The empirical investigation of information structure

Stavros Skopeteas

1. Preliminaries[1]

The underdetermination of theory by empirical data is an inherent property of scientific research, also beyond the study of language. For instance, it was widely held at the time of Copernicus that his theory and Ptolemaic theory do not necessarily make different predictions concerning the available astronomical data at that time (see Newton-Smith 2000: 532). Nevertheless, it is virtually impossible to carry out a scientific investigation that is free from theoretical considerations. First, even if empirical data exist independently of theories, their relevance for the scientific knowledge can only be judged in the light of a particular theory (see Feyerabend 1983: §3). Second, it is not possible to report on observations without reference to background assumptions that are derived by inductive procedures and are necessarily theory-specific. Hence, an astronomer adopting a geocentric theory such as Ptolemy and an astronomer adopting a heliocentric theory such as Copernicus may observe the same empirical phenomenon, i.e., the perceived position of the sun, but keep track of different observations.

The challenge of the present article is to report about the methods of investigating information structure abstracting away from the theoretical possibilities that are used in order to interpret the empirical data. The empirical question is what are the linguistic forms and the information-structural concepts that constitute the building blocks of a research paradigm. For example, accent as a formal property was presumably only established by Paul (1880), and it seems that Dretske (1972) was first in claiming a specific category of contrastive focus. Given a set of linguistic forms and a set of information-structural properties, we can address the question whether there is a correlation between *form* and *function*, such as 'movement to the preverbal position' ~ 'focus' or 'left dislocation' ~ 'topic'. The observation of such a correlation only allows for weak statements concerning the form-function correspondence. Such a correlation raises the next empirical question: is the functional concept a necessary and/or sufficient condition for the

occurrence of the form at issue? The respective kind of evidence is required in order to qualify statements about the association between form and function.

Empirical statements concerning the association between form and function are supported by different types of linguistic evidence. There are principally two types of observations that lead to the conclusion that particular formal properties of the utterance depend on contextual properties. They are presented in (1) and (2), whereby E_i is a generalization over linguistic expressions and C_j is a generalization over contexts.

(1)   A property E_i of linguistic expressions *occurs* if a contextual condition C_j holds true.

For instance, it has been observed that passive voice in English more frequently occurs when the patient constituent is part of the given information (see Mathesius 1975: 156). Statements of this type result from observational research on corpora (see section 2) and can be experimentally tested in studies on language production (see section 3).

(2)   A property E_i of linguistic expressions *triggers the intuition of felicity* if a contextual condition C_j holds true.

For instance, it has been observed that an utterance with a narrow focus domain on the subject constituent (manifested through the nuclear stress on this constituent) is felicitous in the context of constituent questions on the subject (see Krifka 2002: 295). Statements of this type prove the congruence of particular formal properties of the utterance with particular contexts and are based either on the intuition of an (ideal) native speaker or on rating experiments carried out by a representative sample of native speakers (see section 4).

The types of data talked about in (1) and (2) require several paradigms of empirical investigation. These range from qualitative and quantitative observational studies on naturalistic data to experimental studies on language production or perception. Current debates on linguistic methodology led some researchers to the exclusion of particular types of evidence from the scope of grammatical studies. For instance, Newmeyer (2003) argues that corpus-frequencies are irrelevant for grammatical models, while Sampson (2007) claims that the intuition of grammaticality is a delusion. However, contemporary linguistic research arrives at the consensus that different methodological approaches are rather complementary in that they shed light on different aspects of the research object, a view that is clearly reflected in the contributions of a recent forum hosted by the *Zeitschrift für Sprachwissenschaft*, issue 28 (see especially Featherston 2009, Haspelmath 2009). The present article supports this line of thought in showing that the generalizations

on information structure that are obtained by means of a particular data-gathering method can only partially be replicated through a different method. A significant correlation between properties of linguistic expressions and information structural concepts in any type of data reveals an empirical phenomenon, which is a challenge for any theoretical attempt on this issue. By consequence, exclusion of a data type is equivalent to narrowing down the scope of descriptive adequacy of a linguistic theory.

Main emphasis is given to phenomena on the morphological and syntactic levels (e.g., word order, passivization, clitic doubling, etc.); the reader is referred to Chen (this volume) for the empirical study of prosodic issues. The present article discusses straightforward reflexes of the context on the form of the utterance, as manifested in observational research on corpora or experimental research on native speakers' production or intuition. Studies that explore the processing mechanism of the same range of phenomena (such as experiments on reading times or neurophysiological studies) are not included in the present summary; the reader is referred to Cowles (this volume) for further discussion.

In the following summary of empirical research, no sharp distinction is drawn between qualitative and quantitative data, since the majority of hypotheses can be examined in both types of evidence. Relying on qualitative data implies the assumption that singular observations are not affected by the variation that is involved in linguistic phenomena. Hence, the crucial difference lies on the reliability of these types of linguistic evidence, since generalizations out of singular observations are very likely to be biased – especially if they are obtained by native speakers that are initiated into the targets of the investigation.

2. Occurrence in context: Naturalistic evidence

This section deals with observational research on naturalistic data, i.e., corpora of spoken or written discourse. The range of observable phenomena in naturalistic discourse contains those information structural concepts that can be identified in a text corpus by means of operational definitions applying on observational data. As it is argued below, not every informational structural concept can be identified in observational data.

2.1. Overview

The most frequently used information structural concepts in corpus studies relate to the discourse status of the referents (given, new, and related distinctions). For instance, Prince (1981: 243) supports her taxonomy of new, inferrable, and evoked

(as well as subtypes thereof) referents with quantitative evidence from the frequency of occurrence of these concepts in oral and written text. Birner (1994: 244) distinguishes between given/new in discourse and given/new in hearer's perspective in order to find out the licensing factors of word order inversion in English. Weber and Müller (2004) use a binary distinction of givenness: referents are coded as given if they are mentioned within the last two sentences, otherwise they are coded as new. In a study of dative alternation in English, Bresnan and Hay (2007) classify the referents as 'evoked' (mentioned within the previous 10 lines of discourse), 'situationally-evoked' (first or second person), and 'non-given'.

A paradigm of empirical investigations of the influence of contextual factors on the choice of order and voice evolved within the framework of Givón (1994), based on previous ideas of Cooreman (1987). In this framework, discourse status is conceived as a scalar notion (i.e., anaphoric distance to the antecedent in the pretext, counted in n of sentences) and is combined with a cataphoric measure, which captures the persistence of the discourse referent in the subsequent text (counted in n of occurrences in the 10 subsequent clauses). This quantitative framework has been used in order to identify the contextual conditions that license voice alternations, e.g., active/direct vs. passive vs. inverse in Kutenai (see Dryer 1994) and particular word order configurations such as clitic left dislocation in Modern Greek (see Roland 1994). Furthermore, the role of anaphoric relations in combination with their structural properties is also a central part of the empirical studies that have been conducted within the framework of Centering Theory (see Grosz and Sidner 1986; Walker, Joshi, and Prince 1998, eds.; Beaver 2004; Poesio et al. 2004; and Stede, this volume). In this framework, any utterance establishes a set of referents that are hierarchically ordered on the basis of structural properties (argument hierarchy). The choice of the discourse prominent referent among the members of this set depends on several strategies for the maintenance of discourse coherence.

