, Dimroth & Braun 2013). The speakers were thus somehow 'led' to employ prosodic cues, even if they might have possibly preferred other strategies in authentic discourse.

12. Interestingly, a similar experiment on Italian and Dutch (Turco 2014: 153) showed that polarity focus in Italian can be achieved by means of a pitch accent on the verb.

4.2 The IS-prosody-syntax triangle

As we have seen in § 4.1, there seems to be a competition between prosody and syntax as strategies for marking IS. Vallduví (1991) proposed distinguishing between 'plastic' languages, which preferentially rely on prosody (such as English) and 'non-plastic' languages, which use syntax (such as French). However, it now seems widely accepted that both levels, albeit independent, are interconnected and that a typological distinction between 'intonation languages' and 'word-order languages' must be substantially revised (cf. Dufter & Gabriel 2016; Face & D'Imperio 2005; García & Uth 2018).

Once a direct interrelation between prosody and syntax is taken for granted, it follows that syntactic operations consisting either in displacing syntactic constituents from the sentence core to the left or right periphery (such as in the case of *dislocation, fronting* and *subject inversion*), or in assigning them a different syntactic function (such as in *passive* constructions and *cleft sentences*),[13] also have as a consequence the placing of these elements under different prosodic conditions. Thus, it is difficult to decide whether certain syntactic structures (e.g., clefts or dislocations) convey IS features directly or whether they do so by virtue of prosodic conditions related to these constructions.

At first sight, there seem to be arguments in favor of both positions. The possibility of overruling the 'natural' alignment between certain syntactic and certain IS-related functions (e.g., focalizing the subject by means of a pitch accent, while de-accentuating the remaining elements, as in Example (7) above) seems to go against the assumption of a direct relation between syntax and IS at least in some languages. On the other hand, however, a syntactic function like *subject* is characterized by a set of features that seem to make it a 'natural' topic, such as its preverbal position, its reduced morpho-phonological weight in some languages,[14] its privileged status in controlling agreement on the verb, etc. (for a comprehensive account, see Keenan 1976). In diachronically oriented, functionalist approaches, the subject has even been considered as a grammaticalized topic, cf. Li & Thompson (1976).

---

13. While in the passive form, the direct object is promoted to a subject function (the 'original' subject constituent being instead suppressed or 'demoted' to an adjunct function), in cleft constructions, a constituent of the corresponding 'canonical' sentence is promoted to the predicate complement of the matrix clause, while the rest of the sentence is embedded in a relative clause (or a pseudo-relative, cf. De Cesare 2017: 560 for a discussion) that presents its content as presupposed.

14. As has been acknowledged in the typological literature, the nominative tends to be expressed in most languages with a case system by the bare nominal stem, in contrast to other cases; cf. among others Primus (2010: 309).

More recently, Evans & Levinson (2009: 440) have observed that the existence of a grammatical subject "is an efficient way to organize a language's grammar because it bundles up different subtasks that most often need to be done together", such as topic, agent, and the syntactic 'pivot' of the sentence.

Relying on a natural alignment between syntactic and IS functions, it is possible to account for the IS-related effects that are observable, for instance, in the passive form. By means of this structure, it is possible to topicalize a constituent by assigning it a subject function. Considered from this perspective, syntactic operations such as passivization or clefting make it possible to overcome a first-order association between syntax and IS functions. However, this line of thinking does not help in deciding whether such effects rely on a direct relation between syntactic and IS functions, or whether the latter stem from the prosodic features of syntactic functions.

This means that we are caught in a 'triangle', in which IS, syntax and prosody are difficult to disentangle. However, since it seems in principle possible (as shown above by the *polarity focus* example in the experiment of Turco, Dimroth & Braun 2013) to express IS-related categories through prosodical means alone, one can claim that the IS-prosody relation is 'privileged' and possibly more fundamental at a cognitive level (see also the observations in Larrivée, this volume and Lombardi Vallauri, this volume).

4.3 'Non-canonical' syntax: Left dislocation and cleft sentences in French and Italian

In this section, we focus on two syntactic structures, left dislocation and cleft sentences, by discussing 'paradoxical' evidence that seems to go against the descriptions usually found in the literature. In particular, we shall discuss left dislocation occurrences in which the dislocated element exhibits both topical and focal functions, as well as cleft sentences used as 'non-focalizing' constructions.

4.3.1 *Left dislocation*

Left dislocation is usually described in terms of IS as a construction in which the dislocated element is assigned a topic function (cf. Frascarelli 2017 for a recent survey on Romance languages). In the case of non-subject dislocation, this structure makes it possible to overrule the 'natural' topic-subject association. By contrast, it might appear 'paradoxical' to dislocate subjects, which normally receive the topic function without resorting to any syntactic displacement.

In such cases, the motivation behind the use of left dislocation, as observed by Frascarelli (2017: 476), is to realize a topic shifting or a contrastive topic. Therefore,

one could say that dislocation has the function of providing dislocated subjects with additional pragmatic values in certain contexts, as we can see in Example (8):[15]

(8) A: *C'est c'que faisaient vos parents ?*
'That is what your parents did?'
B: *Oui ma mère, elle travaillait dans l' vingtième rue d'la Réunion*, […] *et mon père était artisan*
'Yes, my mother, she worked at 20, rue de la Réunion, […] and my father was a craftsman'
(*CFPP* corpus, Brunetti & Avanzi 2017: 39–40; our translation)

The left dislocated subject NP *ma mère*, 'my mother' is used contrastively, in opposition to the contextual alternative *mon père*, 'my father' (i.e., it is a *contrastive topic*, cf. also § 2.2).

Example (8) is in some way reminiscent of Example (2) above, in that it displays the same discourse strategy, namely, splitting up a more general QUD ('What did your parents do?') into a sequence of partial QUDs ('What did your mother do?'; 'What did your father do?'). Splitting up a QUD into sub-questions means that contrastive topics can be at the same time discourse-new and embedded in an accessible semantic domain: in our example, an alternative set which is inferable through the NP *mes parents*, 'my parents'. Moreover, contrastive topics refer to a closed set of *alternatives* that specify the 'frame' for the utterance to be true.

In fact, contrastive topics are described as structures that share features of both topic and focus (e.g. Krifka & Musan 2012: 30; cf. also § 3 above).[16] This would be a paradox only in a theory that fails to distinguish the two independent levels of topic-comment and focus-background. It is not paradoxical at all, however, if we accept the idea that discourse structure is not just a linear concatenation of QUDs, in which the topic of each sentence refers to the immediately preceding question, but rather a complex, hierarchically organized structure; and if we consequently admit that the topic-comment and the focus-background distinctions are two independent levels within IS (as argued in § 3).

Another fact that might seem paradoxical is that (left and right) dislocation can also convey other different IS functions, such as the marking of polarity focus. As shown in Garassino & Jacob (2018), both left and right dislocations in French, Italian and Spanish typically appear in contexts in which the QUD at issue concerns the truth (vs. non-truth) of a previously uttered proposition:

---

15. For the sake of space, we do not address here the status of subject dislocations in null-subject languages, such as Italian and Spanish. The reader is referred to Dufter & Gabriel (2016: 436).

16. According to Kratzer & Selkirk (2020: 5), contrastive topics receive a feature [FoC], shared with other instances of focus, such as exhaustive and mirative focus.

(9) A: *Non è questione che il tempo non te l'ho dato*; ***io te l'ho dato il tempo***
'It's not that I didn't give time to you; I gave it to you, time'
(*LIP* corpus, Frascarelli 2003: 557)

(10) A: *no, niente, eh, trovare una soluzione*
'Nothing, well, we must find a solution'
B: *ah va be', **la soluzione gliela troviamo***
'Ah well, the solution, we're going to find it'
(*LIP* corpus, Brunetti 2009: 763)

In (9) and (10), not only the dislocated constituent, but also the entire propositional content of the dislocations is made up of discourse-given or inferable material. The dislocation occurrences in these examples also express either a polarity contrast with respect to the previous discourse (such as in the right dislocation in (9), *te l'ho dato* 'I have given it to you' vs. *non te l'ho dato* 'I have not given it to you'), or, as in the left dislocation in (10), a confirmation of the truth of the proposition, which is left open in the previous turn (*gliela troviamo* 'we find it' in connection with *trovare una soluzione* 'finding a solution'). In the remainder of this section, for the sake of space, we shall only consider left dislocations.

In order to explain this 'paradoxical' use of a left dislocation, we refer to two aspects of the dislocation construction that may contribute to this polarity focus reading. The first is a structural one: namely, that the structure allows for the finite verb to appear in the rightmost position, where it receives a nuclear pitch accent conveying focality (similar to German or English, where polarity focus can be marked via accentuation of the finite verb or the auxiliary). The second effect pertains to the discourse level. Consider the following example taken from the *Europarl* parallel corpus (Koehn 2005):

(11) *Monsieur le Président,* […] *il faut rendre honneur à la présidence française, il faut rendre honneur au président Chirac, qui a été au charbon, qui a combattu et qui a vaincu sur sa vision de l'Europe, parce que, **lui, il a une vision***
'Mr. President […] we should honor the French Presidency, we should honor President Chirac. He was at the coalface, he fought and conquered for his vision of Europe – because he does have a vision'
(*Europarl*, Garassino & Jacob 2018: 238)

As suggested by the official English translation, the French left dislocation in (11) seems to correspond to a focused auxiliary *do* (for an analysis of English *do* in terms of polarity focus marking, cf. Wilder 2013). A pragmatic analysis reveals that the use of left dislocation in this example obeys a rhetorical strategy in which the hearer inferentially reconstructs the information that there are other politicians who, unlike Chirac, do not have a vision of Europe. This intuition can be made explicit by analyzing the discourse structure by means of QUDs, as in (12):

(12) Q_1: {*Which politicians have a vision of Europe?*}
> $Q_{1.1}$: {*Does politician A have a vision of Europe?*}
> $Q_{1.2}$: {*Does politician B have a vision of Europe?*}
> $Q_{1.3}$: {*Does politician C have a vision of Europe?*}
> $Q_{1.4}$: {*Does Chirac have a vision of Europe?*}

The French left dislocation realizes a contrastive topic (cf. § 2.1), in which alternative subject referents are contrasted. These alternatives, although not explicitly mentioned in discourse, are easily inferable because of the discourse configuration in which a hierarchically higher QUD entails a series of polar sub-questions, such as $Q_{1.1}$, $Q_{1.2}$, etc., and in which only the sub-question concerning Chirac is supposed to elicit a positive answer.

The QUD structure makes it possible to infer these sub-questions and, consequently, to force the hearer/reader to accommodate the presupposition that politicians A, B, C, … do exist. Therefore, an observation that might appear paradoxical at first glance, namely, the occurrence of dislocation in polarity focus contexts, finds its explanation in a discourse strategy involving a rhetorically complex development of the QUD structure.

4.3.2 *Cleft sentences*

Cleft sentences are usually described as focalizing devices (cf. among many others Lambrecht 2001), as in the following made-up French example (for a survey on cleft constructions in Romance, cf. De Cesare 2017):

(13) *Ce n'est pas contre les réfugiés,* ***c'est contre la procédure d'asile qu'il faut lutter***
'It is not against the refugees, it is against the asylum procedure that we should fight'

In this example, the cleft constructions help focalize the contrast between the PPs *contre les réfugiés* and *contre la procedure d'asile*. In light of the discussion in § 2.3, we can identify here a contrastive-corrective focus, characterized by a closed alternative set.

However, as shown by the following examples (gathered from the *Europarl-direct* corpus, Cartoni & Myers 2012 and Koehn 2005, and a corpus of online newspapers, *Contrast-It*, De Cesare 2011–2018), one finds other occurrences in the data which cannot be described as focalizing devices:

(14) *Le rapport proposé aujourd'hui redonne sa place au politique. Il est un des pas vers une politique de l'emploi et du développement durable.* ***C'est ceci qui nous conduira à le voter***
'The report proposed today re-establishes a place for politics. It is one of the steps towards a policy of employment and sustainable development. This is what persuades us to vote in favor of it' (*Europarl-direct*, Vachetta)

(15) *La precedente serie di banconote, uscita nel 1995, aveva, invece, sottolineato le peculiarità culturali della Svizzera.* ***Era per quel motivo che sui 200 franchi trovavamo Ramuz, sui 100 Giacometti, e sui biglietti da 10 Le Corbusier***

'The previous set of banknotes released in 1995 highlighted, on the contrary, the cultural specificities of Switzerland. For that reason, one could find Ramuz on 200 CHF banknotes, Giacometti on 100 CHF banknotes and Le Corbusier on 10 CHF ones [lit. It was for that reason that one could find Ramuz on 200 CHF banknotes etc.]' (*Contrast-It*, *repubblica.it*, 01.10.2011)

As observed by many researchers (starting with Prince 1978; see also Declerck 1984, 1988; Dufter 2009; Garassino, this volume; Hedberg & Fadden 2007; Jacob 2015; Karssenberg & Lahousse 2018), the clefts in (14) and (15) show a 'reversal' of their IS partitions compared to more prototypical examples, such as (13).[17] In 'non-focalizing' occurrences such as (14) and (15), the content of the cleft clause (i.e., the subordinate clause) is discourse-new information, whereas the clefted constituent (i.e., the constituent following the copula in the matrix clause) is merely an anaphoric 'bridge' to the previous text.

Non-focalizing clefts are anything but rare in the data. In fact, several corpus-based works have recently shown that they may actually represent the most frequent examples of the construction, at least in certain languages and genres.[18] This is shown in, among others, Dufter (2009) on cleft sentences in English, German and the Romance languages within the *Europarl* corpus; Roggia (2009), De Cesare et al. (2016), Borreguero Zuloaga (2016) on Italian data drawn from different sources; Garassino (2016) on Italian and English data from online newspapers.[19]

Again, we are faced with a seeming paradox, since non-focalizing clefts may be the most common occurrences of a structure that is usually described as focalizing the cleft constituent and treating the cleft clause as backgrounded information. How

17. Additionally, cleft sentences can have an all-comment/all-focus configuration, as in *c'est avec plaisir que je vous présente mon beau-fils*, 'it is with pleasure that I present my son-in-law to you'. However, these structures should be considered with caution since they mostly have a formulaic function (cf. De Cesare 2020; Garassino, this volume; Hasselgård 2010; Patten 2012).

18. Empirical inquiries on other types of cleft constructions also revealed that these structures allow for a broad spectrum of IS configurations. This is the case, for instance, for *wh*-clefts in English (Hedberg & Fadden 2007; cf. also Garassino 2014 on the IS configurations of reverse-*wh* clefts) and *il y a*-clefts in French (cf. Karssenberg 2018; Karssenberg & Lahousse 2018).