A number of corpus studies examine focus-related properties. Herring and Polillo (1995) investigate the question of whether focus is expressed in the preverbal position in V-final languages (based on corpus data from Sinhala and Tamil). Their observations on the placement of focused referents are based on the criterion of 'new mentions', i.e., they capture instances of new information focus. A further concept that can be effectively identified in corpora is the concept of contrast. Brunetti (2009a) presents a qualitative corpus study on focus fronting in Italian and Spanish. The analysis of the contexts in which these utterances occur shows that only a subset of the fronted foci relates to an antecedent in discourse that is contrasted or corrected by the focused referent. A further focus-related concept is exhaustivity, which is used by Wedgwood, Pethő, and Cann (2003) in the analysis of Hungarian corpus data. An occurrence of focus-fronting is classified as 'non-exhaustive' if it contains an expression implying that the fronted referent is

not the only member of the set of relevant referents for which the presupposition holds true, e.g., 'primarily', 'for the most part' or 'least of all'. Another property related to focus is the occurrence of focus-sensitive particles, such as 'only', 'also', and 'even'. Matić (2003) observes the occurrence of focus-sensitive particles in focus-fronted constituents in Albanian, Greek, and Serbo-Croatian and Brunetti (2009a) makes similar observations in Italian and Spanish corpora.

Apart from the studies that examine the impact of givenness on topicalization, some empirical studies consider the concept of delimitation as found in contrastive topics and frame-setters (see Krifka and Musan, this volume, section 5.2). Speyer (2007: 104–110) describes the types of elements in the German prefield that violate the expectations of topic continuity as postulated within the theorems of Centering Theory. In these utterances, he identifies instances of contrastive topicalization and frame setting (among other things). Brunetti (2009b: 283f.) reports that contrastive topicalization is among the factors that may induce non-canonical orders with non-subject topics in Italian and Spanish texts.

The overview of corpus studies presented in this section reveals an asymmetry with respect to the use of information structural concepts in observational studies. We may observe that an overwhelming number of studies examine givenness as a contextual factor, while there are no empirical studies based on the notion of aboutness. This asymmetry relates to inherent limitations of the observational data, and these limitations reflect essential differences in the nature of information structural concepts.

The first difference relates to the fact that some information-structural concepts relate to properties of the common ground, while others relate to properties of the target utterance. Hence, categories relating to givenness distinctions are observable in the corpus: given referents can be identified by the previous mentionings in the context, and inferrable referents can be deduced on the basis of assumptions about bridging inferences (see Clark 1977, Prince 1981). On the other hand, the exhaustive interpretation of particular structural configurations is an interpretational property of the utterance and not a contextual property, hence there are no contexts that require an exhaustive expression. Hence, the identification of exhaustive expressions does not relate to the examination of the context. Rather, it relates to selectional restrictions of the construction at issue (e.g., combination with adverbials such as 'primarily'), i.e., to the potential of the constructions at issue to host information that is explicitly not exhaustively identified. In contrast to studies on givenness, there is not a paradigm of investigations collecting the contexts that induce an exhaustive expression and reporting the range of linguistic expressions that are attested in these contexts.

The second difference lies in the availability of observable properties that can be operationalized for the identification of an information structural concept. The givenness of a referent can be quite straightforwardly identified in the corpus

by means of an operational definition: e.g., a referent is decoded as given if it is mentioned in the last two sentences as in Weber and Müller (2004), or in the last ten lines of discourse as in Bresnan and Hay (2007), etc. A similar operationalization is not possible for the concept of aboutness: this concept reflects an intuition about the relation between the topic constituent and its comment; hence, it does not refer to a contextual property that can be observed in corpus data.

2.2. Empirical statements

The choice of the appropriate data from naturalistic discourse depends on the empirical statement at issue. A type of statement that is usually tested in corpora is that some properties of linguistic expressions, e.g. particular word orders, depend on certain contextual conditions. A basic approach is to collect tokens of a particular type of linguistic expression in a text corpus and to identify the range of contexts in which the target expression occurs. The findings of such corpus studies verify *existential hypotheses*, i.e., they examine the possibility of a particular expression to occur under certain contextual conditions, as given in (3).

(3)   A property E_i of linguistic expressions may occur, if a contextual condition C_j holds true.

For instance, Herring (1994: 121) collects utterances with postverbal material in written and spoken Tamil narratives and identifies three contextual conditions in which this structural configuration can occur: (a) afterthoughts, (b) background material, and (c) focused material. In a similar vein, Brunetti (2009b) collects utterances with non-canonical orders in the corpus, i.e., utterances whose word order cannot be accounted for by the structural configuration or the semantic properties of the referents (e.g., non-canonical orders of the arguments of experiencer verbs) and identifies the information structural properties of the fronted constituents: contrastive/corrective focus, contrastive topic, topic shift.

The observation of the possibility of a linguistic expression to occur in a particular context can refute *universal negative hypotheses*, i.e., hypotheses of the type presented in (4).

(4)   A property E_k of linguistic expressions does not occur when a contextual condition C_l holds true.

For instance, a number of accounts conclude that movement to the preverbal position is associated with an exhaustive operator (see Szabolcsi 1981, Horvath 1986, Kiss 1998). Empirically, this statement implies that every instance of

movement to the preverbal position requires an exhaustive interpretation. This is a falsifiable universal hypothesis that has been tested in the Hungarian national corpus by Wedgwood, Pethő, and Cann (2003). In order to refute the hypothesis of exhaustive interpretation, the authors extract a number of sentences (approx. 1000) involving a constituent in the preverbal position from the corpus and examine their contextual properties. The critical sentences for the refutation of the universal hypothesis of an exhaustive operator are those that involve a non-exhaustively identified referent, in particular sentences that involve the expressions 'for the most part', 'least of all', 'primarily', etc. in the preverbal position. The occurrence of non-exhaustively identified referents in the preverbal position, as exemplified in (5), is evidence against the universal hypothesis at issue.

(5)　　*A　　küldöttségben Chris　Patten,　az　unió　külügyi*
　　　the　delegation-in　Chris　Patten　the　union　foreign
　　　biztosa　　　　　　　*mellett helyet　kap　Javier Solana, akiket*
　　　commissioner-POSS.3SG　beside　place　gets　Javier Solana　whom
　　　útjukra　többek között　Anna　Lindh svéd　külügyminiszter
　　　way-on　others among　　Anna　Lindh Swedish　foreign-minister
　　　kisér　majd　el.
　　　accompany FUT　VM
　　　'In the delegation, Javier Solana will be included in addition to Chris Patten, the foreign commissioner of the EU, and they will also be accompanied by among others the Swedish foreign minister Anna Lindh.'
　　　(see Wedgwood, Pethő, and Cann 2003: 14)

In a similar vein, Brunetti (2009a) rejects the universal hypothesis that fronted foci in Italian and Spanish are associated with the concept of contrastive focus in presenting tokens of the target construction without a contrasted or corrected antecedent. The methodological approach in the studies of Wedgwood, Pethő, and Cann (2003) and Brunetti (2009a) is similar: since the aim is to refute a universal hypothesis concerning the functional properties of focus fronting, the authors extract tokens of the target construction from the corpus and investigate the range of contexts in which this construction occurs.

　　　Universal negative hypotheses can even be refuted by a single (positive) counterexample in naturalistic discourse. For instance, Davison (1984: 814) claims that only specific indefinites can be topicalized, which amounts to the universal negative hypothesis that non-specific indefinites cannot be topicalized. Interestingly, Gundel (1974: 187) makes exactly the opposite claim: only non-specific indefinites can be topicalized, which implies that specific indefinites cannot do so. Ward and Prince (1991) reject both hypotheses by presenting single counterexamples from naturalistic discourse. Example (6a) presents a topicalized

non-specific indefinite, which is evidence against Davison's hypothesis, and example (6b) a topicalized specific indefinite, which is evidence against Gundel's hypothesis. (But note that there are accentual differences between the sentences, which means that the underling concept of "topicalization" is presumably not a uniform one).

(6)　a.　*Brains you're born with. <u>A great body</u> you have to work at.* [Brooke Shields, in health club commercial] (Ward and Prince 1991: 170)
　　b.　<u>*Several of these questions*</u> *I will try to answer – but, let me emphasize, from a personal rather than a general viewpoint.* [Nixon 1962:xiii] (Ward and Prince 1991: 171)

Another case of a refutation of a universal hypothesis can be found in Gundel, Hedberg and Zacharski (2005), who examine third person personal pronouns that do not have an explicit antecedent in discourse. They extract a sample of 2 046 third person personal pronouns from the Santa Barbara Corpus of Spoken American English and show that in some tokens the personal pronoun does not refer to an explicit antecedent, as exemplified in (7). In this example, the third personal pronoun in the answer of B refers to the couple (Trish and her husband) that is inferrable from the introduction of the referent Trish in the common ground.

(7)　A:　*Was it Trish who told me she was pregnant?*
　　B:　*She looked really good. Where are <u>they</u> going to church?* [13.221] (Gundel, Hedberg, and Zacharski 2005)

Beyond citing single counterexamples, the authors report the frequency of pronouns without an explicit antecedent in the data set (330 tokens classified in several types, i.e., 16.3% of the examined tokens of third person pronouns). However, the observation of a single counterexample would have been sufficient evidence to reject the universal hypothesis that every personal pronoun has an explicit antecedent in discourse.[2] Reporting counts is informative in order to reject the intuition that a single counterexample may be accidental. However, in the absence of a baseline that indicates the amount of accidental occurrences, there is no principled way to prove that the reported frequency is sufficient, i.e., it corresponds to a probability of occurrence that is beyond the chance level.

So far we discussed deterministic hypotheses concerning the correlation of properties of linguistic expressions with contextual conditions. We have exemplified empirical situations that confirm the truth of existential hypotheses or

---

[2] The aim of the authors is not to reject the universal hypothesis, but to describe the types of reference resolution of third person pronouns and their frequency in spontaneous discourse. Hence, the discussion about the relevance of the quantitative study for the refutation of the universal hypothesis is independent from the aims of this article.

reject the truth of universal hypotheses. The question is whether the opposite claims can be justified by observational data, i.e. whether it is possible to refute an existential statement as in (3) or to justify a universal hypothesis as in (4) if the expression-context pair does not occur in the data (see also discussion of this problem in semantic research in Krifka 2010, §2.1). For instance, Birner (1994) reports that in a data set of 703 tokens of inversion in English, 533 tokens (75.8%) involve a discourse-old initial constituent and discourse-new final constituent, 141 tokens (20.1%) involve two discourse-new constituents, 29 tokens (4.1%) involve two discourse-old constituents, and no utterance involves a discourse-new initial constituent and a discourse-old final constituent. On the basis of this empirical finding she induces a pragmatic constraint: "the preposed element in an inversion must not be newer in the discourse than the postposed element" (Birner 1994: 245).

This conclusion is confronted with Hume's problem of induction, i.e., with the problem that we are not justified to infer a universal statement out of singular ones. In the words of Karl Popper "no matter how many instances of white swans we may have observed, this does not justify the conclusion that *all* swans are white" (Popper 1934, ch. I, §1). In a strict empiricist viewpoint, universal hypotheses can be never justified, which is equivalent to saying that existential hypotheses can be never refuted. However, this problem concerns the validity of the inference and not the relevance of the empirical statement per se. Hence, the fact that no exceptions to a universal hypothesis (or no justifying instances of an existential hypothesis) are attested is a relevant empirical statement as such. The inference of a universal generalization out of such observations is an unnecessary risk (see Kuhn 1970: 18). However, the evidential basis of linguistic generalizations is particular in that it may be complemented by intuition data, which allow confirming universal generalizations, as argued in section 4 below.

The corpus studies mentioned so far are based on samples of the target expression only. This approach is effective in order to identify the range of contexts in which the target construction does or does not occur. An inherent limitation of these empirical studies comes from the fact that they only observe a subset of the data that is determined by properties of the dependent variable. The interpretation of the findings involves the assumption that the range of encountered contexts deviates from the range of possible contexts that appear in discourse independently of the target expression. For instance, Birner's (1994: 245) observation that the occurrences of inversion in English do no involve utterances with a discourse-new initial constituent and a discourse-old postverbal constituent could also be accounted for through the hypothesis that lower constituents (i.e., the initial ones in inversion) always outrank higher constituents (i.e., the postverbal ones in inversion) in discourse-givenness. The belief that this conflicting hypothesis is false is reasonable and may be supported through reference to empirical studies that show

that the opposite holds, but the reported data does not allow to discriminate between these theoretical options.

In order to empirically prove that a particular property of an expression depends on context, we not only need to know the range of contexts in which the target property occurs but also whether this range significantly differs from the range of contexts in which the the target property does not occur. There are fundamentally two ways to formulate a differential hypothesis of this kind. A straightforward formulation of the functional distinction between two alternative expressions is based on the conditional probability of particular contextual properties to occur, as in (8).

(8)   The conditional probability of a context C_i given a property of expression E_k is greater than the conditional probability of the same context C_i given a property of expression E_l, i.e., $prob(C_i | E_k) > prob(C_i | E_l)$.

The aim of the corpus study of Weber and Müller (2004) is to identify the impact of definiteness, givenness, and pronominalization on the choice between the SVO and the OVS order in German (based on the NEGRA newspaper corpus). In order to obtain comparable frequencies, the authors extract the exhaustive set of tokens of the less frequent option (OVS sentences) in the corpus (n = 625) and an equal random sample of tokens of the more frequent option (625 SVO sentences out of total 2773). In order to observe the effects of givenness, subject and object constituents are coded as given (i.e., mentioned in the two last sentences) or non-given. The reported counts allow for generalizations with respect to the likelihood of the four possible permutations of ±given subjects and ±given objects to apply, when German speakers use SVO or OVS utterances. The empirical findings show that the OVS order is more likely to occur than the SVO order when the object of a sentence is given and the subject new.

The reverse empirical question is how likely it is for a set of properties of linguistic expressions to occur when particular contextual conditions hold true.

(9)   The conditional probability of a property of expression E_i given a context C_k is greater than the conditional probability of the same property E_i given a context C_l, i.e., $prob(E_i | C_k) > prob(E_i | C_l)$.

The hypothesis in (9) cannot be tested in the data set of Weber and Müller (2004) since the sample only includes a subset of the SVO sentences in the corpus. I.e., the likelihood for the restricted order (OVS) to occur in a particular contextual condition cannot be estimated since a subset of the occurrences of the unrestricted option (SVO) are excluded from the sample. This type of hypothesis can be examined in the empirical study on dative alternation by Bresnan and Hay (2007),

that reports the frequencies of two word order options (*give* NP NP; *give* NP PP) in all occurrences of the verb *give* in the 'Origins of New Zealand English' corpus: a non-given recipient is approximately four times more likely than a given recipient to be expressed in the prepositional dative.

3. Occurrence in context: Experimental evidence

This section discusses semi-naturalistic data, i.e., data that are induced by production experiments on information structure (for further experimental approaches see also Cowles, this volume). Similarly to the naturalistic data, data from experimental speech production allows for generalizations concerning the dependence of particular aspects of linguistic expressions on contextual properties.

There are several reasons why experimental data are used to complement or substitute the observations in naturalistic discourse. First, corpora allowing for large-scale quantitative studies are only available for some well-studied languages. Second, many structural phenomena do not depend on a single factor but rather on the interplay of a large number of factors. Take for instance the studies on German word order: structural asymmetries (subject > object), semantic asymmetries such as animacy (animate > inanimate), asymmetries relating to information structure such as definiteness (definite > indefinite), and pure formal asymmetries such as weight (short > long constituent) have an impact on the choice of word order in discourse (see empirical findings in Bader and Häussler 2010). I.e., even in a large corpus the selection of minimal pairs in order to observe the exact effect of each factor in isolation results into a substantial reduction of data set. Furthermore, naturalistic data involve further sources of variation from random factors such as different speakers, different genres, different discourse situations, etc.[3] The methodological contribution of experimental studies is exactly this: the interaction of information structural factors with further relevant factors as well as the influence of random factors can be controlled in the experimental design.