19. Interesting results also come from prosodic research: Frascarelli & Ramaglia (2013), for instance, observe that in their speech corpus of Roman Italian, there are some prototypical examples of cleft sentences, mostly associated with a contrastive reading, that present, as expected, a pitch accent on the clefted element. However, in their corpus there are also less prototypical (but very frequent) examples characterized by unexpected accents or intonational contours (similar results are discussed in Mertens 2012 for French).

can this paradox be solved? According to Prince (1978), this type of non-focalizing clefts, which she labels *informative-presupposition*, fulfils a rhetorical strategy since it presents a new piece of information 'as a fact', i.e., as something known by most people but not necessarily by the hearer/reader, and thus not open to negotiation (Delin 1992). According to several authors, such an effect relies on the presuppositional nature of cleft sentences (Delin & Oberlander 1995; Dryer 1996: 488–491). Following this reasoning, one can claim that both focalizing and non-focalizing clefts presuppose the content of the cleft clause. For instance, in (13) the cleft construction presupposes that one must fight against something and then states that *something* is to be identified with the referent of the clefted element. In the focalizing occurrences, the proposition is usually activated or at least highly accessible. In (14) and (15), the proposition is instead new to the hearer but is presented by the speaker in a presuppositional 'packaging' and, as such, it requires the hearer's accommodation.

It is important to stress that more recent literature has recognized different types of clefts behind the *informative-presupposition* label, so that a more nuanced view is necessary. First, as convincingly argued by Lombardi Vallauri (this volume), it would be a mistake to conclude that all clefts hosting an anaphoric element in their matrix clause cannot be focalizing devices. In fact, there is no contradiction in having a clefted constituent which is both anaphoric and focalized, since discourse-givenness and topicality pertain to distinct and independent levels (see Lombardi Vallauri, this volume for a detailed discussion; cf. also § 3 above).

An example such as (14) can be characterized exactly in this way. Jacob (2015) has also noticed that in (14), the cleft clause, although partially discourse-new, is nonetheless activated in the situational setting, in which the politician is intervening to explain her party resolution to vote. The cleft can thus be considered to answer a covert QUD such as 'What will lead the politician's party to vote?' by means of the anaphoric *ceci*, 'this', which refers to (and provides a conclusion to) the politician's argumentation. In other words, the focus-background partition typical of focalizing clefts also seems to be maintained in this case.

In other occurrences, such as in (15), the propositional content of the structure is entirely discourse-new and the cleft clause can hardly be described as 'presuppositional'. As already observed by Berretta (2002) and Dufter (2009: 104), some instantiations of non-focalizing structures can be described as syntactic devices for ensuring textual cohesion,[20] signaling "that the speaker is moving on to another

20. The clefted element in this type of clefts is usually represented by anaphoric adverbial elements, such as English *thus*, *there*, *then*, French *ainsi*, *là*, Italian *così*, *là*, 'thus, there' (see De Cesare et al. 2016; Lahousse & Lamiroy 2015).

discursive segment", or that the speakers intend "to summarize or resume the argumentation developed so far" (Dufter 2009: 104; cf. also De Cesare et al. 2016; Garassino, this volume; Jacob 2015 for similar observations).[21]

Instead of a categorical opposition between focalizing and non-focalizing clefts, it would probably be more adequate to consider different functional types of clefts viewed along a gradient, especially considering the activation state of the cleft clause (not only in the linguistic context but also in the contextual setting) and the function of the clefted constituent (anaphoric and focal at the same time or only anaphoric). What also emerges from this discussion is that the use of cleft sentences (as well as other exemplars of the cleft network) is not always motivated by IS, but needs to be understood in light of other levels of analysis (in our case, textual cohesion).

Finally, a reviewer also suggests that diachrony should be considered in the explanation of non-focalizing clefts. In fact, Dufter (2008) and Jacob (2015) consider non-focalizing clefts in French to be an outcome of a grammaticalization process originally involving focalizing occurrences. These observations leave space for another source of paradoxes that stem from reanalysis processes inherent to grammaticalization paths. Very often, speakers have different intuitions on the semantic and pragmatic interpretation of certain surface structures, thus triggering a reinterpretation of a linguistic form that may eventually lead to language change ('reanalysis'). This is a possible explanation for non-focalizing clefts.

In conclusion, the protean functional nature of syntactic structures such as cleft constructions and left dislocations should not be considered either paradoxical or very surprising. If we assume that, as in the quotation from Lambrecht presented above in § 1, "the study of information structure focuses on the interaction of sentences and their contexts" (Lambrecht 1994: 9), it naturally follows that the pragmatic functions of such structures are the result of a complex, multifactorial interaction between their core semantic and pragmatic values and specific, strictly context-dependent feature constellations (see also Matić & Nikolaeva 2018).

21. Also observe that the cleft in (14) has a similar textual function.

5. Conclusion and overview of the volume

5.1 As a way of conclusion

Our discussion has shown that the 'paradoxes' discussed in § 4 are only apparent ones and can stem from a variety of possible sources, including language change and several dimensions of synchronic language variation (such as diatopic variation, which is thoroughly discussed in García & Uth 2018). In this chapter, however, we have focused on two specific sources of paradoxes.

The first one is related to the choice of specific experimental methods (e.g., the production tasks used for eliciting polarity and other types of focus examined in § 4.1). The second type, which is more relevant to this volume, is represented by those cases in which paradoxes arise because of inadequate discourse models and/or the failure to take into account other explanatory factors, as shown by our discussion on dislocation (in particular, left dislocation) and cleft sentences in §§ 4.3.1 and 4.3.2.

More generally, these observations suggest that it is necessary to consider 'indirect' effects when we analyze structures that are usually thought to convey IS-related functions. Indirect effects can be conceived as the consequence of general inferential pragmatic mechanisms and rhetorical strategies employed by speakers in authentic interactions. In this perspective, the expression of polarity focus by means of a dislocation is an indirect effect, such as the cohesive function of some occurrences of cleft sentences. Consequently, such effects are the result of using certain structures in specific contexts and, in our view, are still compatible with mainstream IS interpretations.

On the contrary, other scholars have taken a step forward by questioning the legitimacy itself of IS as a field of inquiry (cf. Matić, this volume; Matić & Wedgwood 2013; Ozerov 2018, 2021). These researchers consider categories such as topic and focus to be "procrustean beds" (Matić & Wedgwood 2013: 153) since they force 'top down' explanations on the data, which can lead the analyst to ignore more fundamental aspects. For instance, Matić & Wedgwood observe that the morpheme *baa* in Somali (Matić & Wedgwood 2013: 138–139) can be easily mistaken as a signal of narrow focus, but a careful analysis reveals instead that the functional spectrum of this morpheme is much wider and includes topic-change marking as well as the expression of epistemic values. For Matić and Wedgwood, the danger of treating this morpheme as a 'focus device' is to produce a distorted linguistic analysis that conceals more basic meanings and functions (i.e., *baa* conveying a form of *realis* meaning, "indicating the true […] existence of its complement", Matić & Wedgwood 2013: 139) which may be the ultimate cause of other observed, 'superficial' effects. Ozerov proposes replacing the top-down, universally-oriented methodology of IS with a bottom-up research program, investigating "low-level

interactional and discourse-managing instructions" (Ozerov 2018: 92), which are responsible for a variety of pragmatic effects (including the ones usually ascribed to IS). These ideas have inspired a lively debate in the current literature (cf. Riester 2015 and Riester, Brunetti & De Kuthy 2018 for a critique), which at present seems far from being settled.[22]

A similarly challenging stance, albeit originating from a different theoretical framework, is taken by Kratzer & Selkirk (2020). Their claim is that all the syntactic, semantic, pragmatic and phonological effects that we usually ascribe to IS can be reduced to two syntactic features, namely, [FoC] and [G] (the latter standing for 'givenness'), which can have different prosodic or syntactic realizations across languages and whose meanings consist in signaling contrasts or matches in discourse (cf. Kratzer & Selkirk 2020: 42). Ultimately, for Kratzer & Selkirk, "there is no such thing as information structure that deserves this name". In other words, these authors deny that IS "constitute[s] a field of study of its own, governed by its own principles" (Kratzer & Selkirk 2020: 42).