3.1. Overview

An overwhelming number of studies on language production examine the preference for given information to precede new information. A straightforward implementation of the givenness asymmetries is to elicit semi-spontaneous narratives by means of non-verbal stimuli (pictures or videos) that induce repeated

---

[3] The diversity of sources of variation in the naturalistic data is not a fatal problem: statistical models based on logistic regression can be used in order to calculate the effect of the involved factors (see Bresnan et al. 2007, Bader and Häussler 2010).

mentioning of the (intended) given referent. Such a manipulation is reported in Skopeteas and Fanselow (2009).[4] The speaker is presented a series of two pictures and is instructed to describe the pictures one after the other. The first scene introduces the relevant context; the second scene induces the target utterance in which the effects of givenness may be observed. The contextual manipulation that induces utterances with a given agent and a new patient is exemplified in (10) with illustrative data produced by a native speaker of American English.

(10) Contextual condition "given agent"
　　　Stimulus: {Pict. 1: A man. Pict. 2: The man is attacking a woman.}
　　　Speaker's reaction:
　　　There's a man walking. Now the man is attacking a woman.
　　　(see Skopeteas and Fanselow 2009: 324)

The contextual manipulation that induces utterances with a given patient and a new agent is illustrated in (11), again with data from American English. The effect of givenness may be observed in the choice of a linearization in which the patient constituent precedes the agent constituent. This is achieved by a passive clause in languages such as American English, Canadian French, Dutch, Yucatec Maya, and German. Speakers of other languages such as Georgian, Czech, Hungarian, Konkani (Indo-European), Prinmi (Tibeto-Burman), and Teribe (Chibchan) produce utterances with a non-canonical word order such that the patient precedes the agent constituent (see discussion of the data in Skopeteas and Fanselow 2009).

(11) Contextual condition "given patient"
　　　Stimulus: {Pict. 1: A woman. Pict. 2: The man is attacking a woman.}
　　　Speaker's reaction:
　　　There's a woman who's walking. Now she's attacked by a man from behind.
　　　(see Skopeteas and Fanselow 2009: 324)

　　　There are numerous experimental studies that manipulate givenness by means of stimuli inducing repeated mentionings of the given referent. Prentice (1967) was among the first to mention that in a picture description task entities introduced in an immediately preceding picture were more likely to appear early in the sentence. MacWhinney and Bates (1978) present an experimental study on English, Italian and Hungarian based on the repeated mentioning of referents or actions and examine hypotheses relating to ellipsis, pronominalization,

---

[4] The experimental procedure used in this study is part of QUIS (=Questionnaire on information structure) that is a collection of production experiments for the study of information structure developed within the research institute 632 *Information Structure* at the University of Potsdam and the Humboldt University Berlin (see documentation of the experiments in Skopeteas et al. 2006).

definiteness, and word order. Prat-Sala (1997) examines the interaction between animacy and information structure in the choice of word order in English, Spanish, Catalan, and Brazilian Portuguese and makes use of a contextual manipulation that reveals the impact of discourse saliency on the choice of linearization (see also Prat-Sala and Branigan 2000 on English and Spanish). Hörnig and Féry (2009) present a production study on spatial configurations manipulated through the presentation of animal toys and discuss the effects of givenness on role choice, word order, definiteness, and prosodic structure of locative expressions in German. Herbert Clark also conducted a number of production studies examining the effects of the repeated mentioning of the referents in discourse (see Clark and Brennan 1991, see a picture-based experiment in Isaacs and Clark 1987, and an experiment on descriptions of Tangram figures in Clark and Wilkes-Gibbs 1986). These studies show that during the progress of the conversation speakers establish the discourse referents in the common ground to the effect that the reference to these discourse entities is proportionally simplified to their salience, from complex descriptions to descriptions with a full NP to descriptions with pronominal expressions and finally ellipsis.

Some production studies manipulate the focal attention of the speaker at the moment of the utterance. Tomlin (1995, 1997) elicits descriptions of a film presenting a series of scenes with two fish with the same shape and different colours. The two fish enter the screen from opposite directions and meet at the center of the screen, at which point one of them eats the other. In each scene, an arrow accompanies one of the two fish and the test subject is instructed to keep his/her eyes on the character the arrow points at. The produced data are exemplified in (12)-(13) from Bahasa Indonesia. When the speaker keeps his eyes on the agent constituent, (s)he produces active sentences, as exemplified in (12).

(12) Contextual condition "agent=primed"
Stimulus: {A pink fish pointed at by the arrow eats a white fish.}
Speaker's reaction:
ikan   merah muda ber-temu   dengan   ikan   putih.
fish   red   light   INTR-meet   with   fish   white
Dan   me-makan ikan   putih.
and   ACT-meet   fish   white
'The pink fish meets the white fish. And it eats the white fish.' (see Tomlin 1995: 535)

When the speaker's attention is directed towards the patient constituent, (s)he produces passive sentences, as exemplified in (13). Similar results are obtained in English, Burmese, and Mandarin (see Tomlin 1995: 531–537).

13

(13) Contextual condition "patient=primed"
   Stimulus: {A white fish eats a pink fish pointed at by the arrow.}
   Speaker's reaction:
   ikan   merah muda ber-temu   dengan   ikan   putih.
   fish   red   light INTR-meet with   fish   white
   Dan   di-makan   oleh   ikan   putih.
   and   PASS-meet by   fish   white
   'The pink fish meets the white fish. And it is eaten by the white fish.' (see Tomlin 1995: 535)

Experiments concerning the attention of the speaker are reported in several studies (see Myachykov 2005: 353–358 for an overview). Forrest (1997) observed similar effects by manipulating the speaker's attention to particular parts of the presented scene. Furthermore, speaker's attention can be manipulated through asymmetries in the salience of the presented entities (see Johnson-Laird 1968). Manipulations of speaker's attention or asymmetries in the salience of the presented entities are not information structural concepts but rather perceptual properties of the stimulus. The question is how these manipulations relate to information structure. The related concept is the concept of aboutness: speakers tend to produce expressions about the entity that is at the center of their attention at the critical time of producing the utterance at issue. The expectation that the center of attention is likely to be realized as an aboutness topic is in line with the observation in these studies that the referent at issue tends to appear early in the utterance.

A further experimental technique that is used for the study of information structure is the elicitation of semi-spontaneous answers to questions. Christianson and Ferreira (2005) present an empirical study on the voice and order alternation in Odawa (a dialect of Ojibwe, Algonquian). Verb forms in this language show a tripartite alternation between active/direct, inverse, and passive voice. In the production study, the speaker was presented a picture and was instructed to answer a question relative to it. The types of questions used in this study are exemplified in (14). The semi-spontaneously elicited answers show that the frequency of occurrence of inverse/passive clauses increases depending on the question type: agent question < general question < patient question (see results and discussion in Christianson and Ferreira 2005: 121–132).

(14) Stimulus: {a boy is pinching a girl}
   a.   General question: 'What is happening here?'
   b.   Agent question: 'What is the boy doing?'
   c.   Patient question: 'What is happening to the girl?'

Answers to questions have been used for a wide range of information structural concepts: see Arnold et al. (2000), who investigate the interaction of givenness with constituent weight in the choice between a theme-goal and a goal-theme linearization in postverbal orders in English; see Skopeteas and Fanselow (2010) on the elicitation of contrastive/non-contrastive types of narrow focus, and Skopeteas and Féry (2007) on the elicitation of contrastive topicalization through multiple constituent questions.