The conclusion drawn by Kratzer & Selkirk seems to characterize the opinion of all the researchers cited in this section: it is a warning against substantiating a homogeneous linguistic domain or 'module' and attributing to it a number of phenomena that can neither be shown to be exclusive to such a domain, nor exhaustively explained by the categories and rules attributed to this domain. It is also a plea for a sober and analytical consideration of the observed phenomena and processes. We do not see why linguists should not adhere to this enterprise, which promotes greater simplicity as well as terminological and conceptual parsimony. However, we also do not see what should prevent us from maintaining the label *Information Structure* as a term embracing all the processes and phenomena related to the task of adapting sentences to their discourse environment. In our view, what seems to be at stake for current research on IS is:

- The importance of considering textual and interactional aspects (cf. Pekarek-Doehler, De Stefani & Hoerlacher 2015 on the interactional properties of non-canonical syntactic structures in French);
- The need for keeping apart different notions and levels such as relevance, co-text vs. context vs. background knowledge, anaphora, QUDs, givenness, newness vs. contrastivity, inference vs. accommodation, text cohesion and coherence, etc. It is with respect to all these levels that any specific linguistic form and structure should be analyzed;

22. We should mention that this debate has also concerned the peer review process of Dejan Matić's contribution to this volume, leaving unresolved dissent between the author and some of his reviewers.

– The use of a wide (and varied) spectrum of data, belonging to different text types and genres;
– Theoretical models that are able to offer explicit analyses of the manifold discourse strategies employed by speakers.

The articles contained in this volume, which we shall briefly present below, aim at shedding further light on these and other related issues.

5.2 Contributions to this volume

The contributions collected in this book tackle, from different perspectives and theoretical frameworks, several types of 'paradoxical' evidence regarding IS-related categories or specific syntactic structures typically associated to IS, such as focus fronting and cleft sentences.

The articles are sustained by a strong empirical basis: be it direct, as in the case of the experimental and corpus-based research in the second part of the volume, or indirect, as for the theoretical contributions in the first part, which rely on previous empirical research work carried out by the authors. Although most contributions focus on Germanic and Romance languages, some data from the Slavic language family (Bulgarian, in Gabriel & Grünke) and some non-Indo-European language families (in the typologically oriented papers by Masia and Matić) are also discussed, thus presenting a relatively wide perspective on the examined phenomena.

The nine articles have been divided into two distinct sections. The first part of this volume is of a more theoretical nature and contains the papers of Lombardi Vallauri, Masia and Matić, inspecting the foundation of IS-related categories, such as topic and focus.

Edoardo Lombardi Vallauri, in *Distinguishing psychological Given/New from linguistic Topic/Focus makes things clearer*, addresses, in a sort of dialogue with our introductory chapter, some of the 'paradoxes' we have outlined above. Insisting on the need for keeping givenness as a cognitive category separate from topicality as a linguistic one, the author suggests that there is no 'paradox' at all in French and Italian cleft sentences with an anaphorical element in their matrix clause, since the function of these structures is to focalize already active information. Moreover, he discusses the relation between IS, syntax and prosody, claiming that the expression of IS-related categories through prosody is more basic and fundamental than by means of syntactic structures.

In *Remarks on Information Structure marking asymmetries: The epistemological view on the micropragmatic profile of utterances*, **Viviana Masia** criticizes the widespread practice of defining focus based on alternatives. Masia claims that the function of focus is not to evoke contextual alternatives, but rather to express an evidential meaning. As she argues in her article, the most basic difference between

focus and topic involves a different way of encoding evidentiality. While topics indicate knowledge mutually shared by both speaker and hearer at the moment of utterance (*mutual* evidentiality), focus signals knowledge which is only possessed by the speaker at the time of utterance and for which s/he is the only committed source (*individual* evidentiality).

In similar fashion to Masia's article, **Dejan Matić**, in *Alternatives to Information Structure*, strongly advocates linguistic analyses in which focus and topic are reduced to more basic pragmatic and cognitive categories. Matić, though, goes a step further in his criticism of IS as a field of research. Based on typological work, including his own fieldwork experience, he questions the foundational tenets of mainstream IS. Considering categories such as focus and topic as 'procrustean beds' (cf. also § 5.1), Matić suggests abandoning top-down research altogether and relying instead on bottom-up procedures, based on low-level interactional and grammatical properties.

The second part of the volume is devoted to case studies involving Germanic and Romance languages, of both an experimental and corpus-based nature:

The experimental contribution by **Malte Rosemeyer, Daniel Jacob** and **Lars Konieczny**, *How alternatives are created: Specialized background knowledge affects the interpretation of clefts in discourse*, offers a psycholinguistic study of German cleft sentences. The authors show that the same cleft sentence can be assigned a focal or a non-focal reading depending on the reader's knowledge. If a reader fails to generate a set of alternatives in light of the previous context, s/he will interpret the structure as conveying information focus, whereas if s/he is able to reconstruct a set of contextually relevant alternatives, a contrastive reading is assigned. Moreover, the article addresses the complex interaction between text coherence, the speaker's knowledge about the discourse topic as well as certain structural properties of the cleft construction (such as the syntactic function of the clefted constituent).

The role of alternatives is also critically analyzed in *Is focus a root phenomenon?* by **Karen Lahousse**. The author suggests that the IS of French *c'est*-clefts, *il y a*-clefts and the VS word order plays an important role in explaining their distribution in embedded sentences. In other words, their being 'root phenomena' or not is linked to their being contrastive or non-contrastive structures. Contrastive structures rely on the presence of alternatives in discourse, which makes them operate at the assertion level (i.e., they oppose elements already introduced in the CG), a basic pragmatic property associated to non-root phenomena. By contrast, non-contrastive structures, such as clefts conveying information focus, merely introduce new information to the CG, which is instead associated with root phenomena.

In *The curious case of the rare focus movement in French*, **Pierre Larrivée** tackles the relation between data and theory from a cartographic perspective. According to the author, focus fronting, due to its scarcity in French, shows the tension between the rate of realization of a configuration in a language and its assumed availability

in the syntactic structure. In order to solve this 'paradox', Larrivée claims that we should look closely at the possible constraints of a specific language. In his analysis, he considers the competition between focus movement and the more successful *c'est*-clefts, which is ultimately amenable to the syntax-phonology interface, favoring accent placement on the rightmost element of a prosodic unit (which is made possible in *c'est*-clefts by the post-predicate position of the clefted element), while pre-verbal accent placement is disfavored.

Anna-Maria De Cesare, in *To be or not to be focus adverbials? A corpus-driven study of It.* anche *in spontaneous spoken Italian*, shows that *anche* has a remarkable flexibility in her data, being associated with different pragmatic functions. Moreover, De Cesare analyzes the relation between *anche* and different levels of IS as defined in Cresti's model *Language into Act*. While this adverb consistently binds the focus of a semantic proposition articulated in focus-background (the semantic-presuppositional layer of IS), its co-occurrence with the expression bearing the illocutionary force of the utterance (the so-called pragmatic-illocutionary layer in Cresti's model) is far less stable.