Apart from the experimental paradigms illustrated in this section, several further types of controlled speech production have been used in the investigation of information structure, such as recall techniques (see, e.g., Bock 1977, Bock and Irwin 1980 on givenness effects in several English constructions, Ferreira and Yoshita 2003 on givenness effects in Japanese scrambling), forced-choice experiments (see Vion and Colas 1995 on French clefts), and sentence completion tasks (see, e.g., Kaiser 2006 on the effects of focus and clefting on discourse prominence, Weskott et al. 2006 on the influence of discourse status on the choice between OVS and SVO order in German, Onea and Heusinger 2009 on the contextual conditions that license clitic pronouns in Romanian, Quesada and Skopeteas 2010 on incremental properties of the choice of voice in Teribe).

3.2. Empirical statements

A frequently stated problem of hypothesis-driven investigations is that they draw generalizations out of a small number of data points. Experimental studies have to obey several technical limitations relating to the proportion of targets and fillers, the number of repeated observations per participant, and the reasonable size of an experimental session. These limitations have the effect that the information resulting from experimental findings is restricted to the differences between a few categories. However, linguistic theories often require a large amount of empirical data especially when dealing with multifactorial problems, such as the occurrence of particular properties of linguistic expressions in discourse (see Fanselow 2009: 134f.).

The selection of hypotheses out of a set of potentially relevant factors necessarily implies an inductive step. These empirical studies face a classical problem of the inductive relation between hypotheses and data, which is illustrated by the riddle of Nelson Goodman (1954): Assume that we examine the truth of the statement "all emeralds are green" and after a (sufficient) number of verifications we conclude that this statement is confirmed. Assume now a predicate 'grue' that applies to the observed 'green' entities and also to not yet observed 'blue' entities. The problem is exactly that the statement "all emeralds are grue" is also confirmed by the range of data that we considered so far, i.e., that the empirical data is

underspecified for the hypotheses it potentially verifies. One might argue that 'grue' is an unnatural concept, but the problem of the naturalness of the concepts is independent from the problem of induction (see discussion in Sloman and Lagnado 2005: 97). The following example illustrates how this inductive problem applies on "entrenched" concepts (in Goodman's terms), i.e., concepts that have a past history of use in scientific research.

We observed in section 3.1 that in contexts in which the agent constituent is part of the given information, English speakers tend to produce utterances with agent subjects as in (10), while in contexts in which the patient constituent is part of the given information the same speakers tend to produce utterances with patient subjects. On the basis of this observation, we may conclude that givenness has an impact on hierarchical syntax, such that given referents are encoded in higher syntactic roles (as Bock and Warren 1985 argue for conceptually accessible referents). The data presented so far can be accounted for by the hypothesis "Given information is more likely to be realized in a higher syntactic constituent". However, English is a subject-initial language, and the same range of data can be also accounted for by the hypothesis "Given information is more likely to be realized early in the utterance" (see Clark and Haviland 1977). Both hypotheses are verified through the data, which implies that the empirical data is underspecified for these theoretical options.

In order to discriminate among these two theoretical possibilities we need to consider the critical conditions that test conflicting predictions of the alternative hypotheses. Hence, Bock and Warren (1985) report that the accessible-first principle does not influence the order of coordinated conjuncts, which supports the view that the observed phenomenon relates to hierarchical structure. However, several studies on the influence of discourse factors on word order show that the choice of passive is cross-linguistically in complementary distribution with word order operations. Given-first effects affect syntactic relations only in constructions/languages in which two alternative syntactic realizations of the same propositional content such as passive/active are available. Otherwise the impact of givenness is observed in the choice of word order, as is observed in languages with flexible word order (see Prat-Sala 1997, Tomlin 1997, Skopeteas and Fanselow 2009). The cross-linguistic complementarity between passive and marked word order indicates that passivization is a member of a set of syntactic operations that create a linearization in which the given information occurs early in the utterance.

The conclusion is that the inductive problem is a necessary concomitant of hypothesis-driven investigations. The only way to overcome this shortcoming is to enlarge the data base, i.e., to carry out further empirical studies that disentangle the possible effects of further relevant theoretical concepts.

4. Speakers' intuitions of contextual felicity

We have seen in section 2.2 that an inherent limitation of observational data is that they cannot empirically justify universal hypotheses. The observation that a linguistic expression does not occur under certain contextual conditions in our data set does not allow us to infer that this generalization holds for the totality of occurrences of this linguistic expression in the universe of discourse.

The study of language is particular in that it involves another type of evidence that allows us to justify universal statements. This is the role of intuition data. Linguistic intuition involves a number of different things: Knowledge of grammatically correct forms that are based on more or less general productive rules, but also knowledge of how forms are used in particular contexts. To the extent that speaker's intuition is reliable, a negative judgment confirms the truth of a universal negative hypothesis. The background assumption is that the psychological phenomenon of judging a context-expression pair as infelicitous is a generalization over the infinite number of potential context-expression pairs in discourse. Hence, speakers' intuition is a very particular type of meta-linguistic evidence that allows for proving deterministic hypotheses in overcoming the problem of induction in the observational data.

4.1. Overview

The basic phenomenon is the intuition that different grammatically well-formed structures are either felicitous or not felicitous in particular contexts. The concept of 'contextual felicity' refers to the intuition of the speaker whether the target utterance fits to the expectations created through the context. This intuition reflects the extent to which the target utterance presupposes the propositional content that is introduced through the context and presents an assertion that falls within the range of possible contributions to the discourse that are expected in the context.[5] For instance, Lambrecht (2001) observes that the well-formed target utterances in (15) differ with respect to their felicity in the presented context. He claims that the second target is not felicitous because it evokes an exhaustivity presupposition that is not licensed by the context.

(15) Context: *District attorney to potential juror in the trial of a black man: Do you think you might have any bias that would prevent you from*

---

[5] See Matthewson (2004) on empirical issues on the elicitation of contextual felicity and Rooth (1992. 84-85) and Krifka (2002, 2007) on the felicity of question-answer pairs.

> *reaching a finding of not guilty, given that the defendant is a*
> *black man?*
Target 1: *Why no. I have my NEIGHBOR (who's black).*
Target 2: *Why no. #It's my NEIGHBOR who's black.* (Lambrecht 2001: 505–506)

Kiss (1998) presents a test of exhaustive identification (with reference to Donka Farkas) that is based on the felicity of an utterance that follows the target utterance, see (16)-(17). The basic assumption is that the felicity of the latter utterance depends on whether it is among the logical consequences of the former one. In case the target utterance involves an exhaustive interpretation of the focused constituent, as is the case with the preverbal focus in Hungarian in (16), the negation of a statement rejecting the presupposition of exhaustivity is felicitous.

(16) Target 1: *Mari <u>egy kalapot</u> nézett ki   magának.*
             Mary a   hat.ACC picked out herself.DAT
             'It was a hat that Mary picked for herself.'
   Context: *Nem, egy kabátot is   ki   nézett.*
             no   a   coat   too out picked
             'No, she picked a coat, too.' (É. Kiss 1998: 251)

If the target utterance does not evoke an exhaustive interpretation, as is the case with the postverbal focus in Hungarian in (17), the negation of a statement that rejects the exhaustive interpretation is not felicitous.

(17) Target 1: *Mari ki   nézett   magának   EGY   KALAPOT.*
             Mary out picked   herself.DAT a   hat.ACC
             'It was a hat that Mary picked for herself.'
   Context: *#Nem, egy kabátot is   ki   nézett.*
             no   a   coat   too out picked
             'No, she picked a coat, too.' (É. Kiss 1998: 251)

Judgments of contextual felicity are certainly a very widely used type of evidence in the investigation of information structure (see also Krifka 2010: §2.8). In the last years, a paradigm of experimental studies evolved that is based on repeated observations of the intuition of contextual felicity in speakers' samples. Native speakers are presented pairs of context and target utterances and are instructed to judge whether the target utterance fits to the context. Judgments can be either categorical (target utterance fits or does not fit to the context) or scalar, either using a numerical scale, e.g., from '1 = utterance does not fit to the context at all' to '7 = utterance fits to the context very well' or using magnitude estimation

(Bard et al. 1996), in which case the participant is asked to assign a numerical value to a stimulus that serves as reference point and then to express his/her intuition of acceptability in numerical values relative to the reference point (see discussion of the alternatives in measuring acceptability intuitions in Weskott and Fanselow 2009).