The last two papers in this section deal with the challenges posed by IS in the context of language contact and translation:

Christoph Gabriel and **Jonas Grünke**, in *Unmarked use of marked syntactic structures: Possessives and fronting of non-subject XPs in Bulgarian Judeo-Spanish*, analyze possessive constructions and fronting of non-subject constituents in Bulgarian Judeo-Spanish, showing that these structures are frequent in this variety and are also used in IS-neutral contexts, whereas in Mainstream Spanish they are much less frequent and highly marked. The authors suggest that the situation displayed in Bulgarian Judeo-Spanish can either be the result of the maintenance of Old Spanish structural characteristics or of present-day language contact with Bulgarian. If the latter hypothesis should prove to be correct, this would suggest that the realization of IS is likely to undergo changes in situations of linguistic contact.

In *Translation as a source of pragmatic interference? An empirical investigation of French and Italian cleft sentences*, **Davide Garassino** analyzes French and Italian clefts in a corpus of journalistic texts and European parliamentary transcriptions, also including translations from French into Italian. The author observes that there are cross-linguistic differences in the IS configurations of clefts in the two languages and that Italian translations, at least in certain cases, increase the frequency of IS configurations that are usually dispreferred in original Italian texts (such as broad focus clefts). In general, however, the realization of IS in the target language seems resistant to interference from the source that would result in odd or unexpected pragmatic effects.

Acknowledgments

We wish to thank Anna-Maria De Cesare, Marco García García and Viviana Masia for their valuable comments on a first draft of this article as well as to the authors who participated in this volume and to all our colleagues who provided us with evaluations of the contributions. We would also like to thank the Fritz Thyssen Foundation and the *Dr. Jürgen und Irmgard Ulderup Stiftung* for enabling us to organize the workshop "When Data Challenges Theory: Non-Prototypical, Unexpected and Paradoxical Evidence in the Field of Information Structure" (University of Freiburg im Breisgau, February 15–16, 2019), during which most of the contributions collected in this volume were first presented.

Finally, we would like to express our gratitude to Werner Abraham and Elly van Gelderen for accepting this book in the *Linguistik Aktuell / Linguistics Today* series, and to Kees Vaes, Esther Roth and Ineke Elskamp for their patience and continuous support during all stages of the publication process.

References

Angermuller, Johannes, Maingueneau, Dominique & Wodak, Ruth. 2014. The Discourse Studies Reader: An introduction. In *The Discourse Studies Reader*, Johannes Angermuller, Dominique Maingueneau & Ruth Wodak (eds), 2–14. Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/z.184.002int

Arvaniti, Amalia. In press. The Autosegmental-Metrical model of intonational phonology. In *Prosodic Theory and Practice*. Stefanie Shattuck-Hufnagel & Jonathan Barnes (eds). Cambridge, MA: The MIT Press. <https://www.amaliaarvaniti.info/publications> (11 June 2021).

Asher, Nicholas & Lascarides, Alex. 2003. *Logics of Conversation*. Cambridge: Cambridge University Press.

Austin, John L. 1962. *How to Do Things with Words: The William James Lectures Delivered at Harvard University in 1955*. Oxford: Clarendon Press.

Baumann, Stefan, Grice, Martine & Steindamm, Susanne. 2006. Prosodic marking of focus domains – categorical or gradient? In *Proceedings of Speech Prosody 2006*, 301–304. <https://isca-speech.org/archive/sp2006/sp06_065.html> (11 June 2021).

Baumann, Stefan & Riester, Arndt. 2012. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In *Prosody and Meaning*, Gorka Elordieta & Pilar Prieto (eds), 119–162. Berlin-New York: De Gruyter Mouton. https://doi.org/10.1515/9783110261790.119

Beaver, David I. & Clark, Brady Z. 2008. *Sense and Sensitivity: How Focus Determines Meaning*. Hoboken, N.J.: John Wiley & Sons. https://doi.org/10.1002/9781444304176

Berretta, Monica. 2002. "Quello che voglio dire è che": le scisse da strutture topicalizzanti a connettivi testuali. In *La parola al testo. Scritti per Bice Mortara Garavelli*, Gian Luigi Beccaria & Carla Marello (eds), 15–31, Alessandria: Edizioni dell'Orso.

Bianchi, Valentina, Bocci, Giuliano & Cruschina, Silvio. 2015. Focus fronting and its implicatures. In *Romance Languages and Linguistic Theory 2013. Selected Papers from 'Going Romance'*, Enoch O. Aboh, Jeannette Schaeffer & Petra Sleeman (ed), 3–19. Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/rllt.8.01bia

Bianchi, Valentina, Bocci, Giuliano & Cruschina, Silvio. 2016. Focus fronting, unexpectedness, and evaluative implicatures. *Semantics and Pragmatics* 9(3): 1–54. https://doi.org/10.3765/sp.9.3

Borreguero Zuloaga, Margarita. 2016. Elementi anaforici e frasi scisse nei testi giornalistici contemporanei. In *La lingua variabile nei testi letterari, artistici e funzionali contemporanei (1915–2014): analisi, interpretazione, traduzione*, Giovanni Ruffino (ed), 529–542. Firenze, Franco Cesati.

Brunetti, Lisa. 2009. On links and tails in Italian. *Lingua* 119(5): 756–781. https://doi.org/10.1016/j.lingua.2008.10.005

Brunetti, Lisa & Avanzi, Mathieu. 2017. Discourse properties of French clitic left dislocated NPs and their effect on prosody. Ms. <http://www.llf.cnrs.fr/en/node/5811> (11 June 2021).

Bühler, Karl. 2011. *Theory of Language. The Representational Function of Language*. [Sprachtheorie. Die Darstellungsfunktion der Sprache. Jena: G. Fischer, 1934]. Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/z.164

Büring, Daniel. 2003. On D-trees, beans, and B-accents. *Linguistics and Philosophy* 26: 511–545. https://doi.org/10.1023/a:1025887707652

Büring, Daniel. 2016. (Contrastive) Topic. In *The Oxford Handbook of Information Structure*, Caroline Féry & Shinichiro Ishihara (eds), 64–85. Oxford: Oxford University Press.

Cantini, Andrea & Bruni, Riccardo. 2017. Paradoxes and contemporary logic. In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed). <https://plato.stanford.edu/archives/fall2017/entries/paradoxes-contemporary-logic/> (11 June 2021).

Cartoni, Bruno & Meyer, Thomas. 2012. Extracting directional and comparable corpora from a multilingual corpus for translation studies. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC)*, Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds), 2132–2137. Paris: European Language Resources Association.

Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In *Subject and Topic*, Charles N. Li (ed), 25–55. New York: Academic Press.

Chafe, Wallace. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: The University of Chicago Press.

Clark, Herbert H. 1994. Discourse in production. In *Handbook of Psycholinguistics*, Morton A. Gernsbacher (ed), 985–1021. San Diego: Academic Press.

Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620539

Cresti, Emanuela. 2018. The illocution-prosody relationship and the Information Pattern in spontaneous speech according to the Language into Act Theory (L-AcT). *Linguistik Online*, 88(1). https://doi.org/10.13092/lo.88.4189

Cruschina, Silvio. 2019. Focus Fronting in Spanish: Mirative implicature and information structure. *Probus* 31(1): 119–146. https://doi.org/10.1515/probus-2018-0008

De Cesare, Anna-Maria. 2011–2018. Contrast-It, University of Basel. <https://contrast-it.philhist.unibas.ch/en/home/> (11 June 2021).

De Cesare, Anna-Maria. 2017. Cleft constructions. In *Manual of Romance Morphosyntax and Syntax* [Manuals of Romance Linguistics 17], Elisabeth Stark & Andreas Dufter (eds), 536–568. Berlin-New York: De Gruyter Mouton. https://doi.org/10.1515/9783110377088-015

De Cesare, Anna-Maria. 2020. *È con piacere che … / C'est avec plaisir que …* On Italian and French manner adverbial clefts expressing emotional states. In *Pour une perspective fonctionnelle des constructions syntaxiques marquées* [Special issue], Anna-Maria De Cesare & Mervi Helkkula (eds). *Neuphilologische Mitteilungen* 120: 429–448.

De Cesare, Anna-Maria. This volume. To be or not to be focus adverbials? A corpus-driven study of It. *anche* in spontaneous spoken Italian. In *When Data Challenges Theory. Unexpected and Paradoxical Evidence in Information Structure*, Davide Garassino & Daniel Jacob (eds). Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/la.273.07de

De Cesare, Anna-Maria, Garassino, Davide, Agar Marco, Rocío, Albom, Ana & Cimmino, Doriana. 2016. *Sintassi marcata dell'italiano contemporaneo in prospettiva contrastiva con il francese, lo spagnolo, il tedesco e l'inglese. Uno studio basato sulla scrittura dei quotidiani online*. Frankfurt: Peter Lang.

Declerck, Renaat. 1984. The pragmatics of *it*-clefts and *wh*-clefts. *Lingua* 64(4): 251–289. https://doi.org/10.1016/0024-3841(84)90065-2

Declerck, Renaat. 1988. *Studies on Copular Sentences, Clefts and Pseudo-clefts*. Berlin-New York: De Gruyter Mouton. https://doi.org/10.1515/9783110869330

De Kuthy, Kordula, Brunetti, Lisa & Berardi, Marta. 2019. In *Proceedings of the 13th Linguistic Annotation Workshop*, 113–123. <https://www.aclweb.org/anthology/W19-40.pdf> (11 June 2021). https://doi.org/10.18653/v1/W19-4014

Delin, Judith. 1992. Properties of *it*-cleft presupposition. *Journal of Semantics* 9: 179–196. https://doi.org/10.1093/jos/9.4.289

Delin, Judith & Oberlander, Jon. 1995. Syntactic constraints on discourse structure: The case of *it*-clefts. *Linguistics* 33: 465–500. https://doi.org/10.1515/ling.1995.33.3.465

Destruel, Emilie. 2012. The French *c'est*-cleft: An empirical study on its meaning and use. In *Empirical Issues in Syntax and Semantics* 9, Christopher Piñon (ed), 95–112. <https://sites.google.com/site/eissvolumes/eiss-9> (11 June 2021).

Destruel, Emilie & Féry, Caroline. 2020. Prominence in French dual focus. *Language and Speech* 64(2): 319–345. https://doi.org/10.1177/0023830919892024

Dryer, Matthew S. 1996. Focus, pragmatic presupposition, and activated propositions. *Journal of Pragmatics* 26(4): 475–523. https://doi.org/10.1016/0378-2166(95)00059-3

Dryer, Matthew S. 2006. Descriptive theories, explanatory theories, and basic linguistic theory. In *Catching Language: Issues in Grammar Writing*, Felix Ameka, Alan Dench & Nicholas Evans (eds), 207–234. Berlin-New York: De Gruyter Mouton.

Dufter, Andreas. 2008. On explaining the rise of *c'est*-clefts in French. In *The Paradox of Grammatical Change*, Ulrich Detges & Richard Waltereit (eds), 31–56. Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.293.03duf

Dufter, Andreas. 2009. Clefting and discourse organization: Comparing Germanic and Romance. In *Focus and Background in Romance Languages*, Andreas Dufter & Daniel Jacob (eds), 83–121. Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.112.05duf

Dufter, Andreas & Gabriel, Christoph. 2016. Information structure, prosody, and word order. In *Manual of Grammatical Interfaces in Romance* [Manuals of Romance Linguistics 10], Susann Fischer & Christoph Gabriel (eds), 419–455. Berlin-New York: De Gruyter Mouton. https://doi.org/10.1515/9783110311860-017

É. Kiss, Katalin. 1998. Identificational focus versus information focus. *Language* 74(2): 245–273. https://doi.org/10.1353/lan.1998.0211

Evans, Nicholas & Levinson, Stephen C. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* (32): 429–492. https://doi.org/10.1017/S0140525X0999094X

Face, Timothy & D'Imperio, Mariapaola. 2005. Reconsidering a focal typology: Evidence from Spanish and Italian. *Italian Journal of Linguistics* 17(2): 271–289.

Frascarelli, Mara. 2003. Topicalizzazione e ripresa clitica. Analisi sincronica, confronto diacronico e considerazioni tipologiche. In *Italia linguistica anno Mille. Italia linguistica anno Duemila*, Nicoletta Maraschio & Teresa Poggi Salani (eds), 547–562. Roma: Bulzoni.

Frascarelli, Mara. 2017. Dislocations and Framings. In *Manual of Romance Morphosyntax and Syntax* [Manuals of Romance Linguistics 17], Elisabeth Stark & Andreas Dufter (eds), 472–501. Berlin-New York: De Gruyter Mouton. https://doi.org/10.1515/9783110377088-013

Frascarelli, Mara & Ramaglia, Francesca. 2013. Pseudoclefts at the Syntax-Prosody-Discourse Interface. In *The Structure of Clefts*, Katherina Hartmann & Tonjes Veenstra (eds), 97–138. Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/la.208.04fra

Garassino, Davide. 2014. Reverse pseudo-clefts in Italian and English. A contrastive analysis. In *Tra romanistica e germanistica: lingua, testo, cognizione e cultura / Between Romance and Germanic: Language, Text, Cognition, and Culture*, Iørn Korzen, Angela Ferrari & Anna-Maria De Cesare (eds), 55–74. Bern: Peter Lang.

Garassino, Davide. 2016. Using Cleft sentences in Italian and English. A multifactorial analysis. In *Current Issues in Italian, Romance and Germanic Non-canonical Word Orders. Syntax – Information Structure – Discourse Organization*, Anna-Maria De Cesare & Davide Garassino (eds), 181–204. Frankfurt: Peter Lang.

Garassino, Davide. This volume. Translation as a source of pragmatic interference? An empirical investigation of French and Italian cleft sentences. In *When Data Challenges Theory. Unexpected and Paradoxical Evidence in Information Structure*, Davide Garassino & Daniel Jacob (eds). Amsterdam-Philadelphia: John Benjamins.

Garassino, Davide & Jacob, Daniel. 2018. Polarity Focus and non-canonical syntax in Italian, French and Spanish. Clitic left dislocation and *sì che* / *sí que*-constructions. In *The Grammatical Realization of Polarity Contrast. Theoretical, Empirical, and Typological Approaches*, Christine Dimroth & Stefan Sudhoff (eds), 227–254. Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/la.249.08gar

García García, Marco & Uth, Melanie. 2018. Introduction. Core issues of focus realization in Romance. In *Focus Realization in Romance and Beyond*, Marco García García & Melanie Uth (eds), 1–30. Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.201.01uth

Ginzburg, Jonathan. 2012. *The Interactive Stance: Meaning for Conversation*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199697922.001.0001

Groenendijk, Jeroen & Stokhof, Martin. 1991. Dynamic Predicate Logic. *Linguistics and Philosophy* 14: 39–100. https://doi.org/10.1007/BF00628304

Gundel, Jeanette K. & Fretheim, Thorstein. 2004. Topic and focus. In *The Handbook of Pragmatics*, Laurence R. Horn & Gregory Ward (eds), 175–195. Oxford: Blackwell.