Keller and Alexopoulou (2001) report on two experimental studies on the interaction of word order, clitic doubling, and intonation in Greek. Combinations of these factors are tested in several contextual conditions, i.e., absence of context, all-focus context, subject-focus context, object-focus context, and verb-focus context. The intended contextual conditions are established by means of questions, as exemplified in (18)-(19) for object questions. The background assumption is that in the context of an object question only those answers will be judged as felicitous that involve operations licensed by object focus. The examples (18)-(19) illustrate the effects of clitic doubling on the intuition of contextual felicity. Literature on Greek shows that clitic doubling, i.e., the resumption of a case complement of the verb by a coreferential pronoun, is felicitous when the complement is part of the given information (see Alexopoulou 1999: 46 and references therein). Both examples illustrate an SVO′ answer with accent (′) on the object in the context of an object question. Example (18) does not involve clitic doubling, hence it is expected to be felicitous in this context.

(18) Context: *ti   θa   ðiavási   o   tásos?*
  what   FUT   read:3.SG   DEF:NOM   Tasos:NOM
  'What will Tasos read?'
  Target: *o   tásos   θa   ðiavási*
    DEF:NOM   Tasos:NOM   FUT   read:3.SG
    tin   EFIMERIDA.
    DEF:ACC   newspaper:ACC
    'Tasos will read the newspaper.' (Keller and Alexopoulou 2001)

Example (19) illustrates a target utterance with clitic doubling presented in the same context, i.e., as answer to an object focus question. The effect of clitic doubling on the intuition of contextual felicity is reflected in the average judgments reported in Keller and Alexopoulou (2001): SVO′ answers (without clitic doubling) obtain the value 0.54 (S.E. = 0.05) in the normalized log-transformed data, while SclVO′ answers (with clitic doubling) obtain the value 0.12 (S.E. = 0.08).

(19) Context: *ti θa ðiavási o tásos?*
  Target: *o   tásos   θa   tin   ðiavási*
    DEF:NOM   Tasos:NOM   FUT   3.SG:ACC   read:3.SG

tin   *EFIMERIDA.*
DEF:ACC   newspaper:ACC
'Tasos will read the newspaper.' (Keller and Alexopoulou 2001)

Quantitative studies on contextual felicity have been carried out on several issues relating to the interaction of particular properties of the utterance with the context. For instance, Birch and Clifton (1995) present experimental judgments on the appropriateness of several prosodic patterns in the context of different question types in English; Arnold (1998) examines the effects of English clefts on establishing discourse topics by observing the acceptability of pronominal and lexical mentioning of referents; Keller (2000) reports the findings of a rating experiment on the interaction of word order and pronominalization in several contexts in German subordinate clauses; Skopeteas, Féry, and Asatiani (2009) study the influence of the context on the felicity of several word orders, prosodic patterns and morphological properties (case inversion) in Georgian.

4.2. Empirical statements

A crucial issue with respect to the empirical findings of quantitative studies is the observation that contextual felicity involves gradience. For instance, we have seen in section 4.1 that SVO′ answers are judged to be more felicitous than SclVO′ answers in the context of object questions in Greek. The mean values reported in 4.1 are obtained when accentual prominence is realized on the constituent in question. When the prosodic structure is non-congruent with the context, e.g., when accentual prominence is realized on the subject, then a further negative effect is observed in the values that are assigned by the native speakers (average judgments for S′VO obtain the value 0.2 and judgments for S′clVO obtain the value 0.08, see Keller and Alexopoulou 2001). These findings indicate that the empirical phenomenon of contextual felicity is not categorical (±felicitous) but scalar. The degree of felicity observed in these results may be accounted for through the fact that they reflect the influence of two factors (clitic doubling and accentuation) that have a cumulative effect on the intuition of felicity (see also Sorace and Keller 2003).

The observation that the intuition of felicity is a scalar psychological phenomenon does not imply that the licensing of particular operations for the expression of particular information structural concepts is a gradient notion. Assuming that random sources of variation are outbalanced in the experimental design (e.g., different speaker, different lexicalizations of a construction, etc. see Schütze 1996), the observed gradience in speakers' intuitions may result from the fact that speakers perceive multiple violations of the licensing conditions as

reducing the likelihood that the presented stimulus will occur in the context at issue (see further discussion concerning the difference between acceptability and grammaticality in Schütze 1996 §3.3.1, Vogel 2005, Weskott and Fanselow 2009). Moreover, further sources of variation contribute to the gradience of experimental results. For instance, people differ in their willingness to invent larger contexts than the one actually given, which may boost the acceptance of a linguistic form.

Several recent studies have pointed out that the gradience obtained through speakers' intuition is not directly mapped onto the observational data that we can obtain from corpora (see Featherston 2005, Kempen and Harbush 2005, 2006). The text counts do not show a frequency distribution that corresponds closely to the felicity degree observable in scalar judgments. Text counts rather lead to categorical results, in which the optimal construction is frequently attested and the further alternatives either display very few tokens or are not attested at all. This discrepancy is expected, since in speech production speakers tend to select the optimal candidate, such that suboptimal alternatives have very few chances to occur (see Featherston 2005).

5. Conclusions

This article presented the main paradigms of empirical studies that are used for the investigation of information structure. The particular issue in the methods used for this linguistic domain is that the main evidence for the assumption of information structural concepts comes from the context. Section 2 presented several types of empirical studies that are based on the observation of naturalistic data. Section 3 discussed data from language production experiments. Section 4 discussed evidence from the intuition of contextual felicity – either based on singular observations or on experimental studies.

The two former paradigms of empirical methods, corpus studies and production experiments, constitute alternative approaches to the same phenomenon, namely the speaker's choice of particular properties of linguistic expressions depending on the context. Corpus studies provide evidence from naturalistic data, hence allowing for observation of the context-expression dependencies in real communication. Experimental studies in speech production have the advantage that they allow for controlling the sources of variation and comparing the exact effect of discourse factors that may influence the form of the utterance.

Production data differ from data obtained through speakers' judgments. In this case, we are not dealing with alternative methodological approaches but with different phenomena. Rating experiments are dealing with the intuition of contextual felicity which is a phenomenon independent of corpus frequencies (see section 4.2). Since utterances are complex configurations that may involve choices

at several layers (e.g., prosodic structure, syntax, morphological phenomena, etc.), contextual felicity comes up as a gradient phenomenon reflecting the felicity of several properties of a single expression. The gradience observed in intuition data cannot be directly mapped on the frequency distributions that result from the corpus data, since the latter show a predominance of the most felicitous option in each context.

Intuition data are complementary to production data in two respects. First, they allow us to justify universal negative hypotheses (or to reject existential positive hypotheses), which is not possible through production data due to the inherent limitations of the problem of induction. Second, intuition data reveal the gradience of the interacting factors that cannot be always observed in the production data, since speakers' production is necessarily based on the choice of the single optimal linguistic expression for a given propositional content.

References

Alexopoulou, Theodora 1999 *The syntax of discourse-functions in Greek. A non-configurational approach.* Ph.D. dissertation, University of Edinburgh.

Arnold, Jennifer E.   1998 Reference form and discourse patterns. Ph.D. Stanford University.

Arnold, Jennifer, Thomas Wasow, Anthony Losongco and Ryan Ginstrom 2000 Heaviness vs. newness: the effects of structural complexity and discourse status on constituent ordering. *Language* 76: 28–55.

Bader, Markus and Jana Häussler   2010 Word Order in German: A corpus study. *Lingua* 120.3: 717–762.

Bard, Ellen Gurman, Dan Robertson and Antonella Sorace   1996. Magnitude estimation of linguistic acceptability. *Language* 72: 32–68.

Beaver, David   2004. The optimization of discourse anaphora. *Linguistics and Philosophy* 27: 1–53.

Birch, Stacy and Charles Clifton, Jr. 1995 Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech* 38.4: 365–391.

Birner, Betty J.   1994 Information status and word order: An analysis of English inversion. *Language* 70.2: 233–259.

Bock, J. Kathryn   1977 The effect of a pragmatic presupposition on syntactic structure in question answering. *Journal of Verbal Learning and Verbal Behavior* 16: 723–734.

Bock, J. Kathryn and David E. Irwin 1980 Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior* 19: 467–484.

Bock, J. Kathryn and Richard K. Warren   1985 Conceptual accessibility and syntactic structure in sentence formulation. *Cognition* 21: 47–67.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina and R. Harald Baayen   2007 Predicting the dative alternation. In: Gerlof Bouma, Irene Kraemer and Joost Zwarts (eds.), *Cognitive Foundations of Interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, Joan and Jennifer Hay   2007 Gradient grammar: An effect of animacy on the syntax of *give* in New Zealand and American English. *Lingua*, doi:10.1016/j.lingua.2007.02.007.

Brunetti, Lisa   2009a Discourse Functions of Fronted Foci in Italian and Spanish. In: Andreas Dufter and Daniel Jacob (eds.), *Focus and Background in Romance Languages*, 43–82. Amsterdam: Benjamins.

Brunetti, Lisa   2009b On the semantic and contextual factors that determine topic selection in Italian and Spanish. *The Linguistic Review* 26: 261–289.

Christianson, Keil and Fernanda Ferreira   2005 Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition* 98: 105–135.

Clark, Herbert   1977 Bridging. In: Philip N. Johnson-Laird and Peter C. Wason (eds.), *Thinking: Readings in Cognitive Science*, 411–420. London: Cambridge University Press.

Clark, Herbert H. and Susan E. Brennan   1991 Grounding in communication. In: Lauren B. Resnick, John M. Levine and Stephanie D. Teasley (eds.),

Perspectives on socially shared cognition, 127–149. Washington: American Psychological Association.

Clark, Herbert H. and Susan Haviland   1977 Comprehension and the given-new contrast. In: Roy O. Freedle (ed.), *Discourse Production and Comprehension*, 1–40. Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, Herbert H. and Deanna Wilkes-Gibbs   1986 Referring as a collaborative process. *Cognition* 22: 1–39.

Cooreman, Ann   1987 *Transitivity and Discourse Continuity in Chamorro Narratives*. Berlin, Mouton de Gruyter.

Davison, Alice   1984 Syntactic markedness and the definition of sentence topic. *Language* 60: 797–846.

Dretske, Fred   1972. Contrastive statements. *The Philosophical Review* 81: 411–437.

Dryer, Matthew S. 1994 The discourse function of Kutenai inverse. In: Givón (ed.), 65–99.

É. Kiss, Katalin   1998 Identificational vs. information focus. *Language* 74.2: 245–273.

Fanselow, Gisbert 2009 Die (generative) Syntax in den Zeiten der Empiriediskussion. *Zeitschrift für Sprachwissenschaft* 28: 133–139.

Featherston, Sam 2005 The decathlon model: Design features for an empirical syntax. In: Stephan Kepser and Marga Reis (eds.), *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, 187–208. Berlin: Mouton de Gruyter.

Featherston, Sam 2009 Relax, lean back, and be a linguist. *Zeitschrift für Sprachwissenschaft* 28: 127–132.

Ferreira, Victor S. and Hiromi Yoshita   2003 Given-New Ordering Effects on the Production of Scrambled Sentences in Japanese. *Journal of Psycholinguistic Research* 32.6: 669–692.

Feyerabend, Paul 1983 *Wider den Methodenzwang*. Frankfurt: Suhrkamp.

Forrest, Linda B. 1997 Discourse goals and attentional processes in sentence production: The dynamic construal of events. In: Adele E. Goldberg (ed.), *Conceptual Structure, Discourse and Language*, 149–162. Stanford, CA: CSLI Publications.

Givón, Talmy (ed.) 1994 *Voice and inversion*. Amsterdam/Philadelphia: Benjamins.

Goodman, Nelson 1954 *Fact, Fiction and Forecast*. London: Athlone Press.

Grosz, Barbara J. and Candace L. Sidner 1986 Attentions, intentions, and the structure of discourse. *Computational Linguistics* 12: 175–204.

Gundel, Jeanette K. 1974 The role of topic and comment in linguistic theory. Ph.D. dissertation, Univ. of Texas.

Gundel, Jeanette K., Nancy Hedberg and Ron Zacharski 2005 Pronouns without explicit referents: How do we know when a pronoun is referential? In: Antonio Branco, Tony McEnery and Ruslan Mitkov (eds.), *Anaphora Processing: Linguistic, Cognitive, and Computational Modelling*, 351–364. Amsterdam: Benjamins.

Haspelmath, Martin 2009 Welche Fragen können wir mit herkömmlichen Daten beantworten? *Zeitschrift für Sprachwissenschaft* 28: 157–162.

Herring, Susan C. 1994 Afterthoughts, antitopics, and emphasis: the syntacticization of postverbal position in Tamil. In: Miriam Butt, Tracy Holloway King, Gillian Ramchand (eds.), *Theoretical perspectives on word order in Asian languages*, 119–152. Stanford: CSLI.

Herring, Susan C. and John C. Paolillo 1995 Focus position in SOV languages. In: Pamela Downing and Michael Noonan (eds.), *Word order in discourse*, 163–198. Amsterdam: Benjamins.

Hörnig, Robin and Caroline Féry 2009 Linguistic Markers of Discourse Status in Describing Altered Spatial Layouts. Ms., Univ. of Potsdam.

Horvath, Julia 1986 *FOCUS in the Theory of Grammar and the Syntax of Hungarian*. Dordrecht: Foris.

Isaacs, Ellen A. and Herbert H. Clark 1987 References in conversation between experts and novices. *Journal of Experimental Psychology* 116.1: 26–37.

Johnson-Laird, Philip N. 1968 The choice of the passive voice in a communicative task. *British Journal of Psychology* 59: 7–15.

Kaiser, Elsi 2006 Effects of topic and focus on salience. In Christian Ebert and Cornelia Endriss (eds.), *Proceedings of Sinn und Bedeutung 10*, 139–154, (ZAS Working Papers in Linguistics vol. 44). Berlin.

Keller, Frank 2000 Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality. Ph.D. dissertation, University of Edinburgh.

Keller, Frank, Alexopoulou, Theodora 2001 Phonology competes with syntax: Experimental evidence for the interaction of word order and accent placement in the realization of information structure. *Cognition* 79.3: 301–371.

Kempen, Gerard, Harbusch, Karin 2005 The relationship between grammaticality ratings and corpus frequencies: A case study into word-order variability in the midfield of German clauses. In: Stephan Kepser and Marga Reis (eds.), *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, 329–349. Berlin: de Gruyter.

Kempen, Gerard, Harbusch, Karin 2006 Comparing Linguistic Judgments and Corpus Frequencies as Windows on Grammatical Competence: A Study of Argument Linearization in German Clauses. Ms., Nijmegen and Koblenz.

Krifka, Manfred 2002 For a structured meaning account of questions and answers. In: Caroline Féry and Wolfgang Sternefeld (eds.), *Audiatur Vox Sapientiae: A Festschrift for Arnim von Stechow*, 287–319. Berlin: Akademie.

Krifka, Manfred 2007 The semantics of questions and the focusation of answers. In: Chungmin Lee, Matthew Gordon and Daniel Büring (eds.), *Topic and Focus*. Dordrecht: Kluwer. 139–150.