Hasselgård, Hilde. 2010. *Adjunct Adverbials in English*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511676253

Hedberg, Nancy & Fadden, Lorna. 2007. The information structure of *it*-clefts, *wh*-clefts and reverse *wh*-clefts in English. In *The Grammar Pragmatics Interface: Essays in Honor of Jeanette K. Gundel*, Nancy Hedberg & Robert Zacharski (eds), 49–76. Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/pbns.155.05hed

Heidinger, Steffen. 2018. Acceptability and frequency in Spanish focus marking. In *Focus Realization in Romance and Beyond*, Marco García García & Melanie Uth (eds), 99–128. Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.201.04hei

Höhle, Tilman N. 1992. Über Verum-Fokus im Deutschen. In *Informationsstruktur und Grammatik. Linguistische Berichte Sonderhefte* (Vol. 4), Jacobs Joachim (ed), 112–141. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-663-12176-3_5

Horton, William S. & Gerrig, Richard J. 2016. Revisiting the memory-based processing approach to common ground. In *Memory and Common Ground Processes in Language* [Special Issue], Sarah Brown-Schmidt, Melissa C. Duft & William S. Horton (eds). *Topics* 8(4): 780–795. https://doi.org/10.1111/tops.12216

Jacob, Daniel. 2015. Anaphorische Spaltsätze im Französischen: Grammatik – Text – Rhetorik. In *Informationsstrukturen im Kontrast*, Séverine Adam, Michael Schecker & Daniel Jacob (eds), 101–122. Frankfurt: Peter Lang.

Jacobs, Joachim. 2001. The dimensions of topic-comment. *Linguistics* 39(4): 641–681. https://doi.org/10.1515/ling.2001.027

Kamp, Hans, van Genabith, Josef & Reyle, Uwe. 2011. Discourse Representation Theory. In *The Handbook of Philosophical Logic*, volume 15, Dov M. Gabbay & Franz Guenthner (eds), 124–394. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-0485-5_3

Karssenberg, Lena. 2018. *Non-Prototypical Clefts in French. A Corpus Analysis of "il y a" clefts*. Berlin-New York: De Gruyter Mouton. https://doi.org/10.1515/9783110586435

Karssenberg, Lena & Lahousse, Karen. 2018. The information structure of French *il y a* & *c'est* clefts: A corpus-based analysis. *Linguistics* 56(3): 513–548. https://doi.org/10.1515/ling-2018-0004

Keenan, Edward L. 1976. Towards a universal definition of 'subject'. In *Subject and Topic*, Charles N. Li (ed), 305–333. New York: Academic Press.

Klein, Wolfgang & Von Stutterheim, Christiane. 1992. Textstruktur und referentielle Bewegung. *Zeitschrift für Literaturwissenschaft und Linguistik* 86: 67–92.

Koehn, Philipp. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of Machine Translation Summit X*, 79–86.

Kratzer, Angelika. 1991. The representation of focus. In *Semantik / Semantics. Ein internationales Handbuch der zeitgenössischen Forschung. An International Handbook of Contemporary Research*, Armin von Stechow & Dieter Wunderlich (eds), 825–834. Berlin-New York: De Gruyter Mouton. https://doi.org/10.1515/9783110126969.10.825

Kratzer, Angelika & Selkirk, Elisabeth. 2020. Deconstructing information structure. *Glossa* 5(1): 113, 1–53. https://doi.org/10.5334/gjgl.968

Krifka, Manfred. 2006. Association with focus phrases. In *Architecture of Focus*, Victoria Molnar, & Susanne Winkler (eds), 105–136. Berlin-New York: De Gruyter Mouton. https://doi.org/10.1515/9783110922011.105

Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3–4): 243–276. https://doi.org/10.1556/ALing.55.2008.3-4.2

Krifka, Manfred & Musan, Renate. 2012. Information structure: Overview and linguistic issues. In *The Expression of Information Structure*, Manfred Krifka & Renate Musan (eds), 1–44. Berlin-New York: De Gruyter Mouton. https://doi.org/10.1515/9783110261608.1

Ladd, Robert D. 2008. *Intonational Phonology*, 2nd edn. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511808814

Lahousse, Karen & Lamiroy, Béatrice. 2015. "C'est ainsi que": grammaticalisation ou lexicalisation ou les deux à la fois? *Journal of French Language Studies* 27: 161–185. https://doi.org/10.1017/S0959269515000381

Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representation of Discourse Referents*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620607

Lambrecht, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39(3): 463–516. https://doi.org/10.1515/ling.2001.021

Larrivée, Pierre. This volume. The curious case of the rare focus movement in French. In *When Data Challenges Theory. Unexpected and Paradoxical Evidence in Information Structure*, Davide Garassino & Daniel Jacob (eds). Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/la.273.06lar

Leonetti, Manuel & Escandell-Vidal, Maria Victoria. 2009. Fronting and verum focus in Spanish. In *Focus and Background in Romance Languages*, Andreas Dufter & Daniel Jacob (eds), 155–204. Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.112.07leo

Li, Charles N. & Thompson, Sandra. 1976. Subject and topic: A new typology of language. In *Subject and Topic*, Charles N. Li (ed), 459–489. New York: Academic Press.

Lombardi Vallauri, Edoardo. This volume. Distinguishing psychological Given/New from linguistic Topic/Focus makes things clearer. In *When Data Challenges Theory. Unexpected and Paradoxical Evidence in Information Structure*, Davide Garassino & Daniel Jacob (eds). Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/la.273.01val

Masia, Viviana. This volume. Remarks on Information Structure marking asymmetries: The epistemological view on the micropragmatic profile of utterances. In *When Data Challenges Theory. Unexpected and Paradoxical Evidence in Information Structure*, Davide Garassino & Daniel Jacob (eds). Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/la.273.02mas

Matić, Dejan. This volume. Alternatives to information structure. In *When Data Challenges Theory. Unexpected and Paradoxical Evidence in Information Structure*, Davide Garassino & Daniel Jacob (eds). Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/la.273.03mat

Matić, Dejan & Nikolaeva, Irina. 2018. From polarity focus to salient polarity: From things to processes. In *The Grammatical Realization of Polarity Contrast. Theoretical, Empirical, and Typological Approaches*, Christine Dimroth & Stefan Sudhoff (eds), 9–53. Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/la.249.01mat

Matić, Dejan & Wedgwood, Daniel. 2013. The meaning of focus: The significance of an interpretation-based category in cross-linguistic analysis. *Journal of Linguistics* 49: 127–163. https://doi.org/10.1017/S0022226712000345

Mertens, Piet. 2012. La prosodie des clivées. In *Penser les langues avec Claire Blanche-Benveniste*, Sandrine Caddéo, Marie-Noëlle Roubaud, Magali Rouquier & Frédéric Sabio (eds), 127–139. Aix-en-Provence: Presses Universitaires de Provence.

Neeleman, Ad & Vermeulen, Reiko. 2012. The syntactic expression of information structure. In *The Syntax of Topic, Focus, and Contrast. An Interface-based Approach*, Ad Neeleman & Reiko Vermeulen (eds), 1–38. Berlin-New York: De Gruyter Mouton. https://doi.org/10.1515/9781614511458.1

Onea, Edgar. 2016. *Potential Questions at the Semantic-Pragmatic Interface*. Leiden: Brill. https://doi.org/10.1163/9789004217935

Ozerov, Pavel. 2018. Tracing the sources of Information Structure: Towards the study of interactional management of information. *Journal of Pragmatics* 138: 77–97. https://doi.org/10.1016/j.pragma.2018.08.017

Ozerov, Pavel. 2021. Multifactorial Information Management (MIM): Summing up the emerging alternative to information structure. *Linguistics Vanguard* 7(1). https://doi.org/10.1515/lingvan-2020-0039

Patten, Amanda L. 2012. *The English It-cleft: A Constructional Approach and a Diachronic Investigation*. Berlin-New York: De Gruyter Mouton. https://doi.org/10.1515/9783110279528
Paul, Hermann. 1898. *Prinzipien der Sprachgeschichte*. Halle (Saale): Niemeyer.
Pekarek Doehler, Simona, De Stefani, Elwis, & Horlacher, Anne-Sylvie. 2015. *Time and Emergence in Grammar. Dislocation, Topicalization and Hanging Topic in French Talk-in-interaction*. Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/slsi.28
Primus, Beatrice. 2010. Case marking typology. In *The Oxford Handbook of Linguistic Typology*, Jae Jung Song (ed), 303–321. Oxford: Oxford University Press.
Prince, Ellen F. 1978. A comparison of WH-clefts and *it*-clefts in discourse. *Language* 54: 883–906. https://doi.org/10.2307/413238
Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In *Radical Pragmatics*, Peter Cole (ed), 233–255. New York: Academic Press.
Repp, Sophie. 2016. Contrast: Dissecting an elusive information-structural notion and its role in grammar. *Oxford Handbook of Information Structure*, Caroline Féry & Shinichiro Ishihara (eds), 270–289. Oxford: Oxford University Press.
Riester, Arndt. 2015. Analyzing Questions under Discussion and information structure in a Balinese narrative. In *Proceedings of the Second International Workshop on Information Structure of Austronesian Languages*, Linguistics Dynamics Science Project 2, 1–26. Tokyo: Research Institute for Languages and Cultures of Asia and Africa.
Riester, Arndt. 2019. Constructing QUD trees. In *Questions in Discourse. Vol. 2: Pragmatics*, Malte Zimmermann, Klaus von Heusinger & Edgar Onea (eds), 163–192. Leiden: Brill. https://doi.org/10.1163/9789004378322_007
Riester, Arndt & Baumann, Stefan. 2013. Focus triggers and focus types form a corpus perspective. *Dialogue & Discourse* 4 (2): 215–248. https://doi.org/10.5087/dad.2013.210
Riester, Arndt & Shiohara, Asako. 2018. Information structure in Sumbawa: A QUD analysis. In *Perspectives on Information Structure in Austronesian Languages*, Sonja Riesberg, Asako Shiohara & Atsuko Utsumi (eds), 285–311. Berlin: Language Science Press.
Riester, Arndt, Brunetti, Lisa & De Kuthy, Kordula. 2018. Annotation guidelines for Questions under Discussion and information structure. In *Information Structure in Lesser-Described Languages: Studies in Syntax and Prosody*, Evangelia Adamou, Katharina Haude & Martine Vanhove (eds), 403–443. Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.199.14rie
Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of Grammar: A Handbook of Generative Syntax*, Liliane Haegeman (ed), 281–337. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-011-5420-8_7
Roberts, Craige. 2004. Context in dynamic interpretation. In *The Handbook of Pragmatics*, Laurence R. Horn & Gregory Ward (eds), 197–220. Oxford: Blackwell.
Roberts, Craige. 2012 [1996]. Information structure: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics* 5: 1–69. https://doi.org/10.3765/sp.5.6
Roggia, Carlo Enrico. 2009. *Le frasi scisse in italiano. Struttura informativa e funzioni discorsive*. Geneva: Slatkine.
Rooth, Mats. 1985. Association with Focus. PhD Dissertation, University of Massachusetts at Amherst.
Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1): 75–116. https://doi.org/10.1007/BF02342617
Rooth, Mats. 1996. Focus. In *The Handbook of Contemporary Semantic Theory*, Shalom Lappin (ed), 271–297. Oxford; Blackwell.

Rosemeyer, Malte, Jacob, Daniel & Konieczny, Lars. This volume. How alternatives are created: Specialized background knowledge affects the interpretation of clefts in discourse. In *When Data Challenges Theory. Unexpected and Paradoxical Evidence in Information Structure*, Davide Garassino & Daniel Jacob (eds). Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/la.273.04val5

Shlonski, Ur & Bocci, Giuliano. 2019. Syntactic cartography. In *Oxford Research Encyclopedia of Linguistics*. <https://oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-310> (11 June 2021). https://doi.org/10.1093/acrefore/9780199384655.013.310

Sperber, Dan & Wilson, Deirdre. 1995. *Relevance: Communication and Cognition*, 2nd edn. Oxford: Blackwell.

Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy* 25: 701–721. https://doi.org/10.1023/A:1020867916902

Stalnaker, Robert. 2009. A response to Abbott on presupposition and common ground. *Linguistics and Philosophy* 31: 539–44. https://doi.org/10.1007/s10988-008-9047-9

Torregrossa, Jacopo. 2018. Distinguishing focus and contrast at PF: A view from Italian. In *Focus Realization and Interpretation in Romance and Beyond*, Marco García-García & Melanie Uth (eds), 173–200. Amsterdam-Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.201.06tor

Turco, Giuseppina. 2014. Contrasting Opposite Polarity in Germanic and Romance Languages: Verum Focus and Affirmative Particles in Native Speakers and Advanced L2 Learners. PhD thesis, Radboud University Nijmegen.

Turco, Giuseppina, Dimroth, Christine & Braun, Bettina. 2013. Intonational means to mark verum focus in German and French. *Language and Speech* 56(4): 460–490. https://doi.org/10.1177/0023830912460506

Vallduví, Enric. 1991. The role of plasticity in the association of focus and prominence. *Proceedings of the Eastern States Conference on Linguistics* (ESCOL) 7: 295–306.

Vallduví, Enric & Vilkuna, Maria. 1998. On rheme and kontrast. In *The Limits of Syntax* [Syntax and Semantics 29], Peter Culicover & Louise McNally (eds), 79–108. New York: Academic Press. https://doi.org/10.1163/9789004373167_005

van der Wal, Jenneke. 2016. Diagnosing focus. *Studies in Language* 40: 259–301. https://doi.org/10.1075/sl.40.2.01van

van Kuppevelt, Jan. 1995. Discourse structure, topicality and questioning. *Journal of Linguistics* 31: 109–147. https://doi.org/10.1017/S002222670000058X

Velleman, Leah & Beaver, David I. 2016. Question-based models of information structure. In *Oxford Handbook of Information Structure*, Caroline Féry & Shinichiro Ishihara (eds), 86–107. Oxford: Oxford University Press.

Wilder, Chris. 2013. English 'emphatic do'. *Lingua* 128: 142–171. https://doi.org/10.1016/j.lingua.2012.10.005

Wilson, Deirdre & Sperber, Dan. 2004. Relevance theory. In *The Handbook of Pragmatics*, Laurence R. Horn & Gregory Ward (eds), 607–632. Oxford: Blackwell.

Zimmermann, Malte & Onea, Edgar. 2011. Focus marking and focus interpretation. *Lingua* 121: 1651–1670. https://doi.org/10.1016/j.lingua.2011.06.002

PART I

Theoretical studies