Krifka, Manfred 2010 Varieties of semantic evidence. In: Claudia Maienborn, Paul Portner and Klaus von Heusinger (eds.), *Handbook of Semantics*. Berlin: Mouton de Gruyter.

Kuhn, Thomas S. 1970 Logic of discovery or psychology of research? In: Imre Lakatos and Alan Musgrave (eds.), *Criticism and the Growth of Knowledge*, 1–23. Cambridge: Cambridge University Press.

Lambrecht, Knud 2001 A framework for the analysis of cleft constructions. *Linguistics* 39.3: 413–516.

MacWhinney, Brian and Elizabeth Bates 1978 Sentential Devices for Conveying Givenness and Newness: A Cross-Cultural Developmental Study. *Journal of Verbal Learning and Verbal Behavior* 17: 539–558.

Mathesius, Vilém 1975 *A functional analysis of present day English on a general linguistic basis* [ed. by J. Vachek]. The Hague, Paris: Mouton.

Matić, Dejan 2003 Topics, presuppositions, and theticity: An empirical study on Verb-Subject Clauses in Albanian, Greek, and Serbo-Croatian. Ph. D. dissertation, University of Cologne.

Matthewson, Lisa 2004 On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70.4: 369–415.

Myachykov, Andriy, Russel S. Tomlin and Michael I. Posner 2005 Attention and empirical studies of grammar. *The Linguistic Review* 22: 347–364.

Newton-Smith, William H. 2000 Underdetermination of Theory by data. In: William H. Newton-Smith (ed.), *A Companion to the Philosophy of Science*, 532–536. Oxford: Blackwell.

Newmeyer, Frederick J. 2003 Grammar is grammar and usage is usage. *Language* 79.4: 682–707.

Onea, Edgar and Klaus von Heusinger 2009 Grammatical and contextual restrictions on focal alternatives. In: Daniel Jacob and Andreas Dufter (eds.), *Focus and Background in Romance Languages*, 281–308. Amsterdam: Benjamins.

Paul, Hermann 1880 *Prinzipien der Sprachgeschichte*. Tübingen: Niemeyer.

Poesio, Massimo, Rosemary Stevenson, Barbara Di Eugenio and Janet Hitzeman 2004 Centering: A parametric theory and its instantiations. *Computational Linguistics* 30.3: 309–363.

Popper, Karl   1934 *Logik der Forschung* [¹¹2005]. Tübingen: Siebeck.

Prat-Sala, Mercè 1997 *The production of different word orders: A psycholinguistic and developmental approach*. Ph.D. dissertation, University of Edinburgh.

Prat-Sala, Mercé and Holy P. Branigan   2000 Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language* 42: 168–182.

Prentice, J. L. 1967   Effects of cueing actor vs. cueing object on word order in sentence production. *Psychonomic Science* 8: 163–164.

Prince, Ellen   1981 Toward a taxonomy of given-new information. In: Peter Cole (ed.), *Radical pragmatics*, 223–256. New York: Academic Press.

Quesada, Diego and Skopeteas, Stavros   2010 The discourse functions of inversion: an experimental study on Teribe (Chibchan). *Journal of Pragmatics* (doi:10.1016/ j.pragma.2010.02.004).

Roland, Katy   1994 The pragmatics of Modern Greek voice. In: Givón (ed.), 233–260.

Rooth, Mats 1992 A theory of focus interpretation. *Natural Language Semantics* 1: 75–116.

Sampson, Geoffrey R.   2007 Grammar without grammaticality. *Corpus Linguistics and linguistic Theory* 3.1: 1–32.

Schütze, Carson T.   1996 The empirical base of linguistics: Grammaticality judgments and linguistic methodology. Chicago: University of Chicago Press.

Skopeteas, Stavros Gisbert Fanselow 2009 Effects of givenness and constraints on free word order. In: Malte Zimmerman and Caroline Féry (eds.), *Information Structure from different perspectives*, 307–331. Oxford: Oxford University Press.

Skopeteas, Stavros and Gisbert Fanselow   2010 Focus types and argument asymmetries: a cross-linguistic study in language production. In: Carsten Breul and Edward Göbbel (eds.), *Comparative and contrastive studies of information structure*. Amsterdam, Philadelphia: Benjamins.

Skopeteas, Stavros and Caroline Féry 2007 Contrastive Topics in Pairing Answers: A Cross-Linguistic Production Study. In: Sam Featherston, Wolfgang Sternefeld (eds.), *Linguistic Evidence 2006*, 327–347. Berlin, New York: Mouton De Gruyter.

Skopeteas, Stavros, Caroline Féry, Rusudan Asatiani 2009 Word Order and Intonation in Georgian. *Lingua* 119: 102–127.

Skopeteas, Stavros, Ines Fiedler, Sam Hellmuth, Anne Schwarz, Ruben Stoel, Gisbert Fanselow, Caroline Féry and Manfred Krifka 2006 *Questionnaire on Information Structure (QUIS)*. (Interdisciplinary Studies on Information Structure, vol. 4). Universitätsverlag Potsdam, Germany.

Sloman, Steven A. and Lagnado, David A. 2005 The problem of induction. In Keith Holyoak, Robert Morrison (eds.), *Cambridge handbook of thinking and reasoning*, 95–116. Cambridge: Cambridge University Press.

Sorace, Antonella and Keller, Frank 2005 Gradience in linguistic data. *Lingua* 115: 1497–1524.

Speyer, Augustin 2007 Die Bedeutung der Centering Theory für Fragen der Vorfeldbesetzung im Deutschen. *Zeitschrift für Sprachwissenschaft* 26: 83–115.

Szabolcsi, Anna 1981 Compositionality in focus. *Folia Linguistica* 15: 141–163.

Tomlin, Russell S. 1995 Focal attention, voice and word order: an experimental, cross-linguistic study. In: Pamela A. Downing and Michael Noonan (eds.), *Word Order in Discourse*, 517–554. Amsterdam: Benjamins.

Tomlin, Russell S. 1997 Mapping conceptual representations into linguistic representations: The role of attention in grammar. In: Jan Nuyts and Eric Pederson (eds.), *Language and Conceptualization*, 162–189. Cambridge: Cambridge University Press.

Vogel, Ralf 2005 Degraded Acceptability and Markedness in Syntax, and the Stochastic Interpretation of Optimality Theory. In: Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky and Ralf Vogel (eds.), *Gradience in Grammar*, 246–269. Oxford: Oxford University Press.

Vion, Monique and Annie Colas 1995 Contrastive Marking in French Dialogue: Why and How. *Journal of Psycholinguistic Research* 24.5: 313–331.

Walker, Marilyn A., Aravind K. Joshi and Ellen F. Prince (eds.) 1998 *Centering Theory in Discourse.* Oxford: Clarendon Press.

Ward, Gregory L. and Ellen F. Prince 1991 On the Topicalization of Indefinite NPs. *Journal of Pragmatics* 16: 167–177.

Weber, Andrea and Karin Müller 2004 Word order variation in German main clauses: A corpus analysis. *Proceedings of the 20th International conference on Computational Linguistics.* Geneva, Italy, 71–77.

Wedgwood, Daniel, Gergely Pethő and Ronnie Cann 2003 Hungarian 'focus position' and English *it*-clefts: the semantic underspecification of 'focus' readings. Ms., Univ. of Edinburgh.

Weskott, Thomas and Gisbert Fanselow 2009 Scaling issues in the measurement of linguistic acceptability. In: Sam Featherston and Susanne Winkler (eds.), *The fruits of empirical linguistics, Volume 1: Process*, 229–245. Berlin: Mouton De Gruyter.

Weskott, Thomas, Robin Hörnig, Elsi Kaiser, Caroline Féry, Sabine Kern, Gisbert Fanselow and Reinhold Kliegl 2006 Information Structure and the Anticipation of Discourse Referents. Ms., University of Potsdam.